0. Introduction

0.1 Rationality: A trilogy

This book, *The Normativity of Rationality*, is designed to be the first instalment of a trilogy. There are two planned sequels: a second instalment, *The Norms of Rational Belief*, and a third and final instalment, *The Norms of Rational Choice*.

It is a fundamental assumption of the whole trilogy that there is a single central concept, often expressed among philosophers by the term ‘rationality’, which can be applied both in the theoretical domain, to beliefs and judgments, and in the practical domain, to intentions and choices. This concept appears in many of the classic debates among epistemologists and philosophers of science (in the theoretical domain), and among decision theorists and philosophers who study ethics and reasons for action (in the practical domain). Although the English term ‘rational’ can undoubtedly be used in many ways, in this book and its sequels it is always used to express this central concept. This first book sets out to study this concept in general, abstracting away as much as possible from all the special questions that specifically concern the rationality of beliefs or the rationality of choices. These special questions will be taken up in the two sequels to this book: the rationality of beliefs will be explored in the second book, and the rationality of choices in the third book.

What sort of concept is it that we are concerned with here? What concept does the term ‘rationality’ express, when it is used in this way? One of the main goals of this book is to argue that it is a normative concept. But what does this mean? Roughly, it means that to think rationally is to think properly, or to think as one should think, in certain distinctive senses of ‘properly’ and ‘should’. Rationality is a kind of virtue displayed in some of the mental states (like the beliefs and intentions) that agents have, and in the ways in which agents form and revise those mental states in response to reflection and experience. To say that “it is rational” for a person to think in a certain way implies that it is in a certain sense permissible for the person to think in that way; and to say that a person “is rationally required” to think in a certain way implies that in a certain sense the person ought to think in that way.

It is probably the majority opinion among those who theorize with the concept of rationality that it is a normative concept in something like this sense. But recently, philosophers have mounted a series of objections against this opinion, arguing that rationality is not plausibly interpreted as a normative concept. One of the main tasks that I shall attempt in this book is to explain in more detail what it is for something to be a
normative concept, and on the basis of this explanation, to defend the claim that the notion
of rationality is indeed a normative concept. In so doing, I shall answer the philosophers
who have raised objections against the opinion that rationality is a normative concept.

As I shall also argue, however, there are in fact many different normative concepts. So a
further task is to explain how the notion of rationality differs from, and is related to, these
other normative concepts. As I shall argue at the end of this book, in Chapters 8 and 9, once
we understand what sort of normative concept the concept of rationality is, we will also
discover a general framework that must be exemplified by any account of what it is for a
mental state or a mental event of a given kind to count as rational. This framework will
guide us in our investigations of rational belief (in the second instalment of this trilogy)
and of rational choice (in the third instalment). In this way, this book sets the agenda for its
two planned sequels.

0.2 The distinguishing features of rationality

As I have revealed, I shall argue that there are in fact many different normative concepts,
of many different kinds. According to the position that I shall defend in this book, the
concept of rationality differs from all these normative concepts in three crucial ways.

First, the relevant concept of ‘rationality’ is used primarily to evaluate mental states (like
beliefs or intentions, and collections of beliefs and 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). Derivatively, the concept may also be used to evaluate
actions, given the assumption that a rational action is the execution of a rational plan or
intention. But the primary use of this concept is to evaluate mental states and mental
events of these kinds.

Secondly, 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) is true of rationality. Often, of
course, we evaluate mental states and events at last 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

\footnote{For a useful discussion of the epistemological debate between Internalism and externalism, and the
distinction between mentalist and accessibilist forms of internalism, see Conee and Feldman (2001).}
typically evaluating that mental state or mental event on the basis of how it relates to the external world. But assessing a mental state or event as rational or as irrational is a different kind of evaluation – an evaluation that focuses purely on how that state or event relates to what is going on in the thinker’s mind at the relevant time.

Thirdly, 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 I shall assume in this work to depend 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. As I shall argue, the requirements of rationality are constitutive of types of mental state along both dimensions. 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; the requirements of rationality that apply to belief are part of what makes belief into the distinctive kind of attitude that it is.

This third feature is, as I shall argue, what distinguishes between ways of evaluating mental states that respond to what are sometimes called the “wrong kind of reason” and those that respond to the “right kind of reason” instead. 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 it would be intelligible to use the word ‘good’ in such a way that beliefs that lead to the believer’s being so rewarded can be called “good beliefs to have”. But this way of evaluating beliefs does not track anything constitutive of the very nature of belief like evaluations of whether or not the belief counts as “rational”. So this third feature of rationality also helps to distinguish rationality from other norms that apply to mental states and events.

---

2 I have argued that rationality is a “constitutive norm” in some of my earlier work (Wedgwood 2007, chap. 7). Unlike some others – such as Christine Korsgaard (1997) – who also defend the thesis that rationality is a constitutive norm of this sort, I am not appealing to this sort of “constitutivism” to answer any sort of “normative question”; I am simply proposing that it is a distinguishing feature of rationality, which distinguishes rationality from some other normative standards.

3 The question of how to distinguish between the “right kind of reason” and the “wrong kind of reason” has been much discussed; for an example of an illuminating recent discussion, see Schroeder (2010).
0.3 A problem about rationality

As we shall see, these claims about the concept of “rationality” create a problem. According to internalism, the way in which it is rational for one to think at a given time is determined purely by the facts about the mental states and events that are present in one’s mind at (or shortly before) that time. So the ways in which it is rational for one to think are the ways of thinking that “fit” appropriately with these facts about the mental states and events that are present in one’s mind at the relevant times. In a sense, then, internalism implies that rationality is just a matter of coherence: it is just a matter of having mental states and events that “fit” together in the appropriate way. In making this claim, I am using the term ‘coherence’ in a very general sense: for example, in this general sense, we can make sense of relations of coherence between one’s beliefs and one’s experiences and one’s memories, and between the beliefs that one forms now and the beliefs that one held in the immediate past, and not just between one’s current beliefs. At least in this broad sense of the term, it seems that if internalism is true, rationality is just a matter of coherence.

