Rationality as a Virtue

Ralph Wedgwood

University of Southern California

I. The concept of rationality

There seems to be a concept, which can be expressed in philosophical English by the term ‘rationality’, that plays a particularly central role both in epistemology and in ethics. This concept seems especially prominent in the more formal branches of epistemology and decision theory, since it seems to be a central topic for those branches of philosophy to investigate when beliefs and decisions – and events in which we revise or adjust our beliefs and intentions – count as rational. But it is not completely obvious what sort of concept the term ‘rational’ expresses in these contexts.

The word ‘rational’ derives from a Latin word that refers to the faculty of reasoning (or thinking or calculation). However, most formal epistemologists and decision theorists seem to assume that they are using the word to express a normative concept of some kind – in effect, to express a concept that has something to do with the proper use of this faculty of reasoning. Even if this assumption is correct, we may still wonder: exactly what normative concept is the concept of rationality? How is it related to other normative concepts?

In this essay, I shall offer an interpretation of the concept that the term expresses in these contexts. As I shall argue, when the term is used in this way, the concept of “rationality” is the concept of a certain kind of virtue. Moreover, we can gain insights into several of the concept’s features by exploring some of the classic discussions of virtue within the history of ethics; I shall focus here particularly on discussions of virtue in the works of Plato, Aristotle, Kant, and the 18th-century British moralist Richard Price.

Furthermore, as I shall also argue, interpreting the concept of “rationality” as referring to a kind of virtue helps us to solve some of the problems that arise when we theorize with this concept. For example, this interpretation helps us to understand the relations between “rationality” and “rational requirements”, and the distinction that epistemologists often signal by the terms “propositional” and “doxastic justification”.

Finally, interpreting rationality in this way will help us to answer some of the objections that have been raised against the thesis that the term ‘rational’, as it is used in these contexts in epistemology and decision theory, expresses a normative concept of any kind. In particular, I shall argue that this interpretation helps us to answer the following objection. It has seemed plausible to many formal epistemologists and decision theorists that rationality involves having mental states with certain formal features – such as consistency or probabilistic coherence in one’s beliefs, or preferences that meet certain so-called “axioms” like transitivity, monotonicity, stochastic dominance, and the like.

However, it is not obviously even possible for ordinary agents to have mental states with these formal features. If “rationality” is a normative concept, would not the claim that rationality

---

1 Another philosopher who has argued that rationality is a virtue is Judith Thomson (2008, 80) – although as I note in note 3 below, Thomson’s conception of virtue is somewhat different from mine.

2 See e.g. James M. Joyce (1999, chap. 2).
requires these formal features conflict with the principle that ‘ought’ implies ‘can’? As I shall argue, understanding rationality as a kind of virtue will help us to find a solution to this problem.

II. The structure of virtue

There are many terms that stand for types of virtue. The most famous examples are terms for the four so-called “cardinal virtues” that are discussed in Book IV of Plato’s Republic (427e10): wisdom or prudence, justice, temperance, and courage or fortitude. To these four cardinal virtues, Christian theology later added three “theological virtues” – faith, hope, and charity. While many contemporary moralists might balk at including faith and hope among the fundamental virtues, they would all agree in including something like charity, kindness, or beneficence, alongside the cardinal virtues.

All the adjectival terms that can be used to ascribe these virtues – that is, terms like ‘just’, ‘wise’, ‘courageous’ and so on – can, it seems, be applied both to people and to other items such as acts and decisions. There are just people and courageous people, and also just acts and courageous decisions, and so on. This seems to show that to each of these virtues there in fact correspond at least two properties – one property that can be exemplified by people or other agents, and another property that can be exemplified by particular acts and decisions.

In fact, however, virtue-terms can be applied much more widely. For example, earlier on in the Republic (352e-353c), Plato explicitly speaks of the “virtues” of eyes and ears, and implicitly commits himself to the thesis that there are distinctive “virtues” of horses and pruning knives. It is clear from the context that the virtue of a pair of eyes is whatever feature of those eyes makes them in the relevant sense good eyes; and similarly it is clear from the context that the virtue of a horse or a pruning knife is whatever feature of the item in question makes it in the relevant sense a good horse or a good pruning knife.

A similar point emerges from a famous passage in Aristotle’s Nicomachean Ethics (1098a10), where Aristotle commits himself to the thesis that there is a distinctive “virtue” of harpists. The point that Aristotle is making here is clearly intended to hold generally of all craftsmen – such as flute-players, sculptors, carpenters, and leather-workers (1097a26-31) – and also of parts of the body – such as the hand, the eye, and the foot (1097a32).

In both these passages, Plato and Aristotle are concerned to argue for a connection between the virtues of a thing and the thing’s function. I shall not explore this aspect of their thinking here. I shall focus instead on a more general feature of the way in which Plato and Aristotle seem both to be thinking of virtue in these passages. For our purposes, the crucial feature is that every virtue has two aspects. First, there is one kind of virtue-property which is essentially a disposition, which is in a way good. Second, this dispositional virtue-property is characteristically manifested in performances that are also in a corresponding way good.

This structure, involving good dispositions that are manifested in correspondingly good performances, is found both in the case of the classical cardinal virtues and in the more humdrum

---

5 Judith Thomson (2008, 69) also emphasizes that the term ‘virtue’ can be used broadly (indeed, in some ways, her use is even broader than mine, since unlike me, she does not restrict the term ‘virtue’ to dispositional traits that are manifested in good performances).

4 I interpret the term ἔξις as it appears in Aristotle’s account of virtue (Nicomachean Ethics 1105b20-1106a14) as referring to a disposition. For discussion, see Bostock (2000, 36–8).
cases of good pruning knives and good eyes. First, when a pruning knife possesses the distinctive virtue of pruning knives – that is, when it counts as a good pruning knife – this is because of the dispositions that the knife has, the sharpness of its blade, the manageability of its handle, and so on. Secondly, the goodness or excellence of the pruning knife is manifested in performances in which the pruning knife displays its excellence – that is, presumably, on occasions when the knife proves itself to be an effective and convenient tool to use for pruning vines (and other plants that agents normally use knives to prune).

