Normative Judgement, Rationality, and Reflective Agency

Hugh Michael Barrett

November 2021

A thesis submitted for the degree of Doctor of Philosophy of The Australian National University.

Declaration

This thesis contains no material that has been accepted for the award of any other degree or diploma in any university. To the best of the author's knowledge, it contains no material previously published or written by another person, except where due reference is made in the text.

Word count: 61,383 (excluding the appendix, footnotes, and bibliography)

Hugh Michael Barrett

November 2021

Acknowledgements

I am grateful to my primary supervisor, Nicholas Southwood, for his helpful feedback, encouragement, and considerable patience and forbearance in helping me to get this thesis finished. Thanks also to Phillip Pettit for helpful conversations and written feedback early on in the project. Thanks to everyone who provided feedback on talks I presented at ANU and at the Kioloa graduate student conferences, and to Daniel Stoljar and Seth Lazar for being the other members of my supervisory panel. The ANU is a great place to study philosophy; I am grateful for the countless talks by world-class philosophers that I was able to attend. Some of them I even understood. Finally, thanks to my family—in particular to my wife and best friend Suhasini Sumithra for her constant love, support, and good humour while I persisted with this project for longer than I should have.

This research is supported by an Australian Government Research Training Program (RTP) Scholarship.

Abstract

Normative judgements have distinctive features that call out for explanation. In this thesis I provide an explanation of these distinctive features by defending three main philosophical positions. First, I provide a novel analysis of normative concepts in terms of the concept of rationality. Secondly, I provide an account of the concept of rationality in terms of the role that it plays in the ascription of mental states. And, finally, I defend a cognitivist account of reflective agency, according to which self-governing agents regulate their own mental states by self-ascribing them. The overall picture I will be developing can be summarised as follows: normative judgements are judgements about rationality, which rationally constrain our attitudes because they rationally constrain the self-ascriptions by means of which we regulate those attitudes.

Contents

0 Introduction

Normative judgements have distinctive features that call out for explanation. In this thesis I provide an explanation of these distinctive features by defending three main philosophical positions. First, I provide a novel analysis of normative concepts in terms of the concept of rationality. Secondly, I provide an account of the concept of rationality in terms of the role that it plays in the ascription of mental states. And, finally, I defend a cognitivist account of reflective agency, according to which self-governing agents regulate their own mental states by self-ascribing them. The overall picture I will be developing can be summarised as follows: normative judgements are judgements about rationality, which rationally constrain our attitudes because they rationally constrain the self-ascriptions by means of which we regulate those attitudes.

0.1 Normative Judgement

What is a normative judgement? A normative judgement is a kind of *attitude*—that is, a mental state rather than a non-mental entity such as an utterance or sentence. It turns out to be somewhat difficult, however, to delineate the normative domain in a simple way (Finlay, 2019). In this section, I will outline a general strategy for identifying normative judgements. The strategy is to describe the distinctive rational principles to which normative judgements are subject. Over the course of the following chapters, I will describe these rational principles in detail, and provide an account of normative judgement and reflective agency that vindicates them.

One thing that many have thought to be distinctive about normative judgements is that they all either are, or bear some kind of rational connection to, judgements involving the central normative concepts OUGHT or REASON (Wedgwood, 2013; Schroeder, 2007, pp. 80-81; Darwall, 2001; Raz, 2000; 2010; Tappolet, 2011; Tappolet and Voizard, 2011; Dancy, 2000a; Boghossian, 2005). The following, for example, seem to be paradigm cases of normative judgements:

> The judgement that I ought to call my parents this afternoon.
>
> The judgement that John's family history of heart disease is a reason for him to cut back on red meat.

According to this line of thought, normative judgements about what is obligatory, permissible, forbidden, right, wrong, good, bad, admirable, contemptible, benevolent, malevolent, and so on, are normative in virtue of bearing some rational connection to judgements about what ought to be the case, or what reasons there are. In chapters one and two and the appendix of this thesis, I describe the distinctive normative principles that govern how judgements involving the central normative concepts OUGHT, MAY, MUST, REASON (mass), REASON (count), and fitting-response concepts relate to each other, and provide analyses of these concepts that vindicate these rational principles.

Attempting to identify normative judgements solely by this criterion, however, is complicated by the fact that words like 'ought' and 'reason' can have different meanings in different contexts. 'Ought' and 'reason' can be interpreted in many different ways. For example, according to the standard view in linguistics, 'ought' is semantically incomplete, requiring for its interpretation specification of a particular kind of modality, given either by linguistic means or by the context of use (Kratzer, 1977; 1981; Portner, 2009). For example, 'ought' can express an *epistemic* modality—'There ought to be some beer in the fridge' (i.e., ought, *in view of what I know*…)—various different types of *deontic* modality—'The rich ought to give money to the poor' (i.e., ought, *in view of the requirements of morality*…), 'We ought to pay the real-estate tax' (i.e., ought, *in view of the requirements of the law*…), etc.—*bouletic* modality—'You ought to try this chocolate' (i.e., ought, *in view of your desires*)—and *teleological* modality—'You ought to add more salt to the soup' (i.e., ought, *in view of the goal of making a tasty soup*…) (Portner, 2009, p. 135). In each of these example sentences, 'ought' means something different. When someone refers to the judgement that something 'ought' to be the case, then, there are various different types of judgement to which they might be referring. Which of them is the type of judgement that plays the central role in defining what a normative judgement is?

Broome (2013) provides a useful way of identifying the relevant sense of 'ought'. According to Broome, the relevant sense of 'ought' is the sense of 'ought' that figures in the statement of the rational requirement he calls 'Enkrasia'.

> *Enkrasia, roughly.* Rationality requires of you that, if you believe that you yourself ought [to] F, you intend that you F. (p. 23)

In chapter four, I investigate the ways in which different types of normative judgement are related to other types of mental state through various distinctive enkratic principles. I argue that while, in fact, Enkrasia is false, there are a number of other similar enkratic principles that are true, such that normative judgements can be understood as judgements that are, or are related in the right way to judgements that are, subject principles of this kind.

Finally, normative judgements are also subject to distinctive *epistemic* principles. In chapter two, I discuss two distinctive epistemic principles to which normative judgements are subject. The first principle, Rationalism, says, roughly, that one can know *a priori* whether a normative proposition is true, given sufficient information about the non-normative facts. The second principle is a weakened version of Hume's Law—the thesis that one cannot derive an 'ought' from an 'is'.

Having identified these rational principles, we are now in a position to say roughly what normative judgements are. Normative judgements are judgements that are related to each other by the rational principles identified in chapters one and two of this thesis, that are related to other types of mental states by the enkratic principles identified in chapter four of this thesis, and that are subject to the epistemic principles that are set out in chapter two of this thesis.

0.2 Explaining the Distinctive Features of Normative Judgement

I have just identified a number of distinctive rational principles that apply to normative judgements. My aim in this thesis is to make sense of the fact that normative judgements are subject to these principles. I do so by defending three views: a rationality-first analysis of normative concepts; an account of the concept RATIONAL in terms of its role in interpretation; and a cognitivist account of reflective agency. I describe these below.

0.2.1 A Rationality-First Analysis of Normative Concepts

In chapters one and two, I explain why normative judgements are related to each other by the rational principles they are by providing an analysis of normative

concepts in terms of the concept of rationality. I begin in chapter one by focusing on the concepts MAY, MUST, and OUGHT. I argue that MAY and MUST can be analysed in terms of fitting-response concepts such as PERMISSIBLE, CONCEIVABLE, or DESIRABLE, and I argue that OUGHT can be analysed as a counterfactual MUST, such that to judge that something ought to be the case is, roughly, to judge that, if it were required or not permitted, then it would be required.

According to this analysis of OUGHT, to judge that something ought to be so is, roughly, to judge that it is *closer* to being required than it is to being not permitted. I build on this idea in an appendix by developing what I call the 'closeness' analysis of REASON (mass) and REASON (count). According to the closeness analysis, judgements about how much reason there is for something to be so are judgements about the relative closeness of worlds at which it is permitted, and judgements about how strong a reason some fact is in favour of some response are judgements about the relative closeness of worlds at which that fact explains why that response is permitted. Finally, in chapter two, I provide an analysis of fitting-attitude concepts, such as DESIRABLE and ADMIRABLE, in terms of the concept RATIONAL that is inspired by Rabinowicz' (2008; 2012; 2017) accounts of value and probability relations.

0.2.2 An Interpretationist Account of the Concept of Rationality

In the first two chapters of the thesis, I show that the central normative concepts can ultimately be analysed in terms of the concept RATIONAL. In chapter two I provide what I will call an 'interpretationist' account of this concept in terms of the role that it plays in the ascription of mental states. I identify this role by identifying a set of rational principles that link beliefs about rationality to mental-state ascriptions—namely a set of principles of *charity* or *rationalisation*. For example, there is a principle according to which, if one judges that, given their behaviour and environment, someone is rationally required to be in some mental state, then one is disposed to ascribe that mental state to them. It is the concept that plays this role in regulating the ascription of mental states that I have in mind when I refer to the concept of rationality.

I demonstrate in this chapter that judgements involving this concept are subject to epistemic principles that are analogous to the two epistemic principles identified

above—Rationalism and Hume's Law. Since, according to the analyses of normative concepts just discussed, normative judgements can ultimately be understood as judgements about rationality, this explains why normative judgements are subject to these epistemic principles.

0.2.3 A Cognitivist Account of Reflective Agency

In chapter three, I present what I call a 'cognitivist' account of reflective agency. According to this account, self-governing agents have a capacity to regulate their own mental states by self-ascribing them. After making the case in chapter three that cognitivism about endorsement is a viable position, I demonstrate in chapter four that this account of reflective endorsement, in combination with the previous analyses of normative concepts and account of rationality, explains the various enkratic principles to which normative judgements are subject. By this point, I will have shown that the three positions defended in this thesis together explain the main distinctive rational principles that apply to normative judgements. I conclude by considering some other rational principles that have to do with the relation between normative judgements and other phenomena such as advice, criticism, and blame, and sketch some avenues for future research in this area.