As I have also claimed, the norms of rationality are in a way constitutive of the various different types of mental state. According to the version of this claim that I shall advocate, if one is even to be interpretable as having mental states of the relevant types at all, one must have at least some disposition to conform to those norms. Moreover, these norms apply, not just to a few special kinds of mental states and mental events, but to absolutely all beliefs and intentions, and to all processes of forming or revising such beliefs and intentions.

In this way, the norms of rationality are both pervasive – they apply to all reasoning and all beliefs and intentions whatsoever – and have an inescapable grip on us – we necessarily already have a disposition to conform to them. However, it is puzzling how there can be any norms of this sort that evaluate mental states and mental events purely on the basis of coherence. Why does coherence matter? Surely it is nothing more than just a pretty pattern of mental states. Why it is a matter of any importance whatsoever whether one’s mental states form this pretty pattern or not?

This is, in fact, one of the main problems that has been raised for the interpretation of ‘rationality’ as a normative concept. It is essentially this problem that has been raised
against the view that any important normative notion requires pure coherence by Niko Kolodny (2005). One of the main goals of this book is to solve this problem.

0.4 The solution

My solution to this problem will not be presented in full until the whole trilogy is complete. In a nutshell, however, the core of the solution that I propose can be described, metaphorically, by saying that every type of mental state or mental event that can be rational or irrational has an aim, and thinking rationally is a means to achieving that aim.

Thus, to understand what rationality requires of any particular type of mental state, we need to answer two basic questions. The first question is: Under what conditions do mental states of this type achieve what I have metaphorically referred to as their “aim”? The second question is: What literal sense can be attached to this metaphorical description of rational thinking as a “means” to achieving an aim? Answering the first of these two basic questions will be one of the main tasks of the second and third instalments of this trilogy. In particular, I shall defend versions of some traditional views about both belief and choice: in the second instalment, I shall argue that belief aims at the truth; and in the third instalment, I shall argue that choice aims at the practicable good. As for the second of these basic questions, an initial sketch of an answer will be given at the end of this book; but the answer will not be worked out in full until the end of the second book.

Roughly, I shall argue that this metaphorical talk of “means” for achieving an “aim” is to be understood in irreducibly normative and probabilistic terms. This idea of the “aim” of a type of mental state is to be understood as the idea of a special normative concept that applies to mental states of that type. I shall express this concept by using the term ‘correct’. A correct mental state is a mental state that has got things right, while a mental state that is not correct is incorrect, or mistaken – a mental state that has got things wrong.

The norms of correctness, like the norms of rationality, are constitutive norms: they are part of what makes each type of mental state the specific state that it is. Unlike the norms of rationality, however, the norms of correctness are external: they typically evaluate mental states and events, not just on the basis of their relations to what is going on in the

---

4 Kolodny’s argument has some kinship with Joseph Raz’s (2011: Chap. 8) argument that no important normative notion requires what Michael Bratman (1987: 31) calls “means-end coherence” – although as we shall see in Section 0.6 below, Raz himself does not deny the normativity of rationality.

5 I attempted an initial statement of this idea of the “aim of belief” in Wedgwood (2002). This book and its sequel give what now seems to me a greatly improved statement of this idea.
thinker’s mind, but on the basis of these mental states’ relations to the external world. There are norms of correctness, not just for individual mental states, but also for collections of mental states, and for mental events in which one forms and revises one’s mental states. To have a general label, I shall call all the items that can be in this way evaluated as correct or incorrect ways of thinking.

Among the ways of thinking that are not perfectly correct, some are significantly more seriously or badly incorrect than others. In this sense, incorrectness comes in degrees. We can represent these degrees of incorrectness by giving a score to every way of thinking, according to how badly or seriously incorrect it is, according to a certain scoring rule; this scoring rule would be in effect a way of measuring these degrees of incorrectness. Moreover, just as incorrectness comes in degrees, so too does irrationality. That is, some ways of thinking are more irrational – that is, more seriously or badly irrational – than others. Just as there is, in principle, a way of measuring these degrees of incorrectness, so too there is also, at least in principle, a way of measuring these degrees of irrationality.

Now, suppose that somehow the mental states and events present in the thinker’s mind at the relevant time determine a certain probability function – in effect, the probability function that rationally should be guiding the thinker at that time. There could be many different views about what this probability function is like: some philosophers might follow Timothy Williamson or the “objective Bayesians” in believing that it is the result of conditionalizing a certain special Ur-prior probability function on the thinker’s “evidence”; others might follow “subjective Bayesians” like Richard Jeffrey in thinking of this probability function as the one that corresponds to the beliefs that the thinker actually has; and yet other views are also possible. I shall not try to adjudicate between these views until the second instalment of this trilogy. In this book, I shall simply propose that there is a probability function of this kind.

Putting together these degrees of incorrectness and this probability function, we can define the idea of a way of thinking’s expected degree of incorrectness (defining the idea of an “expected value” in the classical probabilistic style). My central proposal is that a way of thinking’s degree of irrationality is determined by how that way of thinking compares with

---

6 See for the idea of this special Ur-prior, see Williamson (2000, chap. 10); for a defence of objective Bayesianism, see Leitgeb and Pettigrew (2010).

7 For a defence of subjective Bayesianism, see Jeffrey (1984 and 2004).
the available alternative ways of thinking in terms of their expected degrees of incorrectness, according to the relevant probability function.

One way to put this idea is as follows. According to the probability function that rationally should be guiding one’s thinking, irrationality is bad news about correctness; and the more irrational one’s thinking is, the worse the news is (according to this probability function) about one’s thinking’s degree of incorrectness. To think in a perfectly rational way, one’s thinking must give one the best possible news about its degree of incorrectness. This is a more literal interpretation that I propose to defend of the metaphor of rationality’s being a means to the end of getting things right or having correct mental states. One of the main goals of this book is to articulate and defend this proposal, and to argue that it solves the puzzle about rationality that I discussed in the previous section.

**0.5 How central is the concept of rationality?**

As I explained in Section 0.1 above, it is a fundamental assumption of this investigation that there is a single concept of rationality that is central to many of the classic debates among epistemologists and philosophers of science (in the theoretical domain), and among decision theorists and philosophers who study ethics and reasons for action (in the practical domain). This assumption might be questioned.