In general, this pattern, involving both (i) a good disposition and (ii) correspondingly good manifestations of the disposition, seems to be a general feature of all virtues. In this way, every virtue involves at least these two kinds of goodness. In fact, however, there are compelling reasons to think that there are in fact, not two, but three kinds of goodness involved in every virtue.

This point is also familiar from the history of ethics. In particular, in the *Nicomachean Ethics* (1105a17-b9), Aristotle distinguishes between just acts and acts that are done justly. According to Aristotle, an unjust person might, by chance, do a just act; but an act can be done justly only if it is a manifestation of the appropriate disposition – specifically, a disposition that is involved in being (to at least some degree) a just person.\(^5\)

It is clear from the context that Aristotle supposes that the same distinction can be drawn with respect to every virtue that can be expressed in acts at all. On reflection, it seems clearly possible for someone to do a beneficent act by mistake: for example, out of a combination of malevolence and incompetence, an agent might end up doing a beneficent act – without doing it from the virtue of beneficence, by manifesting the dispositions that are constitutive of the virtue of beneficence. Indeed, the same distinction can be drawn with respect to virtues that are not expressed in acts at all, but rather in mental events of other kinds, like choices or decisions. By sheer chance, an imprudent person might make the prudent decision, even though they did not make this decision by manifesting the virtue of prudence.

A very similar distinction was rediscovered later in the history of ethics, by Richard Price in his treatise of 1757, *A Review of the Principal Questions of Morals*. In Chap. VIII of this treatise, he introduces a distinction between “abstract virtue” and “practical virtue”.\(^6\) (One of Price’s main reasons for introducing this distinction is to rebut Hume’s claim from *A Treatise of Human Nature*, III.i.1, that the primary motive of a virtuous action cannot be “a regard to the virtue of that action, but must be some other natural motive or principle”.)

According to Price, an act is abstractly virtuous purely because it is of a type that fits the agent’s circumstances; an act is practically virtuous only if it is done because the agent “regards” the act as abstractly virtuous. Arguably, if you do an act precisely because you regard the act as abstractly virtuous, your act must be the manifestation of a disposition that you have, to respond to your beliefs about which of the available acts are abstractly virtuous by doing one of the acts that you believe to be abstractly virtuous. So Price’s account of the distinction seems also, at least implicitly, to involve the manifestation of appropriate dispositions.

---

\(^5\) Of course, an unjust person might, on a few rare occasions, act justly. But even in this case, acting justly involves manifesting a disposition towards just acts – it is simply that in the case of this unjust person, this disposition is normally blocked or inhibited by counteracting dispositions. Nonetheless, for an unjust person to act justly, they must have at least some disposition towards justice.

\(^6\) The relevant passages of Price’s 1757 work can be found in Raphael (1969, §§ 750–762).
Clearly, it is possible for one to do an abstractly virtuous act by chance, without doing it because one believes it to be abstractly virtuous (which is what Price considers to be the motive of practical virtue) – whereas as we have seen, it is not possible to act in a practically virtuous manner without manifesting the appropriate disposition. Thus, Price’s distinction seems fundamentally to coincide with Aristotle’s.

In general, then, each virtue 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 (a kind of goodness that could in principle be exemplified by performances that result by chance, rather than from this disposition).

It seems intuitively clear that we can draw an exactly parallel distinction in the case of rationality. That is, we can distinguish between (i) rational dispositions, (ii) manifestations of these rational dispositions, and (iii) “abstract rationality”. So, for example, an irrational person might, by chance, make the choice that it is abstractly rational for him to make, without thereby manifesting any rational dispositions that this person possesses.

In fact, this distinction between abstract rationality and the manifestation of rational dispositions seems to be the very same distinction as the one that contemporary epistemologists draw between “propositional justification” and “doxastic justification”. To say that there is propositional justification for a thinker to believe a proposition p is in effect to say that it is abstractly rational for the thinker to believe p. To say that the thinker’s belief in p is doxastically justified is to say that in holding this belief, the thinker is manifesting the appropriate rational dispositions.

This interpretation of the distinction between propositional and doxastic justification might be disputed. But the interpretation can be defended in the following way. As these terms are used by contemporary epistemologists, a thinker can have propositional justification for believing a proposition p even if she never in fact believes p – since to say that she has propositional justification for believing p is just to say that the attitude-type believing p is part of a set of attitudes that optimally “fits” the thinker’s cognitive situation (or “fits her evidence”, as many epistemologists would say). It is also possible for the thinker to believe p at the same time as having propositional justification for believing p, even if it is a sheer fluke that on this occasion the thinker holds a belief that she has propositional justification for. In other words, even if the thinker has propositional justification for a belief that she actually holds, it does not follow that this belief is rationally or justifiedly held by the thinker.

For the thinker to believe p in a rational or justified manner – that is, for this belief to be doxastically justified – it must not be a fluke that the thinker’s belief is one that she has propositional justification for; on the contrary, the thinker’s holding of the belief must be the manifestation of some dispositions that in a non-accidental way tend to result in her holding beliefs that she has such propositional justification for. In this way, the distinction between propositional and doxastic justification lines up exactly with the distinction between the propositions that it is abstractly

---

7 For an interesting contemporary discussion of this distinction within epistemology, see Turri (2010).
rational for the thinker to believe and the beliefs that count as manifestations of the thinker’s rational dispositions.

III. Abstract rationality and rational dispositions

What is the relation between abstract rationality and rational dispositions? There are two main proposals that philosophers have made about how the two are connected:

i. It is abstractly rational for an agent to φ if and only if there is an available way for the agent to respond to her situation that would consist of the manifestation of rational dispositions, and would result in the agent’s φ-ing.

ii. Rational dispositions are all and only those dispositions that reliably lead the agent to respond to her situation in ways in which it is abstractly rational for her to respond.