0.3 Rational Principles

Before I move on to the substantive discussion of normative judgements and the rational principles to which they are subject, I should say something about just what I take rational principles to be. Rational principles are principles about what rationality *permits*, *requires*, or *prohibits*. Our ordinary talk about rationality appears to presuppose that what rationality permits, requires, or prohibits of someone can depend on their particular circumstances. For example, given my circumstances as I write this, rationality permits that I now believe that I am currently sitting inside with my computer in front of me, rationality prohibits that I now believe that I am currently snorkelling at the Great Barrier Reef, and rationality requires that I now prefer receiving anaesthetic for my upcoming dental procedure to enduring the procedure without it.

Most of the main principles under discussion in this thesis, by contrast, are general principles about what rationality permits, requires, or prohibits for people in abstraction from their specific circumstances. These kinds of principle are best introduced by way of some common examples from the literature.

(Belief Consistency Principle) Rationality prohibits simultaneously believing that something is the case and believing that it is not the case.

(Intention Consistency Principle) Rationality prohibits simultaneously intending to do something and intending not to do it.

(Preference Transitivity Principle) Rationality prohibits simultaneously preferring something to something else, and in turn preferring the latter to something else again, but not preferring the first thing to the last thing.

(Instrumental Principle) Rationality prohibits failing to intend to do something that you believe to be a necessary means to an end that you intend to achieve.

These principles (if true) apply to all agents independently of their particular circumstances. What is prohibited in each case is thus prohibited *unconditionally*. It is controversial just how to make sense of the idea of a rational prohibition, permission, or requirement (Broome, 2013), but I take the following definitions to be adequate for current purposes. I will provide a more detailed account of these notions in chapter two.

Rationality (conditionally) permits that *S* φ $=_{df}$ it is possible that *S* φs and *S*'s φ-ing is rational.

Rationality (unconditionally) requires that *S* φ $=_{df}$ it is possible that *S* φs and *S*'s φ-ing is rational, and it is necessary that, if *S* does not φ, then *S*'s not φ-ing is not rational.

Rationality (unconditionally) prohibits that *S* φ $=_{df}$ it is possible that *S* φs, and it is necessary that, if *S* φs, then *S*'s φ-ing is irrational. [1]

[1] I will omit the 'conditionally' and 'unconditionally' qualifiers from here on.

The principles listed above are all principles that rule out certain combinations of *mental states*. This focus on mental states reflects the fact that 'rationality supervenes on the mind' (Broome, 2013). Exactly what this means depends on what is meant by 'rationality'. 'Rationality' may refer to a property that a *person* might have, or to a property that someone's having some *response* might have. Although it is plausible that each of these properties supervenes on the mind, the latter is the property that is relevant here, since the principles under discussion in this thesis concern the rationality of responses rather than people.

The thesis that the rationality of a response supervenes on the mind is, roughly, the thesis that whether someone's having a response is rational or irrational depends solely on facts about their mind. In other words, whether a response is rational or irrational cannot vary as long as the facts about the mind of the person who has that response are held fixed.

(Supervenience of Rationality on the Mind) For any possible worlds w_1 and w_2, agents S_1 and S_2, and response type φ, if S_1 φs at w_1, S_2 φs at w_2, and S_1 at w_1 is a psychological duplicate of S_2 at w_2, then S_1's φ-ing is (ir)rational at w_1, if and only if S_2's φ-ing is (ir)rational at w_2.

Given that rationality supervenes on the mind, it is not surprising that the rational principles that have been discussed so far are about combinations of mental states. One might wonder whether the fact that rationality supervenes on the mind rules out the possibility that there are also rational principles that constrain people's *actions* (Broome, 2013, p. 152), such as the following.

(Direct Instrumental Prohibition) For all *S*, *t*, φ, and ψ, rationality prohibits that:

> *S* intends at *t* to φ;
> *S* believes at *t* that, if they themselves were to ψ then, then because of that they themselves would not φ; and
> *S* intentionally ψs at *t*.

This is a very plausible rational principle that prohibits intentionally sabotaging one's own plans. One might worry that this principle is incompatible with the thesis that rationality supervenes on the mind, however, since whether someone violates it does

not depend solely on what their mind is like. Two psychological duplicates may differ with respect to whether they violate this principle because one of them successfully completes the relevant action, while the other is prevented from doing so by something outside their mind.

This worry rests on a misunderstanding of the thesis that rationality supervenes on the mind. It is based on the assumption that the supervenience of rationality on the mind entails that, if someone has some irrational response, then *the fact that they have that response* depends solely on what their mind is like. But the supervenience of rationality on the mind does not entail this. Rather, what it entails is that, if someone has some irrational response, then *the fact that that response is irrational* depends solely on what their mind is like. The fact that rationality supervenes on the mind is therefore perfectly compatible with the truth of rational principles like Direct Instrumental Prohibition that apply to actions as well as mental states.

Finally, rational principles specify the *coherent* or *intelligible* combinations of mental states and actions. They say which combinations of mental states and actions *make sense* and which ones do not. It is hard to know *what to make of* someone when they violate a rational principle. This chapter is not the place to go into detail about this notion of coherence; I will provide more detail in chapter two. For now, I simply assume that we all have an intuitive grasp on this notion. I also take no stand here on whether these principles are *normative*—that is, whether we necessarily have reason to conform to them (Broome, 2005; Kolodny, 2005; 2007; 2008; Hussain, 2007; Southwood, 2008; Way, 2010b; Kiesewetter, 2017).

0.3.1 Rational Basing Principles

The rational principles that have been used as examples so far are all *synchronic* rational principles. Synchronic rational principles specify how mental states and actions may be coherently related to each other at a single time. *Diachronic* rational principles, by contrast, specify how mental states and actions may be coherently related to each other across time. Rational *basing* principles are an important type of diachronic rational principle. They specify when a mental state or action can be coherently *based on* some other mental state(s), possibly from some previous time (Broome, 2013, §10.4; Way, 2011).

BASING is a folk-psychological concept that refers to a relation that can hold between mental states (and perhaps intentional actions). It is the relation that holds between the belief I just acquired that the post office is now closed and the beliefs that I had just prior to acquiring it that it is now Saturday afternoon, and that, if it is now Saturday afternoon, then the post office is now closed. I cannot provide a simple analysis of the concept of BASING; indeed, I doubt that there is one to be had. So, again, I will simply assume that we all have an intuitive grasp of what it is for one attitude or action to be based on another attitude.

The example just described is an instance of a kind of basing that rationality permits. Following Broome (2013, p. 191), we might try to formulate this basing permission as follows.

(Modus Ponens Permission) For all *S*, *p*, and *q*, rationality permits that:

> *S* believes at some time that *p*;
> *S* believes at some time that $p \supset q$;
> *S* believes at some time that *q*; and
> *S*'s belief that *q* is based on their belief that *p* and their belief that $p \supset q$.

This principle seems plausible at first sight. It says that, for any pair of propositions, one can rationally believe that the second proposition is true based on believing that the first is true, and believing that, if the first is true, then so is the second. The principle, however, is too strong to be a basing permission, and moreover is false. A rational basing permission should say merely that it can be rational for some attitudes to be based on others. The above formulation of Modus Ponens Permission, by contrast, says not only that it can be rational for some attitudes to be based on others, but also that it can be rational to hold those attitudes themselves. This feature of the principle not only makes it stronger than a basing permission should be, but also makes it false, as the following counterexample demonstrates.

It is *not* the case (for some arbitrary *p* and *q*) that rationality permits that:

> *S* believes at some time that $p \,\&\, \neg p$;
> *S* believes at some time that $(p \,\&\, \neg p) \supset q$;
> *S* believes at some time that *q*; and

S's belief that *q* is based on their belief that *p* & ¬*p* and their belief that (*p* & ¬*p*) ⊃ *q*.

This principle is true, since the state that it says is not rationally permitted involves believing a contradiction, and it is plausible that being in a state that involves believing a contradiction cannot be rational. It is therefore a counterexample to Modus Ponens Permission as formulated above. Note, however, that, since the fact that being in a state of this kind is not rational is explained by the irrationality of one of the premise beliefs, it remains a possibility that the conclusion belief's being based on the premise beliefs, by itself, might nonetheless be rational. Indeed, it seems reasonable to think that, in a case of this kind, the *basing* itself is rational even though some of the *beliefs* are not.

If Modus Ponens Permission is to be reformulated as a principle that permits that the conclusion belief be based on the premise beliefs regardless of whether those beliefs are themselves rational, it will need to be expressed as a *conditional* permission. In what follows I will employ the following notions of conditional permission, prohibition, and requirement.

Given that *p*, rationality permits that *S* φ $=_{df}$ it is possible that *p* and *S* φs, and *S*'s φ-ing is rational.

Given that *p*, rationality requires that *S* φ $=_{df}$ it is possible that *p* and *S* φs, and *S*'s φ-ing is rational, and it is necessary that, if *p* and *S* does not φ, then *S*'s not φ-ing is not rational.

Given that *p*, rationality prohibits that *S* φ $=_{df}$ it is possible that *p* and *S* φs, and it is necessary that, if *p* and *S* φs, then *S*'s φ-ing is irrational.

With this account of conditional permission in hand, we may reformulate Modus Ponens Permission as follows.

(Modus Ponens Permission) For all S, p, and q, given that:

> S believes at some time that p;
> S believes at some time that $p \supset q$; and
> S believes at some time that q;

Rationality permits that S's belief that q be based on their belief that p and their belief that $p \supset q$.

This principle says that, for any pair of propositions, it can be rational for a belief that the second proposition is true to be based on a belief that the first is true, and a belief that, if the first is true, then so is the second, regardless of whether the beliefs themselves are rational. This formulation of Modus Ponens Permission avoids the problems with the original formulation. In general, then, basing permissions have the following form.

(Basing Permission/Prohibition Schema) For all S, A_1, … , A_n, and B, given that:

> S holds A_1 at some time;
> …
> S holds A_n at some time; and
> S holds B at some time;

Rationality permits/prohibits that S's attitude B be based on S's attitude A_1, … , and S's attitude A_n.

For the rest of this thesis, however, for the sake of readability, I will express such principles in an abbreviated form as follows.

(Abbreviated Basing Permission/Prohibition Schema) For all S, A_1, … , A_n, and B, rationality permits/prohibits that S's attitude B be based on S's attitude A_1, … , and S's attitude A_n.