First, one might point to the fact that many other terms besides ‘rational’ are used both in epistemology and in the study of practical reason: for example, some philosophers would prefer to speak of beliefs and choices’ being “reasonable” or “unreasonable”, “justified” or “unjustified”, or “warranted” and “unwarranted”, or of their being “supported” by the relevant “reasons”, or the like. Perhaps these locutions express different concepts from my talk of “rationality”? 

Secondly, one might question whether the term ‘rational’, as it is used in these branches of philosophy, is itself univocal. Several philosophers have suggested that there is a distinction between (a) a “substantive” kind of rationality, which consists in being supported in the appropriate way by the “reasons” that the relevant thinker has, and (b) a more “structural” kind of rationality, which consists in meeting certain pure constraints of “coherence”.

---

* For some philosophers who espouse a distinction of this sort, see Scanlon (2007) and Worsnip (2015).
I shall start with the first of these two questions. It is undoubtedly true that these terms – ‘rational’, ‘justified’, ‘warranted’, ‘entitled’, and so on – are used in more than one sense. For example, when Alvin Plantinga (1993) stipulatively uses ‘warrant’ to refer to whatever differentiates a mere true belief from knowledge, it seems that this use expresses a different concept from the intuitive notion of ‘justification’ – since the latter notion, as Edmund Gettier (1963) famously showed, clearly allows for the possibility of a justified true belief that does not count as knowledge. Nonetheless, as I shall argue here, in a very wide range of contexts, philosophers use these terms in such a way that they effectively express the same concepts.

It is a familiar point that the term ‘justified’ can be used to express both propositional justification and doxastic justification. But the very same distinction can be drawn using the term ‘rational’ as well. According to the position that will be defended in this book, the uses of the term ‘rational’ that I am concerned with express a normative concept – specifically, a concept that refers to the proper use of one’s faculties of thinking and reasoning. To say that you “rationally believe” a certain proposition \( p \) is to say that in believing \( p \), you are using these faculties properly in the relevant sense; this corresponds exactly to the contemporary epistemologists’ notion of doxastic justification. To say that “it is rational” for you to believe a certain proposition \( q \) in your current situation is to say that there is a way of believing \( q \) available to you in your current situation that would consist of the proper use of these faculties; this corresponds exactly to the contemporary epistemologists’ notion of propositional justification.

A broadly similar point holds of the use of the term ‘entitled’ in recent work in epistemology. Some epistemologists have recently shifted to speaking, not of “rational” or “justified” beliefs or processes of reasoning, but of beliefs that the thinker in question is “entitled” to have, or processes of reasoning that the thinker is “entitled” to perform. According to these epistemologists, a central part of epistemology is concerned with the

---

9 For a philosopher who argues that there are many different notions of “justified belief” which play importantly different roles, see Swinburne (2003).

10 For an influential recent discussion of this distinction, see Turri (2010). My account of this distinction will be given in Chapter 6.

11 See especially Burge (1993 and 2003) for his most influential discussions of entitlement. Another philosopher who articulates the central questions of epistemology in terms of entitlement is Peacocke (2003).
conditions under which thinkers are in this sense “entitled” to various beliefs and processes of belief-formation or belief-revision.

Here too, I shall argue, the difference between the two approaches is fundamentally terminological rather than conceptual. The terms ‘rationality’ and ‘entitlement’ are used in different philosophical traditions, but so far as I can see, in most philosophical discussions they express the very same range of concepts.

It seems plausible that the term ‘entitlement’ normally expresses a normative concept. The natural home of this term seems to be within systems of institutional rules: for example, if you are “entitled”, under the rules of the university that you work for, to go on sabbatical leave for a certain period of time, then according to those rules, it is permissible for you to go on sabbatical leave for that period; and if you communicate your intention to go on leave in the appropriate way, then the university will normally be obliged to allow you to take that period of sabbatical leave.

Among the epistemologists who use the term ‘entitlement’, it is widely supposed that “entitlement” differs in an important way from “justification”. In fact, different philosophers have proposed several different contrasts between justification and entitlement. For example, Crispin Wright (2004, 167) says that entitlement is “a kind of rational warrant” for accepting a proposition that is not to be identified with “having evidence for its truth”. Yuval Avnur (2012, 299) gives a somewhat different characterization when he says:

*Entitlement* is a positive epistemic status that can apply to either a belief or a belief-forming procedure. Having entitlement does not require any independent evidence or reason for the belief in the first case or for believing that the belief-forming procedure is reliable in the second case. Let us stipulate that evidence that one’s faculties are reliable is independent if and only if it was not produced by (or accessed via) those very faculties.

For Avnur, then, entitlement differs from other kinds of warrant in that entitlement does not require “independent evidence” (although having entitlement is also not inconsistent with having such evidence either).

It is clear that the overwhelming majority of formal epistemologists would agree that there are many cases where it is rational for a thinker to have a certain level of confidence in a proposition $p$, and what makes it the case that this is rational is not any “evidence” – let alone “independent evidence” – that the thinker agent possesses for or against this proposition $p$. For example, according to subjective Bayesians like Richard Jeffrey (2004),
one’s “prior” credences are automatically rational, so long as they are probabilistically coherent; the rationality of these prior credences does not depend on evidence that the thinker has for the truth the propositions in question.

So-called objective Bayesians reject the idea that our prior credences are automatically rational in this way; they may even say that there is only one possible set of prior credences that count as rational (such as the credences that conform to some version of the notorious “principle of indifference”). However, even for these objective Bayesians, it cannot be the evidence that the thinker has that makes these prior credences rational, since these are the credences that it is rational for the thinker to have in advance of all evidence whatsoever. In this way, the features of Avnur’s and Wright’s understanding of “entitlement” that we have just explored do not differentiate their notion of “entitlement” from the formal epistemologists’ notion of “rationality”. So far, then, it seems plausible to assume that the beliefs that a thinker is “entitled to” are precisely those beliefs that it is “rational” for the thinker to hold.