The first of these connections is the one that is highlighted by Bernard Williams’s (1995, 35) talk of a “sound deliberative route” from the agent’s current state of mind to her rationally φ-ing. The second connection is the one that is, in effect, highlighted by the way in which Richard Price characterizes “practical virtue” in terms of “abstract virtue”, or by the way in which contemporary epistemologists characterize doxastic justification in terms of propositional justification.

As a matter of fact, I am inclined to think that both of these two proposals are at least roughly correct: abstract rationality and rational dispositions are indeed connected in something like these two ways. Some philosophers may be tempted to doubt one or the other (or indeed both) of these two connections; but just for the sake of argument, I shall for the rest of this essay proceed on the assumption that both of these two proposals – or at least something like them – are correct.

Some philosophers would make even stronger claims about one of these two connections, by claiming that one of these connections actually defines either the notion of a rational disposition, or the notion of abstract rationality, in terms of the other. On an approach of this kind, one of these two notions is taken as prior to the other, and one of these two connections is taken as definitional – while the other connection is either taken as derivative or rejected as illusory. However, at least on the assumption that something like both of these connections hold (and neither connection is illusory), both of these approaches face some significant problems.

First, consider the approach that takes the notion of a rational disposition as prior, and uses the first of the two connections (i) that I specified above to define the notion of abstract rationality. It seems that this approach would be debarred from characterizing rational dispositions in terms of the abstract rationality of the responses to which they give rise. But how else could we characterize such rational dispositions? Theorists who take this approach seems to face a dilemma. Either they give no characterization of these dispositions – in which case, they are in effect viewing these rational dispositions as mysterious “black boxes”, which somehow emit rational responses in a way of which no theoretical account can be given; or else they give a characterization of these dispositions, but without mentioning the abstract rationality of the responses to which they give rise – in which case they have no account of what these dispositions have in common, in virtue of

---

8 For an example of the first approach, see the definition of reasons that is given by Kieran Setiya (2007, 12). For examples of the second approach, consider the dominant approach among epistemologists, such as Earl Conee and Richard Feldman (1985), which is to define doxastic justification in terms of an allegedly prior notion of propositional justification.
which they all count as rational dispositions. The only way to avoid this dilemma, it seems, is to abandon the idea that the first connection (i) gives a full-fledged definition of abstract rationality.

Secondly, consider the approach that takes the notion of abstract rationality as prior, and uses the second of the two connections (ii) that I specified above to define the notion of a rational disposition. The trouble with this approach is that there now seems to be no way to derive the first connection (i) from the second connection (ii). However, even if it cannot serve as a definition, the first connection (i) still seems roughly plausible. That is, in effect, it is plausible that abstract rationality is constrained by the availability of a way of responding to your situation that would consist of the manifestation of rational dispositions: roughly, it cannot be abstractly rational for you to have a certain attitude unless there is an available way of responding to one’s situation, consisting of the manifestations of rational dispositions, that would lead to your having that attitude. Since this second approach has no way of explaining why abstract rationality is constrained in this way – that is, no way of deriving the first connection (i) from the second connection (ii) – the second approach seems to be in tension with the plausibility of the first connection (i).

For these reasons, it seems to me that a “no-priority” view may be more plausible. On this view, the two notions of abstract rationality and of rational dispositions are equally basic, neither of them being prior to the other – although as I have explained, I shall assume here that the two notions are tightly connected to each other by something like these two connections (i) and (ii).

IV. Rationality and the other virtues

In Section II, we identified a certain pattern in all virtues, including rationality. This pattern involves three kinds of goodness: (i) the goodness of a disposition, (ii) the goodness of the manifestations of that disposition, and (iii) the “abstract” goodness that is non-accidentally exemplified by those manifestations, but could also be exemplified by phenomena that result by chance, and not from good dispositions of this sort. This general pattern in fact seems to be present in all virtues – including the kind of “virtue” that we find in good pruning knives and good eyes, as well as in the virtue of rationality.

However, the most famous examples of the virtues – like the “cardinal virtues” of wisdom, justice, courage, and temperance – are more distinctly agential. These virtues are dispositions of agents. Moreover, it has seemed plausible, at least to many theorists, that the manifestation of these dispositions is “up to” the agent – in the sense that it is whenever an agent manifests either a virtuous disposition or a vicious disposition, it within the agent’s power of the agent to act or think in some ways that accord with these virtues and also within the agent’s power to act or think otherwise. Moreover, agents who manifest virtuous dispositions seem to be praiseworthy in a distinctive way in which only agents can be praiseworthy, and agents who manifest vicious

---

9 There is an analogous problem for the kind of “virtue ethics” – such as that of Rosalind Hursthouse (1999) – that seeks to define virtuous acts in terms of what it is for a person to have a virtuous character.

10 Is the analogue of this first connection (i) plausible for other virtues too? Or could an act be (for example) abstractly just even if it was not possible for anyone to perform this just act by manifesting the characteristic dispositions of a just person? These are important questions, which will decide whether there is a parallel argument for a “no-priority” view in the case of these other virtues; but unfortunately we cannot pause to examine these questions here.

11 This is a point that Hume misses, although it is clearly appreciated in most of the Western tradition of moral philosophy. See Irwin (2008, § 776).

12 I have defended this claim in some of my earlier work; see Wedgwood (2013a).
dispositions seem to be worthy of criticism of a correspondingly agential kind. (To say that manifestations of these virtues are praiseworthy in a distinctively agential way is not necessarily to say that they are morally praiseworthy – in the distinctively interpersonal and relational way in which we praise those who endure great costs or make admirable efforts to help or protect other people. It is just to say that it involves praising agents for the way in which, in their thought and actions, they control what is in their power.)

The virtue of rationality, it seems, is an agential virtue in just this way. The rational dispositions are precisely dispositions of agents. Moreover, it is plausible that whenever an agent thinks either in a rational way or in an irrational way, it is “up to” the agent how exactly she thinks on that occasion, in the sense that it is within her power both to think in some ways that are abstractly rational, and also to think in other ways as well. It is plausible that it is for this reason that the manifestations of rational dispositions are praiseworthy in this distinctively agential way, while the manifestations of irrational dispositions are worthy of a corresponding kind of criticism.