0.3.2 Rational Commitment

In the chapters to follow, I will make reference to a relation that holds between certain types of normative judgement that I will by stipulation call 'rational commitment'. Let us say that judging that some propositions are true rationally commits one to judging

that some other proposition is true just in case four conditions are met: (1) whenever rationality permits simultaneously judging that the former propositions are true, rationality permits also judging that the latter proposition is true; (2) rationality prohibits simultaneously judging that the former propositions are true and judging that the latter proposition is false; (3) rationality permits that judging that the latter proposition is true be based on judging that the former propositions are true; and (4) rationality prohibits that judging that the latter proposition is false be based on judging that the former propositions are true.

In what follows, I will make use of schemata of the following form, each instance of which says that the judgements expressed by the premise sentences rationally commit one to the judgement expressed by the conclusion sentence.

$$\frac{P_1; \ldots ; P_n}{C}$$

When I call a schema of this kind 'valid', I mean that, for each of its instances, the judgements expressed by the premise sentences do indeed rationally commit one to the judgement expressed by the conclusion sentence. When I call a schema of this kind 'invalid', I mean that, for some of its instances, the judgements expressed by the premise sentences do not in fact rationally commit one to the judgement expressed by the conclusion sentence. I will mark some of these schemata with a '×' to flag that they are invalid, and I will mark others with a '?' to flag that there is controversy about their validity.

0.4 Conclusion

Now that we have a sense of what rational principles are, we are in a position to examine the distinctive rational principles to which normative judgements are subject, and to provide an account of normative judgement and reflective agency that vindicates these principles. I will begin this task in the next chapter by examining the concepts MAY, MUST, and OUGHT.

1 The Fitting-Response Analysis of MAY and MUST and the Counterfactual Analysis of OUGHT

In this chapter I provide analyses of the concepts MAY, MUST, and OUGHT. I analyse MAY and MUST in terms of fitting-response concepts, such that to judge that something may be so is to judge that it is a fitting object of some relevant attitude or response, and to judge that something must be so is to judge that it may be so and it is not the case that it may not be so. I argue that OUGHT can be analysed as a counterfactual MUST, such that to judge that something ought to be so is to judge that, if it were not the case that it may be so, or not the case that it may not be so, then it would be the case that it must be so. I argue for these analyses by showing that they satisfy several important desiderata. First, they make sense of the intuitive differences in strength between MUST, OUGHT, and MAY. Secondly, they do not straightforwardly validate or invalidate any controversial rational principles. Thirdly, they make sense of the fact that obligations, recommendations, and permissions can be 'owned'. Fourthly, they make sense of the distinction between ordinary conditional obligations and anankastic conditionals. And fifthly, they make sense of the fact that OUGHT is expressed in several languages using an expression that is syntactically equivalent to the English expression 'would have to'.

1.1 Desiderata

I begin by identifying several desiderata for an analysis of the concepts MAY, MUST, and OUGHT, and along the way point out some views in the literature that fail to satisfy one or other of them. These views are mostly theses about the meaning of the words 'may', 'must', and 'ought', rather than about the concepts MAY, MUST, and OUGHT, but I will evaluate them with an eye to their suitability as theses about the contents of these concepts.

1.1.1 Desideratum 1: Validates the Differences in Strength between Must, Ought, and May

Intuitively, MUST is stronger than OUGHT, which is stronger than MAY. Roughly speaking, to judge that something must be so is to judge that it is *required*; to judge that something ought to be so is to judge that it is *recommended*; and to judge that

something may be so is to judge that it is *permitted*. These differences in strength can be summed up by the following schemata.

$$\frac{\text{It must be that } p.}{\text{It ought to be that } p.}$$

$$\frac{\text{It ought to be that } p.}{\text{It may be that } p.}$$

$$\frac{\text{It must be that } p.}{\text{It is not the case that it may be that } \neg p.}$$

It is tempting to think that MUST and MAY are duals, such that it is also the case that judging that it is not the case that something may be so rationally commits one to judging that it must be so. This would entail that it is never rational to believe that someone faces a 'prohibition dilemma'—a situation in which it is neither the case that they may φ nor may ¬φ, for some φ (Vallentyne, 1989). Since it is not obvious to me that it is never rational to believe that someone faces a prohibition dilemma, I do not assume that MUST and MAY are duals.

Thus, our first desideratum is that the analyses of MAY, MUST, and OUGHT validate these plausible schemata. This desideratum rules out the assumption, which is sometimes found in works of moral philosophy, that the concepts OUGHT and MUST are identical (see Silk, 2015 for examples). That these concepts are distinct is reflected in our ordinary usage of the words 'ought' and 'must'. Consider the following exchange, for example: 'Must I do it today?', 'No, but you ought to.' This exchange is only intelligible if 'ought' and 'must' express distinct concepts. This desideratum also rules out Silk's (2021) account of the difference in strength between 'ought' and 'must', since, according to his view, 'ought' does not imply 'may'.

1.1.2 Desideratum 2: Does not Straightforwardly Entail the Validity or Invalidity of any Controversial Principles

There are a number of controversial putative rational principles that apply to judgements involving the concepts MAY, MUST, and OUGHT. In so far as these

principles are genuinely controversial, an analysis of MAY, MUST, and OUGHT should account for this fact, and so should not straightforwardly entail that any them are valid or invalid. To clarify: I do not require of these analyses that, if they are correct, then the controversial schemata are neither valid nor invalid. Given how I defined these terms in the introduction, this is impossible. Rather, what I require is that, if we assume that these analyses are correct, then it remains controversial whether these schemata are valid; it should not be the case that, if we assume that these analyses are correct, then it is utterly mysterious why there was ever any debate about the validity of those schemata in the first place. The analyses should thus *make sense of* the fact that these schemata are controversial. In what follows, I will identify several such controversial principles.

1.1.2.1 Inheritance

First, consider some controversial 'inheritance' schemata.

Inheritance I: MUST

It must be that *p* and *q*.

It must be that *p*. (?)

Inheritance I: OUGHT

It ought to be that *p* and *q*.

It ought to be that *p*. (?)

The rational principles expressed by the instances of these schemata seem plausible, but are controversial because some believe that they are vulnerable to counterexamples, such as the case of Professor Procrastinate. Jackson and Pargetter (1986) describe the case as follows:

> Professor Procrastinate receives an invitation to review a book. He is the best person to do the review, has the time, and so on. The best thing that can happen is that he says yes, and then writes the review when the book arrives. However, suppose it is further the case that were Procrastinate to say yes, he would not in fact get around to writing the review. Not because of incapacity or outside interference or anything like that, but because he would keep on putting the task off. (This has been known to happen.) Thus, although the best that can happen is for Procrastinate to say yes and then write, and he can do exactly this, what would in fact happen were he to say yes is that he would not write the review. Moreover, we may

> suppose, this latter is the worst that can happen. It would lead to the book not being reviewed at all, or at least to a review being seriously delayed. (p. 235)

One might think that, regarding this case, rationality permits simultaneously judging that Professor Procrastinate must/ought to accept the invitation and review the book, and that it is *not* the case that Professor Procrastinate must/ought to accept the invitation.[2] If this is correct, then these two *prima facie* plausible inheritance schemata are in fact invalid. Things are complicated further, however, by the fact that it seems irrational to judge the *conjunction* of these two claims to be true—that is, to judge that it is the case that Professor Procrastinate must/ought to accept the invitation and review the book and not the case that Professor Procrastinate must/ought to accept the invitation (Von Fintel, 2012). I will thus treat these schemata as controversial.

Next are two inheritance schemata that many deontic logicians and linguists are inclined to accept as valid, but whose instances express rational principles that are *prima facie* implausible.

Inheritance II: MUST

It must be that *p*.

It must be that *p* or *q*. (?)

Inheritance II: OUGHT

It ought to be that *p*.

It ought to be that *p* or *q*. (?)

The rational principles expressed by the instances of these schemata appear implausible in view of Ross' (1944) paradox. Many are inclined to believe, for example, that believing that someone must/ought to mail a letter does not rationally commit one to believing that they must/ought to mail the letter *or burn the letter*. The

[2] In my discussion of these schemata, I will shift freely between locutions of the form 'it must/may/ought to be that *S* φs' and '*S* must/may/ought to φ' for the sake of readability. While many philosophers believe that these two types of locution are equivalent, some deny this (Schroeder, 2011). I have expressed the schemata using locutions of the form 'it must/may/ought to be that *p*'. As far as I can tell, rephrasing these schemata using the other type of locution would not affect their plausibility, and so I take it that shifting between these type of locution in the discussion of them is harmless.

reason such principles seem implausible it that they are in tension with the rational principles expressed by the instances of the following schema.

Free Choice Permission

It may be that *p* or *q*.

It may be that *p*. (?)

The rational principles expressed by the instances of this schema are *prima facie* plausible. For example, it seems plausible that judging that someone may go to the beach or go to the cinema rationally commits one to judging that they may go to the beach (and to judging that they may go to the cinema) (Kamp, 1974). In combination with Inheritance II, however, Free Choice Permission entails that, from the judgement that it must/ought to be that *p*, one may rationally derive the judgement that it may be that *q*, for any arbitrary *p* and *q*.

Assumption	(1)	It must/ought to be that *p*.
(1) × Inheritance II	(2)	It must/ought to be that *p* or *q*.
(2) × MUST/OUGHT entails MAY	(3)	It may be that *p* or *q*.
(3) × Free Choice Permission	(4)	It may be that *q*.

This is an extremely implausible result. It seems, then, that Inheritance II and Free Choice Permission cannot both be valid. Since the rational principles expressed by the instances of Free Choice Permission are *prima facie* plausible and the rational principles expressed by the instances of Inheritance II are *prima facie* implausible, one might think that we can therefore conclude that Inheritance II is invalid. Some argue, however, that the apparent plausibility of the rational principles expressed by the instances of Free Choice Permission is best explained by the pragmatic rules governing the *utterances* that express the relevant judgements, rather than a rational principle governing the judgements themselves, and are inclined to give it up instead (Von Fintel, 2012). So I will treat both Inheritance II and Free Choice Permission as controversial.

1.1.2.2 Agglomeration

Next are some controversial 'agglomeration' schemata.

Agglomeration: MUST

It must be that *p*; It must be that *q*.

It must be that *p* and *q*. (?)

Agglomeration: OUGHT

It ought to be that *p*; It ought to be that *q*.