A different contrast between entitlement and other kinds of warrant is suggested by Burge. According to Burge (2003, 504), the key difference between entitlement and other kinds of warrant is that entitlement does not require that the thinker has “conceptual access” to the entitlement: as he says, “individuals can be epistemically entitled to a belief … without having the concepts needed to understand or even think the entitlement.”

While Burge’s formulations could be read in more than one way, it seems clear that whatever features he is ascribing to entitlement here, they are features that Bayesians and other formal epistemologists would equally ascribe to rationality. According to proponents of Bayesianism, the rational thinker need not have the concepts of probabilistic coherence, or of a warrant, or of a reason, or anything of that sort. She need not have the concept of evidence or even the concept of belief; she need not even have the concept of experience or perception, or the concept of a belief’s being incoherent. According to the subjective Bayesians, all that is required is that the thinker’s beliefs must actually be probabilistically coherent, and must evolve in response to evidence by means of conditionalization. Objective Bayesians would impose further conditions, but none of the conditions that have been proposed require that the thinker must possess any of these concepts. Here too, then,

---

12 For an example of an objective Bayesian approach, see Leitgeb and Pettigrew (2010).

13 For a discussion of how to interpret Burge’s work, see Casullo (2007).
we have found no reasons for distinguishing the notions that are expressed by the term ' entitlement' from those that are expressed by 'rationality'.

It seems that the points that I have made here about the terms 'entitled' and 'justified' could also be made about many of the other terms that are used to evaluate beliefs and other mental states and mental events, such as 'warranted' and 'reasonable', and the like. So the investigation of rationality that I shall undertake here seems clearly to be addressed to the same questions as the investigations that other philosophers have pursued into topics that they labelled using these other terms instead. In this way, even though I shall use the term 'rationality' throughout, I shall be engaging in a discussion with those other philosophers.

There is, admittedly, one significant difference between the tradition that speaks of “entitlement” or “warrant” and the tradition that favours the term ‘rationality’. The former tradition is concerned exclusively with epistemology – with the conditions under which we are entitled to believe, or warranted in believing, various propositions. The latter tradition is just as much concerned with rational choices and rational preferences as with rational beliefs. Offhand, it seems that we can apply the same terms, in the same range of senses, to both beliefs and choices. In what follows, I shall aim to speak about rationality in full generality, encompassing both rational belief and rational choice.

We must now address the second of the two questions that I canvassed above, about my assumption that there is a single concept of rationality that plays a central role in these philosophical debates. Perhaps there are really two different concepts, both expressed by the term ‘rationality’, that figure in these recent philosophical debates – one concept according to which rationality requires mere coherence (“rationality as coherence”, RC) and a second concept according to which rationality consists in having attitudes that are adequately supported or justified by one’s reasons (“rationality as reasons-responsiveness”, RR). It might be suggested that the first conception (RC) tells us to “have coherent attitudes!” – while the second conception (RR) tells us to “have reasonable / justified / warranted attitudes!”

In fact, however – at least given certain extremely plausible further assumptions – this alleged contrast between rationality-as-coherence (RC) and rationality-as-reasons-responsiveness (RR) is in fact a distinction without a difference. One of these plausible further assumptions is the following. First, if (as (RR) implies) rationality consists in having

---

14 This way of characterizing the distinction was urged on me by an anonymous reader for Oxford University Press.
attitudes that are “justified” or “supported” by one’s “reasons”, one must have the relevant reasons at the relevant time – and whether or not one “has” a certain reason at this time depends on the mental states and mental events that are present in one’s mind at that time.\(^{15}\) (If rationality-as-reasons-responsiveness (RR) did not satisfy this condition, then it would be incompatible with the kind of “internalism” that, according to the arguments of this book, is one of the key distinguishing features of the concept of rationality.) Moreover, whether the reasons in question “justify” or “support” an attitude is simply a matter of whether the attitude “fits” with those reasons – which is determined purely by the nature of those reasons and of the attitude in question. In this way, the central demand of rationality-as-reasons-responsiveness is that one should have attitudes that “fit” with the reasons that one “has” – and what count as the “reasons that one has” is determined purely by the mental states and mental events that are present in one’s mind.

However, for the purposes of understanding rationality-as-coherence (RC), we need to understand the notion of “coherence” in a sufficiently broad and general way. When “coherence” is understood in this general way, for a thinker’s mental states and events to “cohere” is just for them to stand in a relation that corresponds to this relation of “fitting” together. Specifically, it is for these mental states and events to stand in a relation that is (a) determined purely by the nature and contents of these mental states and events, and (b) necessary for this combination of mental states and events to be rational. When coherence is understood in this general way, we can, as I have already mentioned, make sense of the idea of relations of coherence between one’s sensory experiences and one’s beliefs – not just relations of coherence between one’s beliefs themselves. Moreover, when we evaluate the coherence of different possible ways of thinking that are available to the thinker, we may treat ways of thinking that involve different sensory experiences as not being “available”, even though ways of thinking that involve different beliefs or different intentions may count as “available” in the relevant way. (So this broad notion of “coherence” can encompass the “foundherentist” approach that has been advocated by Susan Haack (1995).)

Once we understand “coherence” in this broad way, and once we recognize that the reasons that one “has” are determined purely by one’s mental states and mental events, it becomes clear that rationality-as-reasons-responsiveness (RR) cannot demand any more of us than that our attitudes should cohere (in this broad sense) with the mental states and mental events that are present in one’s mind at that time.

\(^{15}\) For the idea of the distinction between reasons that there are for one to have an attitude and the reasons that one has to have the attitude, see especially Schroeder (2008).
events that constitute our possession of the reasons that we have. In this way, then, these plausible assumptions imply that a mental state (or mental event) \( M \) is supported or justified by one’s “reasons” if and only if this mental state (or event) \( M \) “coheres” with the mental states and events that constitute one’s “having” those reasons. In short, the norm “Have a coherent set of mental states and events!” and the norm “Have mental states and events that are justified or supported by the reasons that you have!” turn out to be equivalent. In short, rationality-as-reasons-responsiveness (RR) is not a different notion from rationality-as-coherence (RC). The very idea of the distinction is illusory.\(^{16}\)

For these reasons, then, it seems plausible that we can answer these questions that can be raised about my assumption that there is a single concept of rationality that plays a central role in these philosophical debates.