Among the agential virtues, rationality has certain other distinguishing features. It will not be crucial for the rest of the argument of this essay exactly what the distinguishing features are, but it may be useful, in order to fix ideas, to give a quick sketch of how one could give an account of these distinguishing features.

First of all, the virtue of rationality is manifested only in what in the broadest sense counts as reasoning; in this way, rationality differs from many other virtues, which are also manifested in one’s feelings and actions (as well as in one’s reasoning). We should understand ‘reasoning’ broadly, so that it encompasses not just events of forming and revising one’s attitudes (like one’s beliefs and intentions), but also the way in which one holds and maintains such attitudes at a particular time. This makes it possible for us to distinguish between diachronic rationality – the rationality of changes to one’s system of attitudes over time – and synchronic rationality – the rationality of the system of attitudes that one holds at a particular time.

It might seem strange that I have suggested that rationality is manifested only in reasoning. Surely we can call actions “rational” as well? It is undeniable that in a wider sense, an action that is the execution of a rational intention can be called a “rational action”. But – or so it seems to me – the agent would not be any less rational if she were suddenly paralysed at the last moment, and so could not execute her rational intention. So it seems plausible that rationality does indeed have a special connection to what in this broad sense counts as reasoning.

In other words, in Aristotle’s scheme, rationality is a virtue of thought (ἀρετὴ διανοητική) like theoretical wisdom (σοφία) and practical prudence (φρόνησις): its manifestations consist purely in reasoning of this kind; unlike the manifestations of a virtue of character (ἀρετὴ ἠθική), they do not consist in feelings and actions. In this way, the notion of ‘rationality’ is used primarily to evaluate reasoning – or in other words, to evaluate mental states (like beliefs or intentions) and mental events (like events of judgment or choice or decision, in which we form or revise our beliefs or intentions or other mental states in various ways).

The second distinguishing feature of the virtue of rationality that I shall sketch here (merely, as I said, in order to fix ideas), is that it is in a sense an “internal” virtue of mental states and mental

---

13 For Aristotle’s distinction between virtues of thought and virtues of character, see Nicomachean Ethics (1103a2–11).
events. That is, when we evaluate a mental state or mental event as rational or irrational, we are evaluating it, not on the basis of its relations to the external world, but purely on the basis of its relations to the mental states and events that are present in the thinker’s mind at (or shortly before) the relevant time. That is, the doctrine that epistemologists call “internalism” (and more specifically, the “mentalist” form of internalism) seems to be true of rationality.\footnote{For a discussion of this sort of internalism, see Wedgwood (2002).}

Often, of course, we evaluate mental states and events at least partly on the basis of their relations to the external world. For example, when we evaluate a mental state like a belief as correct, or as having got things right, or when we say that someone has made the right choice, we are typically evaluating that mental state or event on the basis of how it relates to the external world. But according to the proposal that I am sketching here, judging whether or not a mental state or event is rational is a different kind of evaluation – one that focuses purely on how that state or event relates to what is present in the thinker’s mind at the relevant times.

Finally, the third feature of the virtue of rationality that I shall sketch here (just to fix ideas) is that the requirements of rationality applying to a given type of mental state are in a sense constitutive of the nature of that type of state. There are two dimensions along which mental events and states can be classified into types: (a) according to their content (which we may assume here to depend in part on the concepts that figure in that mental state); and (b) according to the type of attitude that these mental states exemplify – where examples of such attitude-types include judgment, belief, choice, intention, and the like. According to the account that I am sketching here, the norms of rationality are constitutive of types of mental state along both dimensions. The requirements of rationality that apply to belief are part of what makes belief into the distinctive kind of mental state that it is; the requirements of rationality that apply to the use of the concept ‘if’ are part of what makes it the particular concept that it is.

This final feature may be what distinguishes the ways of evaluating mental states that appeal to what is sometimes called the “wrong kind of reason” from those that appeal to the “right kind of reason” instead.\footnote{For a discussion of the distinction between the “right kind of reason” and the “wrong kind of reason”, see Schroeder (2012).} A belief might lead to the believer’s being rewarded by an eccentric and powerful being (like a deity or a billionaire) who wants the believer to have that belief; and this would make it intelligible to use the word ‘good’ in such a way that beliefs that lead to the believer’s being rewarded in this way count as “good beliefs to have”. But this way of evaluating beliefs, unlike evaluations of whether or not a belief counts as “rational”, does not track anything constitutive of the very nature of belief. So this third feature could also be a further feature that distinguishes rationality from other virtues that can be exemplified by mental states and events.

In fact, the arguments that will be developed later in this essay will not depend on the account that I have sketched in this section. This sketch has been included here just to fix ideas – and to make it plausible that an account that interprets rationality as a virtue can explain how rationality differs from the other virtues that agents can have.

V. Degrees of virtue, requirements of rationality

Intuitively, it seems clear that virtue and vice come in degrees. Some agents are, on account of their dispositions, more virtuous – and less vicious – than others, while the latter agents are less
virtuous – and more vicious – than the former. In this way, dispositions can be compared as more or less virtuous. Comparisons can also be made between agents with respect to particular virtues: one agent can be wiser, or more just, or more courageous, than another, and so on.

Moreover, it is not only agents who can be compared with respect to the degrees to which they exemplify these virtues. Particular acts or mental states or events can also be compared as more or less virtuous, on the basis of the degree of virtue that is exemplified by the dispositions that these acts or mental states or events manifest. In this sense, one act, or one belief or one decision, can be wiser or more courageous than another.

Similarly, it seems, particular acts and mental states and events can be compared, not only with respect to the degree of virtue that is exemplified by the disposition that they manifest. They can also be compared directly with respect to abstract virtue: some such items can be less abstractly virtuous than others. Even if it is a sheer fluke that the agent did an act that is abstractly wise or just to the degree that it is, the act could be wiser or more just than some other acts that were available. So each of the three kinds of goodness that is involved in any virtue allows for comparisons of degrees of goodness – comparisons that we could express by means of comparative terms like ‘better’ and ‘worse’ and the like.