It ought to be that *p* and *q*. (?)

The rational principles expressed by the instances of these schema are *prima facie* plausible, but are controversial because some believe that they are vulnerable to counterexamples, such as those put forward by Jackson (1985) and Lassiter (2017). Jackson describes his case as follows:

> Attila and Genghis are driving their chariots towards each other. If neither swerves, there will be a collision; if both swerve, there will be a worse collision (in a different place, of course); but if one swerves and the other does not, there will be no collision. Moreover if one swerves, the other will not because neither wants a collision. Unfortunately, it is also true to an even greater extent that neither wants to be 'chicken'; as a result what actually happens is that neither swerves and there is a collision. (p. 189)

One might think that, regarding this case, it is rational simultaneously to believe that it must/ought to be that Attila swerves, to believe that it must/ought to be that Genghis swerves, and to believe that it is *not* the case that it must/ought to be that Attila swerves and Genghis swerves. If this is correct, then these *prima facie* plausible agglomeration schemata are in fact invalid. I will thus treat these principles as controversial.

1.1.2.3 Detachment

Finally, consider some 'detachment' schemata. The best way to start thinking about these schemata is by looking at Chisholm's (1963) paradox.

1.1.2.4 Chisholm's Paradox

Chisholm asks us to consider the following four propositions.

1. Jones ought to go to assist his neighbours.
2. If Jones goes to assist his neighbours, then he ought to tell his neighbours that he is coming.
3. If Jones does not go to assist his neighbours, then he ought not to tell his neighbours that he is coming.
4. Jones will not go to the assistance of his neighbours.

It follows from certain *prima facie* plausible assumptions that, if one holds the judgements expressed by each of these sentences, then one is rationally committed to judging that Jones is subject to obligations that conflict with each other.

5. Jones ought to tell his neighbours that he is coming (given 1 and 2).

6. Jones ought not to tell his neighbours that he is coming (given 3 and 4).

The *prima facie* plausible assumptions that give rise to this paradox are instances of the following schemata.

Factual Detachment

If *S* does not φ, then *S* must/ought not ψ; *S* will not φ.

S must/ought not ψ.

Deontic Detachment

If *S* φs, then *S* must/ought to ψ; *S* must/ought to φ.

S must/ought to ψ.

1.1.2.5 Factual Detachment

Factual Detachment is invalid. Judgements about what someone must/ought to do are always judgements about what they must/ought to do *given some relevant set of circumstances*. If this is made explicit, then Factual Detachment may be restated as follows.

Factual Detachment

Given circumstances *C*, if *S* φs, then *S* must/ought to ψ; *S* will φ.

Given circumstances *C*, *S* must/ought to ψ. (×)

The problem with the instances of this schema is that they say that judging that someone is *conditionally* obligated to do something, given some relevant set of circumstances, and believing that that condition holds, rationally commits one to judging that they are *unconditionally* obligated to do it, given those *very same* circumstances. But this is implausible as long as those original circumstances leave open whether the relevant condition is met. Instead, one might think that the relevant conclusion to draw is that they have an unconditional obligation, given some *updated* set of circumstances—for example, that, given the original circumstances *and* that the relevant condition is met, they ought to do it. But really this is just another way of phrasing the first premise.

1.1.2.6 Deontic Detachment

It seems implausible that Deontic Detachment, as formulated below, is valid.

Deontic Detachment: MUST

Given circumstances *C*, if *S* φs, then *S* must ψ; Given circumstances *C*, *S* must φ.

Given circumstances *C*, *S* must ψ. (×)

Deontic Detachment: OUGHT

Given circumstances *C*, if *S* φs, then *S* ought to ψ; Given circumstances *C*, *S* ought to φ.

Given circumstances *C*, *S* ought to ψ. (×)

The instances of this schema say that judging that someone has a *conditional* obligation to do something, given some set of circumstances, and judging that someone has an unconditional obligation to ensure that the relevant condition is met, rationally commits one to judging that they have an *unconditional* obligation to do it, given those *very same* circumstances. But, again, as long as the original circumstances leave open whether the relevant condition will in fact be met, it is not clear that one may conclude that the person has an unconditional obligation, given those original circumstances. Again, the only apparent conclusion to be drawn is that, given the original circumstances *and* that the relevant condition is met, they ought to do it. But, again, this is really just another way of phrasing the first premise.

1.1.2.7 Axiom K

In the previous formulation of Deontic Detachment, the first premise asserted the existence of a conditional obligation. In a more plausible formulation of this schema, the first premise asserts the existence of an unconditional obligation to make a material conditional true. This schema is known in deontic logic as 'Axiom K' (Cresswell and Hughes, 1996). For ease of expression, I will rephrase the material conditional as a negated conjunction.

Axiom K: MUST

S must not (φ and not ψ); *S* must φ.

S must ψ. (?)

Axiom K: OUGHT

S ought not (φ and not ψ); *S* ought to φ.

S ought to ψ. (?)

The instances of these schemata do not involve any confusion between conditional and unconditional obligations. The rational principles expressed by the instances of these schemata have some *prima facie* appeal. Some believe that these principles are vulnerable to counterexamples, however. Consider the following case, derived from an example from Broome (2013, p. 120). Suppose that, since you have entered a marathon, you must/ought to exercise every day, and, in order to avoid depleting your day-to-day energy levels, you must/ought not exercise every day without also eating heartily every day. But suppose that, in fact, you will not exercise every day; you could, but you will not bother. One might think that, regarding this case, rationality permits simultaneously judging that:

> You must/ought to exercise every day;
> You must/ought not exercise every day without eating heartily every day; and
> It is *not* the case that you must/ought to eat heartily every day.

According to Axiom K, however, this combination of judgements is rationally prohibited. Thus, Axiom K is invalid.

One initial possible rejoinder to this objection is that is that it is difficult to make sense of the relevant judgements in this case without taking MUST/OUGHT to be relativised to

different sets of circumstances in some of them. In order to constitute a counterexample to Axiom K, the contents of the relevant normative judgements would have to take the following form.

1. Given circumstances *C*, you must/ought to exercise every day.

2. Given circumstances *C*, you must/ought not exercise every day without heartily every day.

3. It is not the case that, given circumstances *C*, you must/ought to eat heartily every day.

Given the way the case is described, it is natural to think that the relevant circumstances in 1 leave open whether you exercise hard. It is difficult, however, not to interpret the relevant conditions in 3 as ruling out that you exercise hard—after all, the very reason it is not the case that you must/ought to eat heartily every day is that you will not exercise hard. So it is natural to think that the correct judgements about this case take the following form instead.

1. Given circumstances *C*, you must/ought to exercise every day.

2. Given circumstances *C*, you must/ought not exercise every day without heartily every day.

3a. It is not the case that, given circumstances *C and that you will not exercise*, you must/ought to eat heartily every day.

As initially presented, then, this example does not appear to constitute a counterexample to Axiom K. We may be able to strengthen the example, however, by assuming that, although the relevant circumstances do not rule out your exercising every day, it is the case that, given those circumstances, it is very *unlikely* that you will exercise every day. In that case, it may be rational to believe that, given those circumstances, it is not the case that you must/ought to eat heartily every day. Thus it may well be rational to believe each of 1, 2, and 3. Here again, then, we have a schema whose instances express intuitively plausible rational principles that are subject to apparent counterexamples. I will thus treat Axiom K as controversial.

1.1.2.8 Summary

Based on the results of this discussion, I conclude that it is desirable that an analysis of the concepts MAY, MUST, and OUGHT not straightforwardly entail the validity or invalidity of Inheritance I, Inheritance II, Free Choice Permission, Agglomeration, or Axiom K. Many semantic accounts of 'may', 'must', and 'ought' fail to satisfy this desideratum. Some of these accounts are based on the standard Kratzerian (1977; 1981; 1991) framework for modal verbs in linguistics, and treat 'may', 'must', and 'ought' as quantifiers over sets of possible worlds (see Von Fintel and Iatridou, 2008 for a prominent example). Views of this kind straightforwardly entail the validity of all of these controversial schemata apart from Free Choice Permission. By contrast, some other accounts that analyse 'may', 'must', and 'ought' in terms of the concept of probability (Finlay, 2014; Lassiter, 2011; 2017), goodness (Jackson, 1985) or expected value (Lassiter, 2011; 2017; Cariani, 2009) straightforwardly entail the invalidity of some of these principles. In my view, the fact that these accounts uncontroversially validate or invalidate these intuitively uncertain principles makes them less plausible.

1.1.3 Desideratum 3: Makes Sense of the Idea that Obligations, Recommendations, and Permissions can be 'Owned'

Some authors have observed that obligations, recommendations, and permissions can sometimes be 'owned' by an agent. Consider the following two examples from Broome (2013, p. 12-13):

1. Alison ought to get a sun hat.

2. Alex ought to get a severe punishment.

Broome points out that 1 is naturally interpreted as saying that Alison is the one who is responsible for seeing to it that the requirement is satisfied, whereas the natural reading of 2 is one according to which Alex is not the one who responsible for seeing to it that the requirement is satisfied. As Broome puts it, Alison 'owns' the obligation in 1, whereas Alex does not 'own' the obligation in 2. Following Broome, I will, where necessary, indicate that an obligation or permission is owned by an agent in the following way.

S must/ought/may that *T* φs.

An analysis of MAY, MUST, and OUGHT should provide a way of making sense of the fact that obligations, recommendations, and permissions can sometimes be owned in this way.

1.1.4 Desideratum 4: Makes Sense of the Distinction Between Anankastic Conditionals and Ordinary Conditional Obligations

The next desideratum for the analysis of MAY, MUST, and OUGHT is that it provide a way of making sense of the distinction between ordinary conditional obligations and what linguists call 'anankastic conditionals' (Saebo, 1985). The following is an example of an anankastic conditional.

If Sam wants to go to Harlem, then Sam must take the 'A' train.

Intuitively, this proposition says that Sam's taking the 'A' train is required *in order to* satisfy their desire to go to Harlem. This is different to a conditional obligation of the following kind.

If it is raining, then Sam must take an umbrella.