**0.6 What are normative concepts?**

So far, I have only given a rough indication of what it means to say that “rationality” is a normative concept. What more precisely is it for a concept to be a normative concept?

Many philosophers suppose that the way to answer this question is by postulating that there are some utterly primitive normative concepts – such as the concept of a “reason” for an action or attitude, or the concept that is expressed by certain central uses of ‘ought’ or ‘should’ – and all other normative concepts are to be defined or analysed in terms of these primitive concepts.\(^{17}\)

Methodologically, these philosophers simply rely on our competence with certain words in our language: they simply deploy certain words – like ‘reason’ or ‘ought’ or the like – in the hope that we will somehow cotton on to the allegedly primitive normative concept that these words express in the context. Some of these philosophers seem to think that this methodology is acceptable because (as they assume) there is only one normative concept that is expressed by the term ‘ought’, or only one normative concept that can be expressed by speaking about “reasons” for acting or thinking in various ways – or at most, a small number of such concepts that are easily distinguished from each other. Much of the recent

\(^{16}\)There are admittedly some ingenious arguments – most notably, the argument of Worsnip (2015) – for the claim that RC (rationality-as-coherence) and RR (rationality-as-reasons-responsiveness) must “come apart”. My own view is that at best these arguments show that some kinds of coherence “come apart” from other kinds of coherence. These arguments will be discussed in greater detail in Chapter 2 (Section 2.4).

\(^{17}\)For example, one philosopher who expresses sympathy for the view that all normative concepts can be analysed in terms of the fundamental notion of a “reason” for an action or attitude is Scanlon (2014, 2).
discussion of the normativity of rationality seems to be based on this methodological approach to understanding what it means to call a concept a normative concept.

For example, John Broome (2013, 144) says:

I explained ... what it means to say a requirement on you to \( F \) is normative: it means the requirement is a reason for you to \( F \).\(^{18}\)

In a similar way, Nicholas Southwood (2008, 18) also associates the normativity of rationality with reasons:

For the normativity of rationality is a matter of reasons that are internal to rationality, not reasons that are external to it. It is a matter, if you like, of reasons of rationality, not independent reasons to obey rationality.

A slightly different conception between the normativity of rationality and reasons is proposed by Joseph Raz (2011, 89), who identifies irrationality the “malfuctioning” of our “rational powers” – which Raz (2011, 85) identifies with our capacities for recognizing and responding appropriately to reasons.\(^{19}\)

A good example of a philosopher who understands the normativity of rationality in terms of ‘ought’ rather than in terms of “reasons” is Errol Lord (forthcoming, 8–9), who defends the claim that rationality is “strongly deontically significant”, which he defines as follows:

\[
\text{rationality ... is strongly [deontically] significant if it is such that we always ought to do what rationality requires.}
\]

Unlike these other philosophers, I shall not follow either of these approaches here. First, I shall not take the notion of a “reason” for an action or an attitude as fundamental in this way. As I shall explain in Chapter 4, the language of “reasons” is significantly more complex than many philosophers have recognized. There are in fact many different normative concepts that can be expressed by the term ‘reason’, and none of these concepts is any more fundamental or primitive than any other. In fact, as I shall argue, the most plausible

---

\(^{18}\) For another influential discussion of the normativity of rationality that associates normativity with reasons, see Kolodny (2005).

\(^{19}\) Besides Raz, other philosophers who advocate this sort of “reasons fundamentalism” include Scanlon (2014, 2) and Schroeder (2007). The idea of identifying rationality with “responding appropriately to reasons” is also taken up by Kolodny (2005), who identifies rationality with responding to one’s beliefs about reasons; for some criticism of this idea, see Broome (2013, chap. 5).
account of this language that deals adequately with all this complexity is incompatible with the assumption that the concept of a “reason” is primitive in this way.

Secondly, I shall also not follow the philosophers who (like Errol Lord) look to the concepts that can be expressed by ‘ought’ to identify the fundamental normative concepts. Here, too, as I argue in Chapter 5, we find that ‘ought’ is a systematically context-sensitive term, expressing different concepts on different occasions. Nonetheless, in spite of this great proliferation of ‘ought’-concepts, the broad outline of the classical semantics for deontic modal terms like ‘ought’ and ‘should’ turns out to be correct. For every normative use of a term like ‘ought’ or ‘should’, some value – that is, some way of ranking alternative possibilities as better or worse – is presupposed. Roughly, ‘Ought (p)’ is true if and only if p is true at all the worlds in the relevant domain of available worlds that are ranked as optimal by this value. Although this account of ‘ought’ is rejected by some philosophers, such as Mark Schroeder (2011) among others, it is the standard account among linguists and semanticists.

In general, I propose that what distinguishes normative concepts from all other concepts is the distinctive sort of conceptual role that they play – what we could call, in general, a guiding or regulative role in various kinds of reasoning. There are in fact a huge number of such normative concepts, corresponding to the different kinds of guiding or regulative role that these concepts play in different kinds of reasoning. Nonetheless, among the normative concepts, the most fundamental concepts – or so I shall argue – are broadly speaking evaluative concepts, concepts that stand for the many different ways in which things can be good or bad, or in which one thing can be better or worse than another. As I shall argue, rationality is also a kind of value. The beliefs or intentions or processes of reasoning or ways of thinking that are more rational (or less irrational) are in a certain way better than the corresponding items that are less rational (or more irrational).

Still, there are many different kinds of value – many ways in which things can be good or bad, or better or worse.20 One of the main tasks of the book is to address the question of how exactly rationality differs from all these other values. Still, we will not even understand this question until we appreciate that rationality is itself a value in this way.

In short, the approach that has dominated most of the debate about the normativity of rationality relies on the assumption that there are a small number of primitive normative

---

20 Philosophers have greatly underestimated the enormous variety of different kinds of value. Even philosophers who were at pains to insist on the “varieties of goodness”, like von Wright (1963) and Thomson (2008), give what seem to me to radically incomplete catalogues of all the different values that there are.
concepts – such as the concepts expressed by ‘reason’ or ‘ought’. This assumption cannot be reconciled with the best available account of the nature of normative concepts, or with the linguistic data that needs to be integrated with any adequate account of the nature of these concepts. This is why much of Chapters 4 and 5 of this book is devoted to considering these issues about the nature of normative concepts and the meaning of normative language.