This point seems to apply to rationality just as much to the other virtues. Rationality also comes in degrees. Some pieces of reasoning – that is, some mental states or mental events – are more irrational than others. Again, we can discern a scale of lesser and greater degrees of irrationality with respect to each of the three kinds of rationality that we distinguished above: some reasoning dispositions can be more irrational than others; some particular mental states or events can manifest dispositions that are more irrational than those manifested by other mental states or events; and some mental states or events can themselves be more abstractly irrational than others.

It will clearly be a challenging matter to develop an illuminating account of what makes one piece of reasoning more irrational than another. But it seems undeniable that irrationality comes in degrees in this way. Indeed, it may be that the comparative notion of being less rational than is more fundamental than the non-comparative notion of being rational simpliciter. Perhaps, for a mental state or event to be rational simpliciter is just for it to be maximally rational – that is, to be no less rational than any available alternative.

With respect to many of these different kinds of virtue, we are often especially interested in comparing the degrees of virtue that are exemplified by the different acts or mental events or states that are available to an agent in the particular situation that the agent is in at a particular time. In many situations, some of the available ways of responding to the situation count as more vicious than other such ways; and in some of these situations, some of these available ways of responding will count as minimally abstractly unjust (that is, maximally abstractly virtuous); in other words, these responses are no more abstractly vicious (that is, no less abstractly virtuous) than any available alternative.

For example, some of the available ways of responding to one’s situation might be minimally abstractly unjust. In this case, we can say that justice requires that one should respond in one of these minimally unjust available ways. According to this interpretation, the term ‘required’ in this context means simply needed or necessary. What is required by justice is precisely what is necessary

---

16 I have attempted to sketch such an account in some earlier work; see Wedgwood (2013b).
in order to be as just as possible – that is, what holds in all the available worlds in which one responds in a way that is no more unjust than the way in which one responds in any other available world.\(^{17}\)

What does it mean to say that some way of responding is “available” to a particular agent at a particular time (or that a world in which the agent responds in that way is available to the agent at the relevant time)? It seems, in fact, that this talk of “availability” is context-sensitive. In some contexts, the term ‘available’ is used in a more inclusive way, so that many possible responses count as “available”, while in other contexts, it is used in a more restricted way, so that fewer possible responses count as “available”.\(^{18}\)

When ‘available’ is used in the more inclusive way, the corresponding requirements of justice can be more idealized – since the responses that count as minimally unjust in these contexts may not even count as “available” when the term is used in a more restricted way.\(^ {19}\) When ‘available’ is used in a more restricted way, the corresponding requirements of justice are more “realistic”, and less idealized, since the responses that count as minimally unjust in these contexts may count as more unjust than some alternatives that are viewed as “available” when the term is used in a more inclusive way.

Given that we have a notion of degrees of irrationality, and a notion of the responses that are “available”, we can understand the requirements of rationality in the same way as the requirements of justice:

In each situation, what rationality requires is that one should respond in one of the minimally irrational ways that are available in that situation.

Arguably, this notion of a “rational requirement” is a kind of ‘ought’. To say that you are “rationally required” to respond in a certain way is to say that you in a certain sense “ought” to respond in that way.\(^ {20}\)

The notion of degrees of irrationality also allows us to make sense of a corresponding way of grading the requirements of rationality. In effect, rational requirements can be graded according to how gravely or egregiously irrational it is to violate those requirements. The more egregiously irrational it is to violate a requirement, the more “basic” the requirement is; the less egregiously irrational it is to violate a requirement, the less “basic” the requirement is.

For example, it clearly seems to be a basic requirement of rationality that we should not have very high levels of confidence in propositions that are obviously logically absurd; to violate this requirement is, it seems, quite egregiously irrational. By contrast, it seems to be a less basic

---

\(^{17}\) Cf. John Broome’s (2013, 111–16) distinction between two ways of understanding “requirements” as “property requirements” and as “source requirements”.

\(^{18}\) For the general idea that modal terms like “available” are context-sensitive in this way, see Angelika Kratzer (2012, chap. 2).

\(^{19}\) So the kind of “idealization” that is in question here is not the same as that found in the “ideal gas laws”. It is a genuinely normative kind of idealization – more akin to the kind of “ideal theory” that is found in political philosophy. See e.g. Rawls (1971, 245–6).

\(^{20}\) This way of understanding terms like ‘ought’ has its roots in the work of A. N. Prior (1958); for similar approaches, see David Lewis (1973) and Kratzer (2012). So long as there are some possible worlds where the relevant agent responds to her situation in a minimally vicious or irrational way, this way of understanding terms like ‘rationally required’ and ‘ought’ implies that all the principles of standard deontic logic hold of these terms.
requirement of rationality that we should have no level of confidence that falls short of the maximum possible level in any logical truth, since even if \( p \) is a logical truth, having a level of confidence in \( p \) that falls short of the maximum possible level need not be quite as egregiously irrational as being virtually certain of a contradiction.

This distinction between more and less basic requirements of rationality could play an important role in many different kinds of theory. For example, it may be plausible that it is constitutive of being capable of certain types of mental states that one has a disposition to avoid violating certain particularly basic requirements of rationality that apply to those mental states; but it may be less plausible to suggest that in order to be capable of those types of mental states, one must also have dispositions to avoid violating the less basic requirements of rationality.

At all events, this conception of the different ways of understanding the notion of the “requirements of rationality” may help to explain the sense in which it might be true to claim that rationality requires probabilistic coherence in one’s beliefs, or making choices that maximize some sort of expected value.

In the contexts in which such claims are made, I propose, the relevant notion of “availability” is understood in the most inclusive and least restricted way – so that the bare metaphysical or logical possibility of your responding in a certain way is taken as sufficient for that way of responding to count as “available”. It may be fantastically difficult to have a set of beliefs that is probabilistically coherent, but it is not logically or even metaphysically impossible; so in the relevant contexts, worlds where one has a probabilistically coherent system of beliefs count as “available”. If the worlds where one’s reasoning is no more irrational than any available alternative are included among the worlds in which one’s beliefs are coherent in this way, then this explains why rationality requires probabilistic coherence.