The difference between these two examples is that the antecedent of the first conditional seems to say something about the ideal with respect to which the normative claim in the consequent is to be evaluated. It says roughly that, given that the salient ideal is, or includes, satisfying Sam's desire to go to Harlem, the best option is for Sam to take the 'A' train. The antecedent of the second conditional, by contrast, does not appear to have anything to do with the ideal with respect to which the normative claim in the consequent is to be evaluated. Rather, it says, roughly, that, in view of whatever the salient ideal is, the best option, on the supposition that it is raining, is that Sam takes an umbrella. An analysis of MAY, MUST, and OUGHT should provide a way of making sense of this distinction.

1.1.5 Desideratum 5: Makes Sense of the Fact that Ought is Expressed Linguistically in Many Languages Using the Word for Must Combined with Counterfactual Morphology

Von Fintel and Iatridou (2008) report that the concept OUGHT is expressed in many languages using an expression that is syntactically equivalent to the English expression 'would have to'. That is, it is expressed using the word that is used to express the concept MUST, in combination with counterfactual morphology. Although the focus in this chapter is on the *concept* OUGHT, rather than the *word* 'ought', it is nonetheless desirable that an analysis of this concept make sense of the fact that it is expressed in this way in various languages. Von Fintel and Iatridou (2008) and Silk (2021) provide two of the most prominent attempts in the literature to make sense of this linguistic fact.

1.2 The Fitting-Response Analysis of MAY and MUST

In this section, I will defend a fitting-response analysis of MAY and MUST, according to which, for each kind of modality—for example, deontic, epistemic, or evaluative—there is a particular fitting-response concept in terms of which MAY and MUST can be analysed. Thus, for each type of modality, *M*, there is some fitting-response concept, *A*-ABLE, such that:

(The Fitting-Response Analysis of MAY) For all *p*, to judge that it may$_M$ be that *p just is* to judge that it is *A*-able that *p*.

The fitting-response analysis of MUST is based on the intuitively plausible idea that something *must* be the case just in case it *may* be the case and it is not the case that it may *not* be the case. Thus, for each type of modality, *M*, and its characteristic fitting-response concept, *A*-ABLE:

(The Fitting-Response Analysis of MUST) For all *p*, to judge that it must$_M$ be that *p just is* to judge that it is *A*-able that *p* and it is not *A*-able that ¬*p*.

This analysis of MUST validates the intuitive difference in strength between MAY and MUST: judging that something must be so rationally commits one to judging that it may be so, and to judging that it is not the case that it may not be so. The analysis is

particularly plausible in the case of *deontic* modality, where MAY and MUST can be naturally analysed in terms of the fitting-response concept PERMISSIBLE.

(The Fitting-Response Analysis of the Deontic MAY) For all *p*, to judge that it may be that *p just is* to judge that it is permissible that *p*.

(The Fitting-Response Analysis of the Deontic MUST) For all *p*, to judge that it must be that *p just is* to judge that it is permissible that *p* and it is not permissible that ¬*p*.

The fitting-response analysis is also plausible in the case of the *epistemic* MAY and MUST, which can be naturally analysed in terms of the concept CONCEIVABLE.

(The Fitting-Response Analysis of the Epistemic MAY) For all *p*, to judge that it may be that *p just is* to judge that it is conceivable that *p*.

(The Fitting-Response Analysis of the Epistemic MUST) For all *w* and *p*, to judge that it must be that *p just is* to judge that it is conceivable that *p* and it is not conceivable that ¬*p*.

The fitting-response analysis can also make sense of the *evaluative* MAY and MUST, which can be analysed in terms of the concept DESIRABLE.

(The Fitting-Response Analysis of the Evaluative MAY) For all *p*, to judge that it may be that *p just is* to judge that, it is desirable that *p*.

(The Fitting-Response Analysis of the Evaluative MUST) For all *w* and *p*, to judge that it must be that *p just is* to judge that it is desirable that *p* and it is not desirable that ¬*p*.

The existence of an evaluative *OUGHT* has often been proposed in the literature (see, for example, Humberstone, 1971; Wedgwood, 2007; 2009; Gilabert, 2009; Schroeder, 2011; Southwood, 2016a). It is the kind of concept that is at work when I judge of Larry, who is a good man but terribly down on his luck, that he ought to win the lottery (Schroeder, 2011). Examples of the evaluative MUST also seem to exist, though they are somewhat less natural. I might judge, for example, that it mustn't rain tomorrow, since that would ruin the huge outdoor party I am planning. The evaluative MAY seems to be less common, and examples are difficult to imagine. Perhaps,

having been forced to leave my brand new expensive car out on the street overnight, I might desperately judge that it may rain, just as long as it doesn't hail.

One might worry that my treatment of the evaluative MAY is inadequate based on the following argument.

1. If there is indeed such a thing as the evaluative MAY, then it must be possible, for some *p*, that it evaluatively may be that *p* and at the same time evaluatively may be that ¬*p*.
2. If my analysis of the evaluative MAY is correct, then it is possible, for some *p*, that it evaluatively may be that *p* and at the same time evaluatively may be that ¬*p*, only if it is possible that it be desirable that *p* and at the same time desirable that ¬*p*.
3. But there is no *p* such that it is possible that it be desirable that *p* and at the same time desirable that ¬*p*.
4. Therefore, my analysis of the evaluative MAY is incorrect, or there is no such thing as the evaluative MAY.[3]

It seems to me that this argument can easily be resisted by denying the third premise. It is possible for it to be simultaneously desirable that something be so, and desirable that it not be so. Examples abound. Suppose my friend has invited me to go to a party now. I am torn: it is desirable that I to go to the party, since I would have a fun time, but it is also desirable that I not go to the party, since it is nice and warm at home and I am already in my pyjamas. This premise does seem plausible if we assume that it is desirable that something be so only if it is more desirable that it be so than that it not be so, but this assumption seems false. In the above example, I might conclude upon reflection that, although each of the alternatives is desirable, not going to the party is the more desirable alternative. This does not seem incoherent.

3 I am grateful to an anonymous examiner for raising this objection.

Perhaps the motivation behind the third premise has something to do with the thought that something is desirable just in case it is *fitting* to desire it, where a response's being fitting is a matter of its being *required* or *correct*, in some sense. So, if, for some *p*, it is at the same time desirable that *p* and desirable that ¬*p*, then desiring that *p* and desiring that ¬*p* are each required or correct. This, in turn, might commit me to simultaneously desiring that *p* and desiring that ¬*p*, which is impossible or irrational. The response to this thought is that it is not impossible or necessarily irrational to desire that something be so and desire that it not be so. In the example above, my decision whether to go to the party is a difficult one precisely because I both want to go to the party, and want to not go to the party, and these seem to be the correct, or required, responses to the situation. In any case, one might wonder how I am understanding the concept of fittingness in this chapter. As far as I can tell, nothing I say in this chapter depends on exactly how fittingness is construed. I discuss fittingness in chapter three. To anticipate, according to the view that I develop there, to judge that something is desirable (or fitting to desire) is (roughly) to judge that desiring it is (rationally) *permitted*, rather than *required* or *correct* in some other sense.

According to this analysis, then, judgements about what may or must be so, in the deontic, epistemic, and evaluative senses, are judgements about what is fit to be *permitted*, *conceived*, and *desired*, respectively. The analysis also opens up opportunities to make some more subtle distinctions between different kinds of modality. Some of the modalities I am about to consider may seem somewhat unnatural. I am not committing myself here to the view that all of these modalities are in common use. In so far any of them are in common use, however, it is an advantage of the fitting-attitude analysis that it can easily make sense of them. For example, consider *prescriptive* modality. This is the kind of modality that is relevant in the context of *advice* (Southwood, 2016a). It is natural to think that the prescriptive MAY and MUST might be analysable in terms of the concept ADVISABLE, such that to judge that something may be so *just is* to judge that it is advisable that it be so, and to judge that something must be the case *just is* to judge that it is advisable that it be so, and not advisable that it not be so.

Similarly, consider *practical* modality. This is the kind of modality that is relevant in the context of *practical deliberation* (Williams, 1981, p. 118; Southwood, 2016a). Practical deliberation is a process in which one seeks to make a decision about some outcome which one takes oneself to be capable of *intentionally* bringing about or preventing. There is a practical sense of 'permits' such that it makes sense to analyse the practical MAY and MUST in terms of PERMISSIBLE. Someone permits that some state of affairs obtain, in this practical sense, roughly when it obtains either because they intentionally failed to prevent it, or because they intentionally brought it about. On this understanding of permitting, for example, I permit myself to gain weight if I gain weight intentionally or if I gain weight because I failed to prevent it when it was in my power to do so. I will rely on this practical conception of permitting in chapter four to explain the enkratic principles that apply to judgements involving the concepts MAY, MUST, and OUGHT. I will not try to spell out this concept's necessary and sufficient conditions here, but I will highlight two sufficient conditions that will prove to be important later. For all *S* and *p*, it is plausible that the following are *a priori* truths about this practical kind of permitting.

(Failing to Prevent) If its being the case that *p* causally depends on *S*'s intending that *p*, and *S* does not intend that *p*, then *S* permits that ¬*p*.

(Bringing about) For all *S* and *p*, if its being the case that *p* causally depends on *S*'s intending that *p*, and *S* intends that *p*, then *S* permits that *p*.

According to this analysis of the practical MAY and MUST, then, to judge that something may be the case is, roughly, to judge that it is fitting to bring it about or fail to prevent it from occurring, and to judge that something must be the case is, roughly, to judge that it is fitting to bring it about or fail to prevent it from occurring, and it is not fitting to bring it about that it not be so or fail to prevent it from not occurring.

Next, consider *hypological* modality. This is the kind of modality that is relevant in the context of *criticism* (Zimmerman, 2006; Southwood, 2016a). If there is such a thing as the hypological MAY and MUST, then it is plausible that these concepts can be analysed in terms of the fittingness of a certain kind of *acceptance*. To accept that someone have some response, in this sense, involves being disposed not to criticise

them for having that response. For example, for me to accept, in this sense, that you will borrow my car involves my not being disposed to criticise, blame, or punish you for borrowing my car, should I learn that you have done so. Thus, to judge that something may be the case, on this analysis, *just is* to judge that it is acceptable that it be so, and to judge that something must be the case *just is* to judge that it is acceptable that it be so and not acceptable that it not be so.