In general, the existing discussion about the nature of rationality has for the most part been divided into several separate debates, where the philosophers engaged in one of these debates have rarely paid much attention to what was being discussed in the other debates. Besides the debates about the normativity of rationality that I have mentioned, there has also been a debate within epistemology about the nature of rational or justified belief, which has for the most part been pursued quite independently of the debate that focuses on practical reason and the nature of rational choice and rational intention. Moreover, each of these two debates has also, at least for the most part, been pursued independently of a third debate within the more formal branches of philosophy, where philosophers have sought to study rational belief and rational choice by means of a range of precise mathematical tools (such as the tools that are provided by logic and probability theory).

One of the main goals of this book and its planned sequels is to bring these different debates into a conversation with each other. I am convinced that it is only through synthesizing the insights that have been uncovered in these different philosophical debates into a coherently unified account that we can achieve an adequate understanding of these questions.

0.7 The plan of this book

Although every chapter of this book adds a crucial element to the whole picture of rationality, each chapter is designed to be capable of being read by itself. The overall plan of the book is as follows. The problem that the book aims to address is set out in Chapter 1, and Chapters 2-3 address some of the simpler aspects of the problem. Chapters 4–5 develop accounts of the concepts that are expressed by terms like ‘reason’ and ‘ought’; these accounts reveal the limitations of rival approaches, and help to motivate my solution to the problem. Finally, the core of my solution to the problem is given in Chapters 6–9; a reader who wants to focus on the heart of my theory could concentrate on these four chapters.

Here is a more detailed plan of the whole book. In Chapter 1, I shall set out some considerations that seem prima facie to count in favour of the view that the term ‘rationality’ (as used by formal epistemologists and decision theorists and the like)
expresses a normative concept. Then I shall consider four objections to that view: (a) the first objection is based on examples in which unusual circumstances (like a mischievous demon or the like) will either reward irrationality or punish rationality; (b) the second objection is based on the alleged possibility of rational false beliefs about how one ought to think; (c) the third objection is based on the principle that ‘ought’ implies ‘can’; and (d) the fourth objection based on the worry that rationality requires nothing more than coherence, which is just a pretty pattern of mental states, with no real normative importance of any kind. In this way, Chapter 1 sets up a puzzle: on the one hand, it seems overwhelmingly plausible that rationality is a normative concept, but there also seem to be some powerful objections against the view that it can possibly be a normative concept.

In Chapter 2, I shall set out a preliminary answer to the first two of these four objections. The heart of this answer is the idea that rationality is a normative notion that has the three distinctive features that I focused on in Section 0.2 above: first, rationality is exemplified by mental states and mental events; secondly, the requirements of rationality are constitutive norms; and thirdly, internalism about rationality is true – the requirements of rationality supervene on what is present in the relevant thinker’s mind at the relevant time. This internalist idea is related to the view that there are different kinds of ‘ought’ – including more “subjective” and more “objective” kinds of ‘ought’. Indeed, the notion of a “rational requirement” implies what is in effect the most subjective of all kinds of ‘ought’ – a sense of ‘ought’ on which the truth about how an agent “ought” to think supervenes purely on the mental states and events in the agent’s mind at that time. As this chapter explains, these features of rationality can form the basis of an answer to the first two objections raised in the previous chapter.

In Chapter 3, I offer an answer to the third of the four objections that were raised in Chapter 1 – that is, the objection that is based on the principle that ‘ought’ implies ‘can’. This chapter gives a straightforward solution to this objection. It defends the claim that the notion of a rational requirement implies a kind of ‘ought’, and the principle that ‘ought’ implies ‘can’, by arguing that there is a sense of ‘can’ such that whenever an agent is rationally required to think in a certain way at a certain time, it is true that that agent “can” think in that way at that time. Since the kind of ‘ought’ implied by the notion of a rational requirement is not a practical ‘ought’, which applies directly to action, but a kind of ‘ought’ that applies to thinking and reasoning, being rationally required to think in a certain way does not imply that one can think in that way at will. However, being rationally required to think in a certain way does imply that thinking in that way consists of a certain kind of exercise of one’s mental capacities, and that one has an opportunity at the relevant time for that kind of exercise of one’s capacities.
With Chapter 4, I begin my long response to the fourth and final objection – the objection that is based on the thought that it is not clear why coherence matters or has any real normative significance. In this chapter, I respond to a question that many readers will be eager to make. Surely, many philosophers will think, normativity is all about reasons. If I am going to argue that rationality is a normative concept, surely I will need to explain the connection between rationality and reasons? In this chapter, I explain why I will not make any notion of “reasons” central to my discussion. Philosophers who focus centrally on “reasons” have been repeatedly led astray by the complexities of the meaning that the word ‘reason’ has in English. In fact, there are many more different notions that can be expressed by the word ‘reason’ than most philosophers have recognized; and the best account of how all these different notions are related to each other implies that these notions are not primitive or basic normative concepts at all.

In Chapter 5, I return to the idea, which was first introduced in Chapter 2, of the distinction between the “objective” and the “subjective” forms of ‘ought’, and to the idea that the notion of a rational requirement implies a highly subjective kind of ‘ought’. In this chapter, I give a semantic account of the truth conditions of sentences involving these different kinds of ‘ought’. In this way, this chapter is a contribution to the burgeoning literature on the semantics of so-called “weak necessity” modals like ‘ought’ and ‘should’. The semantic account proposed in this chapter centrally involves the ideas of a probability function, and of ranking possibilities according to their expected value, which will play important roles in the account of rationality that is given in Chapter 9. But the big lesson of this semantic account is that if any normative concepts count as the most fundamental, they are the concepts of all the various kinds of values. This is because according to this account, it is in terms of values – or expected values, according to some contextually relevant probability function – that the relevant possibilities are ranked.