As I explained above, contexts where the relevant notion of “availability” is this extremely inclusive and unrestricted notion are contexts that involve the most idealized notion of the requirements of (abstract) rationality. In general, then, I propose, the claims of formal epistemology and decision theory are highly idealized claims about abstract rationality. As we might put it, they are claims about perfect abstract rationality. Such idealization treats every metaphysically (or even barely logically) possible response to one’s cognitive situation as available, and to be perfectly rational is to exhibit one of the least irrational of this immense range of possible responses.

VI. Idealization and rational dispositions

The proposals of the previous section go a long way towards solving the problem that I raised at the outset – the problem that formal theories of rationality typically claim that rationality requires one’s overall set of mental states to have certain formal features (like probabilistic coherence and the like), even though it is not obviously even possible for ordinary agents to have mental states with these formal features. The previous section offered an interpretation of what it means to say that “rationality requires” something; and this interpretation provided a principled way of seeing how in some contexts the kind of “possibility” that must be exemplified by everything that can truly be said to be “required by rationality” may be a very weak kind of possibility indeed.

However, there are some complications that have not yet been adequately addressed. As I explained in Section III, I am assuming that abstract rationality and rational dispositions are related by something like the two connections (i) and (ii) that I have specified. The connection that is
particularly relevant here is the first connection (i). As I explained, this connection implies that abstract rationality is constrained by the availability of ways of responding to one’s situation that would consist of the manifestations of rational dispositions. That is, roughly, it cannot be the case that it is abstractly rational for you to $\varphi$ at a time $t$ unless there is available to you a way of responding to your situation at the relevant time, consisting of the manifestations of rational dispositions, which would result in your $\varphi$-ing at $t$.

Presumably, this use of the term ‘availability’ can also be understand in a maximally inclusive and unrestricted sense. When the term is understood in this sense, a way of responding to your situation counts as “available” if it is barely metaphysically possible for you to respond to your situation in this way.

Undoubtedly, our actual dispositions are imperfect – even though they are dispositions to be guided by some of the genuine requirements of rationality (and especially by the most basic requirements). However, it is at least metaphysically possible for our rational dispositions to be less imperfect than they are. We can conceive of a series of possible dispositions of this kind, where each of these dispositions is less imperfect than its predecessor. The worlds in which your dispositions are less imperfect than they actually are, and in which you manifest these less imperfect dispositions, can in a sense be regarded as “available”. So it may seem possible to reconcile even a highly idealized conception of the requirements of rationality with the idea that what abstract rationality requires of us is constrained by what we attain through available manifestations of rational dispositions.

This inclusive interpretation of “availability” does not trivialize the idea that the requirements of abstract rationality are constrained in this way. Arguably, however much our dispositions were to improve, it would still not be metaphysically possible for flesh-and-blood creatures like us to have infinite minds. If that is right, then abstract rationality cannot require that we have a system of degrees of belief that involves attitudes towards infinitely many propositions. So, we cannot be rationally required to have degrees of belief that constitute a complete “subjective probability” function. At most, we are required not to have a system of degrees of belief that cannot be extended into such a complete “subjective probability” function. In that sense, we may be rationally required to have probabilistically coherent beliefs, but not to have a complete infinite system of beliefs.

Thus, it seems possible for a perfectly rational thinker to have a system of beliefs that is “gappy”, in the sense that there are some propositions towards which the thinker has no attitudes at all, even though she is quite capable of having attitudes towards these propositions. For such thinkers, it seems plausible that there is kind of diachronically rational thinking which consists in filling in these gaps, through a process of inference. Presumably, such rational processes of inference will involve the thinkers’ responding to their considering a proposition $p$ that they had previously never had any attitudes towards, by forming a new system of beliefs that respects all the constraints of coherence and includes an attitude towards that proposition $p$.

Such processes of inference raise further questions about whether it is metaphysically possible for a thinker to conform to what many theorists have regarded as requirements of rationality. The problem is that such diachronically rational processes of inference will involve one event (one’s

\[\text{For the idea of interpreting probabilism as a requirement of “coherent extendibility” in this way, see especially Joyce (1999, 97–104).}\]

\[\text{For more on this conception of inference, see Wedgwood (2012).}\]
considering the relevant proposition) causing another event (one’s forming an appropriate attitude towards the proposition), and there seem to be certain metaphysically necessary limits to all such causal processes, at least in the minds of flesh-and-blood creatures like us. If a process whereby one event causes another event is the manifestation of a rational disposition, the two events are presumably related to the disposition as *stimulus* and *response*. But it seems plausible that – at least in flesh-and-blood creatures like us – whenever a disposition of this sort is manifested, there must always be some *time lag* between the stimulus and the response.

Let us take a simple example, in which the proposition *p* that you consider now for the very first time is a logical truth. In this case, there will inevitably be a time lag between your considering *p* and your coming to believe *p* with the degree of confidence that is required by probabilistic coherence – namely, maximal confidence. During this time lag, it may seem that you will inevitably have some level of confidence towards *p* (perhaps a vague or indeterminate level of confidence) – and this level of confidence will presumably be different from the attitude of maximal confidence that is required by probabilistic coherence. So, during this time lag, your system of beliefs or levels of confidence will be probabilistically incoherent. Thus, it seems inevitable that if you ever consider a logical truth that you had never considered before, your beliefs will not be probabilistically coherent at all times. Should we conclude that rationality does not require probabilistic coherence?

Here is a solution to this problem that seems to fit with the other proposals that we have offered. Rationality requires probabilistic coherence among one’s *settled* beliefs, of the kind that form part of one’s enduring system of beliefs, but it does not require that *provisional* attitudes, of the kind that one has towards this logical truth *p* during the time lag between considering *p* and coming to have an attitude of maximum confidence in *p*, must be probabilistically coherent. At most, rationality requires that one should form this attitude of maximum confidence towards *p* as quickly as is metaphysically possible after considering *p*.