Finally, consider *rational* modality. The kind of modality I have in mind here is the kind of modality that is relevant in the context of *forming attitudes*. It is natural to think that this kind of modality is analysable in terms of the concept REASONABLE. To judge that someone may have some attitude, on this analysis, *just is* to judge that it is reasonable for them to have that attitude, and to judge that someone must have some attitude *just is* to judge that it is reasonable for them to have that attitude and not reasonable for them not to have that attitude. This is consistent with the fitting-response analysis of MAY and MUST to the extent that REASONABLE can be interpreted as a fitting-response concept. We might think of something's being reasonable in this sense as its being fitting for it to be 'reasoned' in the sense of being based on reasoning. For example, its being reasonable that I believe in extra-terrestrial life, on this view, would consist in its being fitting that I believe in extra-terrestrial life based on reasoning.

This would provide an attractive *constructivist* account of rational modality. Southwood (2018) argues that constructivism with respect to some normative domain is best understood as the view that the facts about that domain are ultimately explained by facts about *reasoning* that is *correct*, in some sense. For example, Kant (1959) (at least on one way of interpreting him) and Korsgaard (1996; 2008; 2009) defend a constructivist account of normativity according to which rationality requires reasoning in accordance with the Categorical and Hypothetical Imperatives, and all normative facts are ultimately grounded in facts about reasoning that is correct by these lights. Another example is provided by Street (2008; 2010), according to whom the rational requirements governing practical reasoning are constitutive of the attitude of normative judgement itself, and facts about practical reasons are ultimately grounded in facts about reasoning that is correct by these lights. According to the account of rational modality I have just presented, facts about what may or must be

the case, in the rational sense, are ultimately understood in terms of facts about reasoning that is correct in the sense of being *fitting*. For this reason, this account of rational modality can be appropriately categorised as constructivist.

One might worry that some of these analyses cannot be correct because they take the non-gradable concepts MAY and MUST to be analysable in terms of gradable concepts such as DESIRABLE, ADVISABLE, and REASONABLE. The worry is that, while it is intelligible to judge that something is more or less desirable, advisable, or reasonable than something else is, it does not make sense to judge that something is 'more or less may' than something else is. The proponent of the fitting-response analysis has a straightforward response here. Although according to these analyses the non-gradable MAY and MUST are indeed analysable in terms of gradable concepts, they are analysable in terms of the *positive* forms of these gradable concepts. Consider the gradable concept EXCITED, for example. EXCITED has a comparative form, such that it is intelligible to judge, for example, that Sam is more excited than Tom is. EXCITED also has a positive form however, such that it is intelligible to judge simply that Sam is excited. This judgement is plausibly equivalent to something like the judgement that Sam's level of excitement is at least as great as some relevant threshold. This positive form of EXCITED does not itself admit of degrees—Sam's level of excitement either is at least as great as the relevant threshold, or it is not. Thus, since according to the analyses presented above MAY and MUST are analysed in terms of the positive forms of the concepts DESIRABLE, ADVISABLE, and REASONABLE, the fact that these concepts are gradable while MAY and MUST are not does not pose a problem.

1.2.1 Controversial Principles

Does this analysis straightforwardly validate or invalidate any of the controversial principles identified above? It does not, for two reasons. First, according to this analysis, MAY and MUST are analysed in terms of different fitting-response concepts depending on the relevant type of modality. These fitting-response concepts seem to vary in their inferential properties, such that, while some of them have inferential properties that would make some of these principles valid, others do not. Secondly, some of the inferential properties of these fitting-response concepts are themselves controversial.

Consider Inheritance I and II, for example. It seems plausible that we could establish the validity of these principles, given the fitting-response analysis of MAY and MUST, if the following schemata were valid.[4]

Inheritance I: *A*-ABLE

It is *A*-able that *p* and *q*.

It is *A*-able that *p*.

Inheritance II: *A*-ABLE

It is *A*-able that *p*.

It is *A*-able that *p* or *q*.

Are these schemata valid? They seem fairly plausible when applied to the concept CONCEIVABLE. Judging that it is conceivable that *p* and *q* does seem rationally to commit one to judging that it is conceivable that *p*; and judging that it is conceivable that *p* seems rationally to commit one to judging that it is conceivable that *p* or *q*. On the other hand, these schemata seem implausible when applied to the concept DESIRABLE. The Professor Procrastinate case provides a good example: it seems rational with respect to this case simultaneously to judge that it is desirable that Professor Procrastinate accept and write and to judge that it is not desirable that he accept (because he will not write). With regard to the concept PERMISSIBLE, however, things are not so clear. Does judging that it is permissible that *p* and *q* rationally commit one to judging that it is permissible that *p*? And does judging that it is

---

[4] Inheritance I: MUST: Suppose, for some arbitrary *p* and *q*, that it must be that *p* and *q*. Then, given the fitting-response analysis, (i) it is *A*-able that *p* and *q*, and it is not *A*-able that ¬(*p* and *q*), which is so just in case (ii) it is not *A*-able that ¬*p* or ¬*q*. From i it follows, given Inheritance I: *A*-ABLE, that (iii) it is *A*-able that *p*. Contraposing Inheritance II: *A*-ABLE, it follows from ii that (iv) it is not *A*-able that ¬*p*. Given the fitting-response analysis, it follows from iii and iv that it must be that *p*.

Inheritance II: MUST: Suppose, for some arbitrary *p* and *q*, that it must be that *p*. Then, given the fitting-response analysis, (i) it is *A*-able that *p*, and (ii) it is not *A*-able that ¬*p*. From i it follows, given Inheritance II: *A*-ABLE that (iii) it is *A*-able that *p* or *q*. Contraposing Inheritance I: *A*-ABLE, it follows from ii that it is not *A*-able that ¬*p* and ¬*q*, which is holds just in case (iv) it is not *A*-able that ¬(*p* or *q*). Given the fitting-response analysis, it follows from iii and iv that it must be that *p* or *q*.

permissible that *p* rationally commit one to judging that it is permissible that *p* or *q*? These principles seem no less controversial than the original Inheritance I and II.

It is plausible that these differences boil down to differences between the concepts CONCEIVE, DESIRE, and PERMIT. It seems to be an *a priori* truth that, if someone conceives of its being the case that *p* and *q*, then they thereby conceive of its being the case that *p*; and, similarly, if someone conceives of its being the case that *p*, then they thereby conceive of its being the case that *p* or *q*. But there do not seem to be any parallel *a priori* truths about desiring: desiring that *p* and *q* does not entail desiring that *p*—I could desire that Professor Procrastinate accept and write without desiring that he accept, for example—and neither does desiring that *p* entail desiring that *p* or *q*. And, again, with respect to PERMIT, things are not so clear. Does permitting that *p* and *q* entail permitting that *p*? Does permitting that *p* entail permitting that *p* or *q*? The answers are not obvious. It makes sense, then, that these inheritance principles are controversial when applied to the concept PERMISSIBLE.

Now consider Free Choice Permission. On this analysis, this schema is equivalent to the following.

Free Choice Permission: *A*-ABLE

It is *A*-able that *p* or *q*.

It is *A*-able that *p*. (?)

Whichever fitting-response concept we focus on, this schema seems no less controversial than the original Free Choice Permission schema.

Next, consider Axiom K: MUST. It seems plausible that we could establish that this schema is valid if something like the following schema were valid.[5]

[5] Assume, for some arbitrary *p* and *q*, that it must be that *p* and it must be that ¬(*p* and ¬*q*). Given the fitting-response analysis, it follows that (i) it is *A*-able that *p*, (ii) it is not *A*-able that ¬*p*, (iii) it is *A*-able that ¬(*p* and ¬*q*), and (iv) it is not *A*-able that *p* and ¬*q*. Given Agglomeration: *A*-ABLE, it follows from i, ii, and iii that it is *A*-able that *p* and ¬(*p* and ¬*q*), which is to say (v) it is *A*-able that *q*. Suppose it is *A*-able that ¬*q*. Then it follows from that and iii and iv, given Agglomeration: *A*-ABLE, that it is *A*-able that ¬*q* and ¬(*p* and ¬*q*), which is to say that it is *A*-able that ¬*p*, which contradicts ii. Thus, (vi) it is not *A*-able that ¬*q*. Given the fitting-response analysis, it follows from v and vi that it must be that *q*.

Agglomeration: *A*-ABLE

It is *A*-able that *p*; It is not *A*-able that ¬*p*; It is *A*-able that *q*.

It is *A*-able that *p* and *q*.

Is this schema valid? It is fairly plausible when applied to the concept CONCEIVABLE. If it is conceivable that *p*, inconceivable that ¬*p*, and conceivable that *q*, then it does seem to follow that it is conceivable that *p* and *q*, since any conceivable scenario at which *q* is the case would have to be one at which *p* is also the case. The schema is also quite plausible when applied to the concept PERMISSIBLE. If it is permissible that *p*, not permissible that ¬*p*, and permissible that *q*, then it does seem to follow that it is permissible that *p* and *q*. It is less plausible when applied to the concept DESIRABLE, however. If it is desirable that *p*, not desirable that ¬*p*, and desirable that *q*, it does not seem to follow that it is desirable that *p* and *q*.

Finally, consider Agglomeration: MUST. It is plausible that we could establish that this schema is valid if Agglomeration: *A*-ABLE and the following schema were valid.[6]

Distribution

It is *A*-able that *p* or *q*.

It is *A*-able that *p* or it is *A*-able that *q*.

Again, this schema seems plausible when applied to CONCEIVABLE and PERMISSIBLE, but not when applied to DESIRABLE. I thus conclude that it is sufficiently dubious whether the fitting-response analysis of MAY and MUST validates the controversial schemata presented above that we can consider our second desideratum satisfied.

[6] MUST: Assume, for some arbitrary *p* and *q*, that it must be that *p* and it must be that *q*. Given the fitting-response analysis, this entails that (i) it is *A*-able that *p*, (ii) it is not *A*-able that ¬*p*, (iii) it is *A*-able that *q*, and (iv) it is not *A*-able that ¬*q*. Given Agglomeration: *A*-ABLE, it follows from i, ii, and iii that (v) it is *A*-able that *p* and *q*. Suppose it is *A*-able that ¬*p* or ¬*q*. Then, given Distribution, it follows that (vi) it is *A*-able that ¬*p* or (vii) it is *A*-able that ¬*q*. Since vi contradicts ii and vii contradicts iv, it follows that it is not *A*-able that ¬*p* or ¬*q*, which is to say that (viii) it is not *A*-able that ¬(*p* and *q*). Given the fitting-response analysis, v and viii entail that it must be that *p* and *q*.