If, as Chapter 5 in effect implies, the concepts of the various kinds of values are particularly fundamental, what values are relevant to understanding rationality? Chapter 6 starts to address this question, by proposing that rationality is itself a kind of value – indeed, more specifically, that rationality is a certain kind of virtue. Unlike the virtues that are primarily manifested in the agent’s actions or emotions, rationality is a virtue that is primarily manifested in the agent’s reasoning – that is, in the agent’s mental states and mental events. As with the other virtues, there are at least two connected values: (a) the value of rational dispositions of thought and reasoning, and (b) the value of the mental states and events that result from such dispositions. In fact, however, we can also draw a further distinction, between (b(i)) mental states or events that it is “abstractly” rational for a thinker to have, and (b(ii)) mental states that a thinker rationally holds, or mental events of the thinker’s
reasoning rationally. An irrational person might by a lucky accident form a belief that it is “abstractly” rational for her to form; but to be reasoning rationally, the thinker must actually manifest some appropriately rational dispositions. As I explain, this is in fact the very same distinction that epistemologists have come in recent years to express by contrasting “propositional justification” and “doxastic justification”.

Other fundamental features of other virtues are also exemplified by rationality. Rationality, like the other virtues, comes in degrees: some decisions and some beliefs are more irrational than others. It is this that enables us to understand the idea of the requirements of a virtue – such as the requirements of rationality or the requirements of justice – as equivalent to a kind of ‘ought’. At the same time, this account of what it means to speak of the requirements of rationality allows us to revisit some of the issues that were considered in Chapter 3 about the kind of ‘can’ that is implied by the rational ‘ought’.

In chapter 7, I address two questions that are raised by the kind of “internalism” that I first introduced in Chapter 2. First, what exactly is the distinction between the “external world” and what is “present in the thinker’s mind” that I invoked in characterizing “internalism”? Secondly, even if there are some normative concepts that evaluate mental states and mental events purely on the basis of how they relate to what is present in the thinker’s mind, why are such normative concepts of any importance to us? In effect, the goal of this chapter is to explain the sort of “internalism” that I am claiming to be true of rationality. The key idea is that internal norms of this sort can guide our thinking directly. More external norms can guide us only indirectly, by means of our having beliefs or thoughts about those external norms; by contrast, internal norms can guide us directly – we can think in a certain way precisely because that way of thinking really is a rational way for us to think. As I shall argue, this point explains why the concept of rationality is such an important concept for central branches of philosophy like epistemology and ethics.

In chapter 8, I turn to the big question that has haunted us since the first chapter. Even if, as Chapter 7 has argued, we can be “directly guided” by the internal requirements of rationality, what is the point of being guided by such an internal norm? How can any such internal norms matter? The goal of this chapter is to clarify the meaning of this question. According to the clarification that I give, answering this question would involve giving a general account of the value of rationality which makes it intelligible how it can play the fundamental normative role that it does. Then I explore some rival approaches to understanding the value of rationality: the approach that appeals to mathematical theorems to the effect that thinkers who violate certain requirements of rationality are vulnerable to what is known as a “Dutch book”, and the approach that is based on the idea
that these requirements of rationality are “constitutive” of agency or of being an interpretable thinker in the first place. Neither of these approaches, as I argue, can give a satisfactory solution to our problem.

In Chapter 9, I offer my own answer to the question that was explored in Chapter 8. I propose that the value of rationality is, in a way, not completely free-standing. Any adequate account of rationality must appeal to another radically different value: the value of correctness, which has a fundamental connection to rationality. As I argue in this chapter, the connection between the two concepts is probabilistic. In a slogan, irrationality is bad news about correctness; and the more irrational one’s thinking is, the worse the news is about correctness. It is this that justifies the metaphor that rationality is the “means” towards a more external “aim” – the aim of correctness, or of getting things right in our thinking.

Finally, in the Conclusion, I explain how the general account of rationality given here sets the agenda for the accounts of rational belief and of rational choice that I aim to give in the second and third instalments of this projected trilogy.

0.8 Some remarks on method: Properties and concepts

My fundamental concern, in this whole trilogy, is with the question of what it is for a mental state or event to count as rational. That is, in effect, my concern is with the property of rationality itself – with what makes it the case that a mental event or mental state has this property.

Nonetheless, I shall spend a large part of this first book of the trilogy speaking, not about the property of rationality, but about the concept of rationality. The property and the concept of rationality are crucially distinct. The property of rationality is the feature of ways of thinking that makes those ways of thinking count as rational rather than irrational. A person’s mental states and events could be rational even if the person never thinks about whether or not they are rational – indeed, they could be rational even if the person does not even possess the concept of rationality. Roughly, the distinction is this: we use the concept of rationality to think about the property of rationality; the property is exemplified by the items that are themselves rational, and the concept is that constituent of thoughts about rationality in virtue of which those thoughts count as being about rationality.

If my fundamental concern is with the property of rationality, why do I plan to spend so much time talking about the concept of rationality? There are two main reasons for this.
First, we need to identify which property we are concerned with. The word ‘rational’ expresses many different concepts in different contexts. So we shall have to probe the different meanings that this word and other related terms can have, in order to be confident that we have identified the particular property that concerns us.

Secondly, it seems plausible that, in this case, some truths about the nature of this property are conceptual truths. That is, certain propositions about the nature of the property of rationality are themselves guaranteed to be true by the nature of the concept that we use to pick out the property. So studying the concept of rationality can be one way of finding out about the property that the concept stands for.

This is not to say that all truths about rationality can be discovered purely by investigating the concept of rationality. Indeed, the features of rationality that we shall investigate in the second and third instalments of this trilogy are features that cannot be ascertained purely by studying the concept of rationality; they will be features that can be discovered only by combining our knowledge of these conceptual truths with further considerations concerning the nature of the relevant types of mental states and events, like belief and choice. Still, I believe, the claims about rationality defended in this first instalment of the trilogy are all conceptual truths, propositions that are guaranteed to be true by the nature of the concept of rationality.