So far, we have been identifying certain limits on the requirements of rationality that flow from the principle that these requirements are constrained by what can be attained through possible manifestations of rational dispositions. In the remainder of this section, I shall explore a further argument for thinking that the requirements of rationality are limited in this way – even though in the end, I shall suggest that there is a way of resisting this argument.

This argument for thinking that there are yet further limits on the requirements of rationality is inspired by a famous argument that is due to Timothy Williamson (2000, Chapter 4). Specifically, the argument starts out from the observation that all manifestations of rational dispositions involve some causal connection between mental states or mental events, and it seems plausible that all such causal connections depend on factors that can vary *continuously* from one case to another.

Specifically, it seems that our dispositions to believe propositions depend on causal factors that can vary continuously in this way. There is some respect of similarity such that, for every degree to which things can be similar in that respect, there is a *series* or *spectrum* of possible stimuli in which each stimulus is similar to the immediately preceding and succeeding stimuli to at least that degree. On this spectrum, the way in which our dispositions respond to each stimulus can differ at most very *slightly* from the way in which they respond to the immediately preceding and succeeding stimuli. Thus, the way in which our dispositions to have beliefs respond in each of these cases can
differ at most very slightly from the way in which they respond in the immediately preceding and succeeding cases – even if the proposition believed differs in truth value between the two cases.\(^2\)

This point about our dispositions to have beliefs looks likely to make trouble for some views about the requirements of rationality. For example, consider the view that rationality requires correctly introspecting the truth of any proposition about one’s conscious mental states that one actually considers. Specifically, for any conscious mental state \(M\), take the first-person present-tensed proposition that you could express by saying something like ‘I am in mental state \(M\)’ (using a canonical psychological description ‘\(M\)’ to pick out this mental state \(M\) ); let us call this proposition the “\(M\)-self-ascribing proposition”. Now suppose that rationality requires that you must meet the following two conditions: first, whenever you are in mental state \(M\), if you have any attitudes towards the \(M\)-self-ascribing proposition at all, you must believe it; and secondly, whenever you are not in that mental state \(M\), you must not have any high degree of confidence in this \(M\)-self-ascribing proposition.

An analogue of Williamson’s argument seems to show that however much your dispositions improve, there always will be some instances of these requirements that your dispositions cannot lead you to conform to. Consider a series of cases where each case differs at most very slightly from the immediately preceding and succeeding cases, in terms of the stimuli that trigger our dispositions to have beliefs, and where in at least some of these cases, the proposition that the thinker is considering is true, and in others of these cases, the proposition that the thinker is considering is false. On this series, there will be at least one pair of adjacent cases where the proposition considered in the first case is true and the proposition considered in the second case is false.

Suppose that in each of these two cases, the thinker does have some attitudes towards the relevant \(M\)-self-ascribing proposition. (If the thinker has no attitudes whatever towards this proposition, then although the thinker may vacuously conform to the rational requirement, it can hardly be said to be the thinker’s rational dispositions that lead the thinker to conform to the requirement.) If the thinker has attitudes towards this proposition, there are two possibilities. First, it might be that in the first case the thinker does not believe the \(M\)-self-ascribing proposition; then the thinker fails to conform to the first condition of this rational requirement, because she has attitudes towards this true proposition but does not believe it.

Alternatively, it might be that the thinker does believe the \(M\)-self-ascribing proposition in the first case. But then the thinker will be at most fractionally less confident of the relevant \(M\)-self-ascribing proposition in the second case too (because the thinker’s dispositions can only respond to the two cases in at most very slightly different ways). So in the second case, the thinker fails to conform to this second condition of the rational requirement – because she has a high level of confidence in an \(M\)-self-ascribing proposition that is false. So, either way, if the thinker has any attitudes towards the propositions in question, the thinker’s dispositions cannot lead to her conform to this requirement.

In short, it seems that we could not possibly have rational dispositions that would lead us to have markedly different levels of belief in two adjacent cases in any series of cases of this sort – even if

\(^2\) This point is sometimes denied. In principle, why couldn’t there be some mental dispositions that respond to some “tipping point” with hair-trigger accuracy? In my opinion, it is unrealistic to hope that such super-sensitive dispositions are even metaphysically possible in the case of all the requirements of rationality that I shall discuss below. Unfortunately, however, I cannot defend this opinion here; so I shall simply have to assume that it is correct for the sake of the present argument.
the proposition in question in the first case is true, while the proposition in question in the second case is false.

One might be tempted to say that this simply shows that rationality cannot require us to have markedly different levels of belief in any two adjacent cases of this kind. But in fact there are some powerful reasons for thinking that abstract rationality does require markedly different levels of belief in some cases of this sort. For example, consider first-person present-tensed propositions about what rationality requires of you (that is, propositions that one might express by saying something of the form 'Rationality requires me now to φ'). Suppose that it were possible that, even though rationality in fact requires you to φ, you rationally believe the false proposition that rationality requires you not to φ. Then you would be caught in a "rational dilemma" – a case in which you are bound to respond irrationally, no matter how you respond: you either comply with your rational belief about what rationality requires, by refraining from φ-ing, in which case you violate the rational requirement to φ; or else you comply with the rational requirement to φ, in which case you are guilty of a kind of akrasia, by responding in a way that you rationally believe to be irrational. So it seems that it cannot be rational for you to believe a false proposition about what rationality requires of you in this way.²⁴

Moreover, it seems that this argument can be generalized so that it implies that it cannot be rational to place any degree of confidence at all in a false proposition about what rationality requires of you; for if you placed any degree of confidence in the false proposition that φ-ing is not just irrational, but terribly and atrociously irrational, then so long as you were also convinced that if φ-ing is in fact what rationality requires, not φ-ing is at worst only very slightly irrational, it could still be in a sense akratic for you to φ – in the sense that the response that is supported by your beliefs about what rationality requires of you would involve not φ-ing. So any degree of confidence in a false proposition of this sort seems to be enough to give rise to a "rational dilemma".