1.2.2 Ownership of Obligations and Permissions

The fitting-response analysis of MAY and MUST implies that these concepts have more structure than they might otherwise be thought to have, since fitting-response concepts are themselves plausibly composed out of the concept of fittingness and the concept of a particular kind of response. One advantage of this is that it provides a way of making sense of the idea that permissions can be 'owned'. The extra structure provides a way of making sense of the idea that the responsibility for satisfying the relevant permission or prohibition falls on a particular person.

Consider prescriptive modality, for example. To judge that something may be so in the prescriptive sense is to judge that it is advisable, in the sense of its being fitting (for a suitably situated advisor) to advise *some relevant advisee* that it be so. In this case, it is natural to think of the relevant advisee as the one who is responsible for the advisable state of affairs' obtaining. Next, consider practical modality. To judge that something may be the case in the practical sense is to judge that it is permissible, in the sense of its being fitting *for some relevant agent* to bring it about or allow it to happen. In this case, it is natural to think of that agent as the one who is responsible for that state of affairs' obtaining. Similarly, to judge that someone may have some attitude in the rational sense is to judge that it is reasonable, in the sense of its being fitting they have that attitude, based on *their* reasoning. In this case, it is natural to think of that agent as the one who is responsible for having the attitude. Finally, consider hypological modality. To judge that something may be the case in the hypological sense is to judge that it is acceptable, in the sense of its being fitting to be disposed not to criticise *some relevant agent* for its being the case. In this case, it is natural to think of the person who is the object of the would-be criticism as the one who is responsible for that state of affairs' obtaining. We may thus consider our third desideratum satisfied as well.

1.2.3 Anankastic Conditionals

Another advantage of the extra structure that is provided by this analysis is that it allows for two different ways of understanding conditional normative judgements. To use the deontic MUST as an example, in ascribing to someone the judgement that, if

p, then it must be the case that *q*, one might be ascribing either of the following judgements to them.

> The judgement that it is fitting, conditional on *p*, to permit that *q*, and not fitting, conditional on *p*, to permit that ¬*q*.
>
> The judgement that it is fitting to permit, conditional on *p*, that *q*, and not fitting to permit, conditional on *p*, that ¬*q*.

Whereas the first judgement is about the *conditional fittingness* of permitting that *q* and ¬*q*, the second judgement is about the fittingness of *conditionally permitting q* and ¬*q*. That there are these two ways of interpreting this kind of judgement is a benefit of this view, since it allows us to make sense of the distinction between ordinary conditional obligations and anankastic conditionals. Judgements about anankastic conditionals are judgements about the fittingness—conditional on the antecedent's being true—of permitting that some state of affairs obtain. Thus, to judge that, if Sam wants to go to Harlem, then she must take the 'A' train *just is* to judge that it is fitting for Sam—conditional on her wanting to go to Harlem—to permit that she take the 'A' train, and that it is not fitting for Sam—conditional on her wanting to go to Harlem—to permit that she not take the 'A' train. Judgements about non-anankastic conditional obligations, by contrast, are judgements about the fittingness of permitting—conditional on the antecedent's being true—that the relevant state of affairs obtain. Thus, to judge that, if it is raining, then Sam must take an umbrella *just is* to judge that it is fitting for Sam to permit—conditional on its raining—that she take an umbrella, and it is not fitting for Sam to permit—conditional on its raining—that she not take an umbrella.

1.2.4 Summary

The fitting-response analysis of MAY and MUST thus satisfies the relevant desiderata set out above. It validates the differences in strength between MUST and MAY; it does not straightforwardly validate any controversial principles; it makes sense of the idea that obligations and permissions can be 'owned'; and it makes sense of the distinction between ordinary conditional obligations and anankastic conditionals. In the next section, I provide an analysis of OUGHT in terms of MAY and MUST.

1.3 The Counterfactual Analysis of OUGHT

I have just provided an analysis of MAY and MUST in terms of fitting-response concepts. In this section, I provide an account of OUGHT that takes our fifth desideratum seriously, and analyses OUGHT as a counterfactual MUST. As presented, this simple counterfactual analysis looks like it is vulnerable to counterexamples. I will show that the way to avoid these counterexamples is to specify that the counterfactual is to be interpreted relative to a particular kind of closeness relation between worlds—namely one that prioritises similarity with respect to relevant modal facts. While the resulting analysis is elegant and has a number of attractive features, it fails to satisfy one of the relevant desiderata. I conclude by considering a variation on it which promises to satisfy all of the desiderata but has problems of its own. Addressing the problems with the counterfactual analysis of OUGHT is thus an avenue for future research. There is a natural way of extending the counterfactual analysis of OUGHT so that it also accounts for the concept REASON, but I leave the discussion of that possibility to an appendix so as not to break up the flow of the overall argument I am developing, and because of its similarity to the analysis under discussion here.

1.3.1 The Simple Counterfactual Analysis

The simple counterfactual account of OUGHT is as follows.

(Simple Counterfactual Analysis of OUGHT) For all *p*, to judge that it ought to be that *p* *just is* to judge that, if it were not the case that it may be that *p*, or not the case that it may not be that *p*, then it would be the case that it must be that *p*.

In other words, according to this account, to judge that something ought to be so is to judge that, if it were not *optional*, then it would be *required*. If we assume that counterfactuals can be analysed in terms of the closeness of worlds, then we can also give this view the following rough gloss: something ought to be so just in case it is closer to being required than it is to being not permitted.

This analysis meets our first desideratum. According to this analysis, since OUGHT is analysed in terms of MUST, MUST is, in a sense, more fundamental than OUGHT is. That MUST is in this sense more fundamental than OUGHT is a natural conclusion to draw from the fact that, in many languages other than English, OUGHT is expressed

using the word for MUST with counterfactual morphology. It is plausible that the analysis meets the third and fourth of our desiderata. Since, according to this analysis, OUGHT is simply a counterfactual MUST, we may assume that, if we can make sense of the idea that obligations that are ascribed using the concept MUST can be owned, then we can make sense of the fact that recommendations that are ascribed using the concept OUGHT can be owned in the same way. Similarly, if we can make sense of the distinction between ordinary conditional obligations and anankastic conditionals involving the concept MUST, we may assume that we can make sense of this distinction in the same kind of way when the conditional judgements involve the concept OUGHT rather than MUST.

1.3.1.1 Differences in Strength

This account arguably also meets our first desideratum: according to this account, OUGHT is stronger than MAY, and MUST is stronger than OUGHT.

OUGHT is stronger than MAY

If it were not the case that it may be that *p* or not the case that it may not be that *p*, then it would be the case that it must be that *p*.

It may be that *p*.

To see that this schema is valid, suppose for *reductio* that, for some arbitrary *p*, the relevant counterfactual is true, but it is not the case that it may be that *p*. It follows that the antecedent of the counterfactual is true, but, since ¬MAY implies ¬MUST, the consequent is false. Thus the counterfactual itself is false, contrary to our assumption. This proof presupposes that *modus ponens* is valid for counterfactual conditionals—that is, that, for all *p* and *q*, $(p > q) \supset (p \supset q)$—but this principle is not controversial.

MUST is stronger than OUGHT

It must be that *p*.

If it were not the case that it may be that *p* or not the case that it may not be that *p*, then it would be the case that it must be that *p*.

One might argue that this schema is valid based on the following reasoning. Suppose, for some arbitrary *p*, that it must be that *p*. Then, since MUST implies

¬MAY¬, it is not the case that it may not be that *p*, and so the antecedent of the relevant counterfactual is true. Since both the antecedent and consequent of the counterfactual are true, it follows that the counterfactual itself is true. This proof presupposes, however, that Conjunction Conditionalization is true—that is, that, for all *p* and *q*, (*p* & *q*) ⊃ (*p* > *q*). Conjunction Conditionalization is a controversial principle. A number philosophers have argued that it is counterintuitive and subject to counterexamples (for example, Bennett, 1974; 2003, §92; Fine, 1975; McDermott, 2007; Nozick, 1981, pp. 680-681). Walters and Williams (2013), however, have mounted a strong argument in favour of Conjunction Conditionalization by showing that it is entailed by other plausible logical principles, in conjunction with a plausible thesis about 'irrelevant semifactuals'—subjunctive conditionals whose consequent is true and whose antecedent is irrelevant to the truth of its consequent. Since defending Conjunction Conditionalization is beyond the scope of this thesis, however, I will simply accept at this point that, in so far as Conjunction Conditionalization is controversial, this counts against the counterfactual analysis of OUGHT.

Conjunction Conditionalization is valid on Lewis' (1973) influential analysis of counterfactuals due to the fact that he takes the similarity relation to be *strongly centred*, such that no other world is *as similar* to a world as that world is to itself. This makes sense if the similarity of worlds is construed as some kind of *overall* similarity. Lewis considers modifying his analysis by permitting the similarity relation instead to be merely *weakly* centred, such that no other world is *more similar* to a world than that world is to itself (p. 29). This permits other worlds to be as similar to a world as it is to itself. This makes sense if the similarity of worlds is taken to consist in similarity only *in certain respects*, and would render Conjunction Conditionalization invalid.

It is worth noting that, on this modified account, for all *p* and *q*, (*p* & *q*) ⊃ (*p* > *q*) is still true as long as similarity with respect to whether *p* and *q* are true is part of what determines how similar two worlds are. If, as I suggest below, the similarity relation that is relevant in the counterfactual analysis of ought is one according to which similarity between worlds consists at least partly in similarity with respect to facts about what *may* be the case at those worlds, then, on this modified Lewisian analysis, my reliance on Conjunction Conditionalization in showing that OUGHT implies MUST would still be legitimate. Since assessing the merits of this restricted

version of Conjunction Conditionalization would take us too far afield here, however, I will not pursue this line of thought any further.