Some philosophers might object that my claims about the distinguishing features of the concept of rationality cannot possibly be conceptual truths. After all, every one of these claims – that rationality is a normative concept, that internalism is true of rationality, and that it has in some sense an external aim – is denied by some philosophers, who are apparently competent users of the concept in question. However, this objection is based on a view of concepts that is being rejected here – a view according to which possessing a concept necessarily involves already knowing or appreciating these conceptual truths. On the view of concepts that I prefer, possessing a concept consists in the concept’s playing the appropriate conceptual role in one’s thinking. A concept can play this role in one’s thinking even if one is not consciously aware of its doing so.

In general, the most reliable method for studying concepts is not simply to introspect, and to figure out what one already believes. Instead, a crucial part of the data for studying concepts is linguistic. This is because the easiest way for us to identify concepts is by means of the words that we use to express them. However, even this way of identifying concepts is not straightforward, since virtually none of the words that express philosophically interesting concepts is univocal. Moreover, although almost all the words that express such concepts are in some way polysemous or context-sensitive, it is not always obvious to us
whether two uses of one of these words involve the same sense, and express the same concept, or whether instead they involve different senses, and express different concepts, from each other. This creates a danger that can in principle affect our thought in any domain of investigation, but is particularly acute in philosophy: our linguistic intuitions can lead us astray if we wrongly assume that two uses of a word express the same concept when in fact they express different concepts. (In principle, we can also miss out on important information if we wrongly assume that two uses express different concepts when in fact they express the same concept; but in this case the danger is merely that we will fail to appreciate some truths rather than that we will be positively misled.)

For these reasons, then, it is methodologically crucial to consider the linguistic data carefully. This is not because questions about the meaning of particular English words are of fundamental importance for philosophy. (Such questions really belong to empirical linguistics rather than to philosophy.) It is just that without attention to language we cannot be confident of correctly identifying all of the different concepts that we need to be concerned with, or of understanding their relations and connections to each other.

In this way, I shall spend part of this book exploring the nature of our concept of rationality, and of the words that can be used in English and other languages to express this concept; but my ultimate focus throughout will be on the nature of the property of rationality itself.

For this reason, although I shall build on my earlier work on normative concepts, my focus in this work will be somewhat different. In my earlier work, I was interested in the nature of normative thought itself; I explained normative thought in terms of the use of normative concepts, which on my account are fundamentally individuated by their distinctive conceptual role; in terms of Frege’s famous distinction, a concept’s conceptual role corresponds to the level of sense rather than the level of reference. In this book, my interest in normative concepts is primarily at the level of reference – that is, my interest is primarily in the truth conditions of normative thoughts and utterances rather than in any other dimensions of their meaning. I shall study the concept of rationality, not primarily in order to discover the nature of this concept itself, but in order to see what light can be shed on the nature of the property of rationality that the concept stands for.

The method appropriate for such a conceptual investigation is not fundamentally different from the method that is appropriate for any other sort of constructive philosophical theorizing. The goal is to develop a comprehensive theory, in some detail, and to make it plausible that the theory has some striking advantages: first, the theory is internally coherent; secondly, it can explain the phenomena and solve the problems that the theory is
called upon to explain and solve; and thirdly, it has the resources to solve any further problems to which the theory itself gives rise, and to defend itself against all the serious objections that can be raised against it.

Making it plausible that this theory has these advantages is, of course, by no means a demonstration that the theory is correct. To provide such a demonstration one would have to compare the theory to all possible alternative theories, and to show that the theory gives a better explanation of the phenomena than all of those competitor theories. It will certainly not be possible for me to compare my theory to all possible competitors in this way.

Indeed, in this book, my focus will principally be on developing a comprehensive theory, in some degree of detail. As I see it, there are three main kinds of philosophical research that are practised today:

a. Some philosophical research remains close to everyday thought (or to the thought of non-philosophical inquirers such as scientists or lawyers or the like), and explores how such everyday thought gives rise to problems and puzzles that philosophers need to address.

b. Some philosophical research focuses on constructive theory-building, developing a detailed conception of a certain domain.

c. Some philosophical research focuses on the debate between such constructive theories, arguing that one side in the debate has more “plausibility points”, as David Enoch (2013) would put it, than its rivals.

It seems clear that all three kinds of research are necessary. But it plausibly also makes sense for there to be a division of labour, in which some philosophers focus more on one kind of research, and other philosophers focus on the other kinds. (This sort of division of labour is certainly found in many of the natural sciences, where some scientists pursue “theoretical” work, developing specific theories in precise detail, and others pursue more “experimental” work, testing which of the theories does the best job of explaining the data that needs to be explained.)

This book focuses mostly on the second kind of research (b) – on constructive theory-building. Along the way, in constructing the theory, I try to bring out the overall coherence of the theory, and to highlight some of the data that the theory could explain; but significantly less space is devoted to arguing that my theory is preferable to the alternatives.
My own view is that, of these three kinds of philosophical research, the contemporary literature already contains enough of the third kind (c) – that is, of arguments that effectively amount to haggling over how many “plausibility points” should be awarded to each of the already-familiar positions that are discussed in these debates. Instead, I believe, what is needed is more research of the first and second kinds – introducing either (a) new problems and puzzles or (b) new theories for philosophers to consider. This is why my book was mostly focused on this second type of research (b). It is my hope that this type of research will be a worthwhile contribution to the philosophical debate.

As I have explained above, my account will involve the following three central theses: (1) rationality is a normative concept; (2) rationality is internal (what is rational supervenes on the mental states and events that are present in the relevant thinker’s mind at the relevant time); (3) rationality has, in some sense, an external aim or goal. A further claim that plays an important, though less central, role in my argument is: (4) the thesis that the requirements of rationality are constitutive of the types of mental states involved. Each of these four theses is familiar, and has been defended by other philosophers. Indeed, each of the first three theses ((1), (2), and (3)) has probably been accepted by the majority of philosophers who have worked on these topics. Nonetheless, many philosophers have doubted whether all these views are correct, or can be reconciled with each other; and few philosophers have attempted to work out a detailed theory that explains how all of these ideas can fit together.

In a sense, then, this work does not strive to be revolutionary. It is the theories that oppose mine on one or more of these central theses that count as revisionist. The goal of my theory is not to bring about a radical revolution in the study of rationality, but rather to synthesize many insights that many philosophers have had over the years into a coherent comprehensive theory.