Furthermore, at least if your beliefs are probabilistically coherent, any degree of doubt in a true proposition about what rationality requires of you will involve attaching at least some degree of confidence to a false proposition. Thus, if rationality requires probabilistic coherence, and also forbids you to place any degree of confidence in a false proposition about what rationality requires, it will also forbid any degree of doubt in the truth about what rationality requires. In short, rationality seems to require that if you have any attitude at all towards propositions about what rationality requires of you, you must have complete confidence in the true propositions, and complete disbelief in the false propositions.

There also seem to be other examples of this phenomenon. For example, consider a set of propositions involving only the essential concepts of a decidable mathematical theory, such that each of these propositions is either provably true or provably false. Any probabilistic theory will imply that rationality requires that for every one of these propositions p, if you have any attitudes towards p at all, you must have complete confidence in p if p is true, and complete disbelief in p if p is false. But again, it may be possible to construct a spectrum of cases such that in terms of what triggers your dispositions to believe, each case is extremely similar to the immediately preceding and succeeding cases, but in some cases the proposition in question is true, while in some other cases the proposition in question is false. So yet again, it seems that you could not have any rational dispositions that would lead you in every case to conform to this requirement of rationality.

²⁴ This argument is due to Michael Titelbaum (forthcoming).
In this section, I have been exploring the assumption that what abstract rationality requires of us is constrained by the possibility of our responding to our situation in an abstractly rational way by manifesting our rational dispositions. The Williamson-inspired considerations that I have just surveyed may seem to suggest that, given some of the necessary features of our belief-forming dispositions, this assumption imposes some severe restrictions on what abstract rationality can require of us. In the remainder of the paper, I shall argue that there is a way of interpreting this assumption that will prevent it from imposing such severe restrictions.

For the sake of argument, let us accept that it is indeed impossible for any thinkers to have the kind of super-sensitive dispositions that could in every case lead them to conform to the requirements of rationality that we have been considering. Still, it may be possible for us somehow to weaken the assumption that the requirements of abstract rationality are constrained by the possibility of attaining the rationally required response through manifesting one’s rational dispositions.

Our original formulation of the assumption was this: Whenever abstract rationality requires you to φ, there is a possible way for you to respond to your situation, consisting of the manifestation of rational dispositions, that will result in your φ-ing. The trouble with this formulation was that our rational dispositions seem only capable of leading us to conform to requirements of rationality in a proper subset of the full range of cases to which these requirements apply. However, we can make sense of measuring how closely this proper subset of the range of cases approximates to being the full range. Then we could reformulate the constraint as follows: For every set of rational requirements, which apply to a range of possible cases \( R_1 \), and for every degree of closeness to the full range of cases, there is a more limited range of cases \( R_2 \) that approximates to the full range to at least that degree, such that one could have a disposition to conform to the requirements in each case in that range \( R_2 \).

In other words, even if no finite agents can have dispositions that reliably lead them to conform to the requirements of rationality in every case, they can at least have dispositions that take them arbitrarily close to conforming to those requirements in every case. If we imagine your dispositions becoming more and more discriminating, these dispositions could get within any degree of approximation to conforming in every case – even if they could never achieve such perfect conformity across all possible cases. In a sense, we could think of perfect abstract rationality as the limit towards which such a sequence of improvements in your dispositions converges.

Thus, for example, even if it is not metaphysically possible for our dispositions to believe \( M \)-self-ascribing propositions in response to our cognitive situation to be perfectly sensitive – so that we always believe these proposition when they are true and disbelieve them when they are false – our dispositions could, in principle, become ever more and more sensitive – so that the range of cases in which we fail to conform to this requirement becomes ever narrower as our dispositions improve.

In fact, the history of ethics contains an antecedent for this conception of ideal rationality. In the Critique of Practical Reason, Kant argues that moral perfection is an ideal that we could only ever approximate to – although if we had endless opportunity for improvement, we could get ever closer and closer to that ideal.25 In effect, this is how Kant reconciles an uncompromisingly idealized conception of moral perfection with the idea that all rational beings have dispositions that can lead them to conform to at least some of the requirements of morality.

---

The same approach seems to be the most promising way to reconcile an idealized conception of rationality (such as a conception that requires probabilistic coherence or the like) with the idea that rationality has an essential link to the dispositions of actual rational agents.

VII. Conclusion

I have proposed here that the notion of “rationality”, as it is used by formal epistemologists and decision theorists and the like, is the notion of a certain kind of virtue. Conceiving of rationality as a virtue in this way helps us to understand several important features of this notion.

First, it allows us to understand the connections between what I have called abstract rationality and the manifestations of rational dispositions – a distinction of which the epistemologists’ distinction between propositional and doxastic justification is just one instance.

Secondly, it allows us to see how rationality comes in degrees, with some agents – and some acts and attitudes – being less rational than others, and the connection between this notion of degrees of rationality and the context-sensitive language of “requirements of rationality”.

Finally, it helps us to answer one of the objections that I mentioned at the beginning – the worry that if rationality is a normative notion, then the precise theories of rationality that have been developed by formal epistemologists and decision theorists all collide with the principle that ‘ought’ implies ‘can’. In some contexts, the kind of ‘can’ (or ‘available’) that is implied by ‘ought’ is a very weak kind of ‘can’ (or ‘available’); and the connections between rationality and the kind of ‘ought’ that is expressed by the language of “requirements of rationality” is context-sensitive in a precisely corresponding way. Rationality does indeed have a fundamental link to dispositions that ordinary agents have; but it is still a lofty ideal, to which we can aspire, and approximate, but which we can never expect fully to achieve.26

References


26 Earlier versions of this essay were presented at a conference at Cornell in honour of Gail Fine and Terence Irwin, at the Analytic Philosophy workshop at the University of Texas at Austin, and at the University of Nebraska. I am grateful to all those audiences for extremely helpful comments.