I should also note here that, as stated, the counterfactual analysis of OUGHT is only plausible if the principle of Simplifying of Disjunctive Antecedents (SDA) is false. SDA is the thesis that, for all p, q, and r, $((p \vee q) > r) \supset (p > r)$. If the counterfactual analysis of OUGHT is correct and SDA holds, then every judgement to the effect that something ought to be so is false. According to the counterfactual analysis of OUGHT, the judgement that something ought to be so is equivalent to a judgement of the form $(\neg p \vee \neg q) > (p \,\&\, \neg q)$ (at least as long as the analysis of MUST in terms of MAY is correct). If SDA is true, then this entails $\neg p > (p \,\&\, \neg q)$, which in turn plausibly entails $\neg p > p$. Since this is cannot be true, it follows that the original judgement must be false. SDA is controversial; although it seems intuitively plausible and some philosophers have explicitly endorsed it (see, for example,Nute, 1975; Studtmann, 2003, pp. 56, n. 52), it is in tension with some other very plausible logical principles—for example, the principle of the Interchange of Logical Equivalents (Ellis et al., 1977; Lewis, 1977). Again, since establishing the falsity of SDA is beyond the scope of this thesis, I will simply accept here that, to the extent that SDA is plausible, this counts against the counterfactual analysis of OUGHT I have presented.

1.3.1.2 Counterexamples and Closeness

One apparent problem with this analysis is that it seems vulnerable to counterexamples. For example, suppose that I have won the lottery, and that because of that I now have enough money to buy a house. Suppose that I *ought* to use the money to buy a house, although I do not *have to*—investing the money in the stock market is also a permissible, though less ideal, option. It seems coherent regarding this case to deny that, if it were not the case that I may buy a house, or not the case that I may not buy a house, then it would be the case that I must buy a house. After all, given how easily I could have failed to win the lottery, it seems plausible that the closest worlds at which either I am not permitted to buy a house or not permitted not to buy a house are worlds at which I did not win the lottery and am therefore not permitted to buy a house, since I cannot not afford it. Thus we have a case in which something ought to be so, but it is not the case that, if it were not the

case that it may be so, or not the case that it may not be so, then it would be the case that it must so, *contra* the simple counterfactual analysis of OUGHT.

I suggest that the best way for the advocate of the simple counterfactual analysis to respond here is to draw on the idea that counterfactual conditionals can be interpreted in different ways in different contexts. While the advocate of the simple counterfactual analysis need not be committed to any particular analysis of counterfactuals, to illustrate this point, it will be instructive here to think of counterfactuals in terms of worlds that are ranked by some relevant similarity relation, as in the familiar Lewis-Stalnaker account of counterfactuals (Lewis, 1973; Stalnaker, 1968). Lewis (1973) discusses the context-sensitivity of counterfactuals in addressing a pair of counterfactuals from Quine (1960, p. 222):

> If Caesar had been in command [in Korea] he would have used the atomic bomb.
>
> If Caesar had been in command [in Korea] he would have used catapults.

According to Lewis, which of these counterfactuals is true depends on which priorities or interests are most salient at the context of utterance.

> In one context, we may attach great importance to similarities and differences in respect of Ceasar's character and in respect of regularities concerning the knowledge of weapons common to commanders in Korea. In another context we may attach less importance to these similarities and differences in respect of Caesar's own knowledge of weapons. The first context resolves the vagueness of comparative similarity in such a way that some worlds with a modernized Caesar in command come out closer to our world than any with an unmodernised Caesar. It thereby makes the first counterfactual true. The second context resolves the vagueness in the opposite direction, making the second counterfactual true. Other contexts might resolve the vagueness in other ways. A third context, for instance, might produce a tie between the closest worlds with unmodernised Caesars and the closest worlds with modernized Caesars. That context makes both counterfactuals false. (p. 67)

My suggestion, then, is that, assuming a Lewis-Stalnaker-style analysis of counterfactuals, the advocate of the simple counterfactual analysis of OUGHT can avoid counterexamples of the kind presented above by clarifying the nature of the relevant similarity relation. Regarding to the example above, I suggested that, since my winning the lottery was a complete fluke, there are worlds that are overall very similar to the actual world at which I did not win the lottery. But note that these worlds

differ significantly from the actual world with respect to the relevant *normative* facts. There is significant disparity between these worlds with respect to what I *may* and *must* do: for example, we might suppose that, at the world in which I win the lottery, I may quit my job, and must make large charitable donations, while at the nearby worlds at which I do not win the lottery, I must not quit my job, and I may donate much less to charity. Similarly, there is significant difference between these worlds with respect to the *desirability* of outcomes: for example, while at the world in which I win the lottery it is very desirable that I buy a house, at the nearby worlds in which I did not win the lottery and would struggle to make the loan repayments, it is very undesirable that I buy a house. These worlds also differ significantly with respect to the non-normative facts that *explain* these normative facts; given that the normative facts supervene on the non-normative facts, the significant difference in the normative facts is underwritten by a significant difference in the relevant non-normative facts.

It is thus open to the advocate of the simple counterfactual analysis of OUGHT to avoid counterexamples like the one presented above by specifying that the closeness relation between worlds that plays a role in the interpretation of the relevant counterfactual is one that at least gives significant weight to, if not reduces to[9], similarity between worlds with respect to *relevant modal facts*, or with respect to the non-modal facts that explain these modal facts. In this example, the relevant modal facts are *normative* facts—for example, facts about what *may* or *must* be the case, facts about the degree to which it is *fitting* to hold certain attitudes toward certain objects, and, perhaps also facts about what *explains* these other normative facts. With this understanding of the relevant closeness relation in mind, it seems much more plausible that in the example above, if it were not the case that I may buy a house or not the case that I may not buy a house, then it would be the case that I must buy a house. This is because we may assume that only a fairly small change in the normative facts (or the non-normative facts that explain the normative facts) would be required in order to make it the case that I must buy the house—for example, increasing the desirability of buying a house a little, or making the option of

[9] If the closeness relation reduced to similarity with respect to the normative facts, then this would mean giving up Strong Centering, and thus Conjunction Conditionalization.

investing in the stock market a little less desirable. On the other hand, given how desirable buying the house in fact is relative to the alternatives, it would take a more substantial change in the normative facts to make it the case that I am not permitted to buy it.

1.3.1.3 Controversial Principles

1.3.1.4 Inheritance

With regard to the controversial inheritance principles, the simple counterfactual analysis of OUGHT satisfies our second desideratum: it does not by itself entail the validity or invalidity of these principles. Rather, this analysis validates these principles only if similar principles hold for MAY and MUST.

Inheritance I: OUGHT

If it were not the case that it may be that *p* & *q* or not the case that it may not be that *p* & *q*, then it would be the case that it must be that *p* & *q*.

If it were not the case that it may be that *p* or not the case that it may not be that *p*, then it would be the case that it must be that *p*. (?)

The simple counterfactual analysis validates this schema if the following two controversial schemata are valid.[10]

---

[10] To see this, let us assume an analysis of counterfactuals in terms of the relative closeness of possible worlds and, for the sake of simplicity, assume that the limit assumption holds (Lewis, 1973; Stalnaker, 1968). Assume for reductio that there is some *p*, *q*, and *w* such that all of the closest worlds to *w* at which it is not the case that it may be that *p* & *q* or not the case that it may not be that *p* & *q* are worlds at which it must be that *p* & *q*, and at some closest world *v* to *w* at which it is not the case that it may be that p or not the case that it may not be that *p*, it is not the case that it must be that *p*. It follows that, at *v*, it is not the case that it may be that p, and so, given Inheritance I: MAY, it is not the case that it may be that *p* & *q.* Since the closest worlds to *w* at which it is not the case that it may be that *p* & *q* or not the case that it may not be that *p* & *q* are worlds at which it must—and, therefore, may—be that *p* & *q*, there must be some world, *u*, that is closer to *w* than *v* is at which it must be that *p* & *q*. Given Inheritance I: MUST, it follows that, at *u*, it must be that *p*, and so it is not the case that it may not be that *p*. But then *v* is not one of the closest worlds to *w* at which at which it is not the case that it may be that *p* or not the case that it may not be that p after all. It follows by reductio there is no *p*, *q*, and *w* such that, at *w*, it ought to be that *p* & *q*, and it is not the case that it ought to be that *p*.

Inheritance I: MUST

It must be that *p* and *q*.

It must be that *p*. (?)

Inheritance I: MAY

It may be that *p* and *q*.

It may be that *p*. (?)

Next, consider Inheritance II: OUGHT.

Inheritance II: OUGHT

If it were not the case that it may be that *p*, or not the case that it may not be that *p*, then it would be the case that it must be that *p*.

If it were not the case that it may be that *p* or *q*, or not the case that it may not be that *p* or *q*, then it would be the case that it must be that *p* or *q*. (?)

This analysis vindicates Inheritance II: OUGHT just in case the following two controversial schemata are valid.[11]

Inheritance II: MUST

It ought to be that *p*.

It ought to be that *p* or *q*. (?)

[11] To see this, assume for *reductio* that there is some *p*, *q*, and *w*, such that all of the closest worlds to *w* at which it is not the case that it may be that *p*, or not the case that it may not be that *p*, are worlds at which it must be that *p*, and at some closest world *v* to *w* at which it is not the case that it may be that *p* or *q*, or not the case that it may not be that *p* or *q*, it is not the case that it must be that *p* or *q*. It follows that, at *v*, it is not the case that it may be that *p* or *q*, and so, by Inheritance II: MAY, it is not the case that it may be that *p*. Since the closest worlds to *w* at which it is not the case that it may be that *p* or not the case that it may not be that *p* are worlds at which it must—and, therefore, may—be that *p*, there is some world, *u*, that is closer to *w* than *v* is at which it must be that *p*. Given Inheritance II: MUST, it follows that, at *u*, it must be that *p* or *q*, and so it is not the case that it may not be that *p* or *q*. But then *v* is not one of the closest worlds to *w* at which at which it is not the case that it may be that *p* or *q* or not the case that it may not be that *p* or *q* after all. It follows by *reductio* there is no *p*, *q*, and *w* such that, at *w*, it ought to be that *p*, and it is not the case that it ought to be that *p* or *q*.

Inheritance II MAY

It may be that *p*.

It may be that *p* or *q*. (?)

1.3.1.5 Agglomeration and Axiom K

Unfortunately this account ultimately fails to meet our second desideratum, inasmuch as it appears straightforwardly to entail the invalidity of Agglomeration: OUGHT and Axiom K: OUGHT, even if Agglomeration: MUST and Axiom K: MUST are valid.

Agglomeration: OUGHT

If it were not the case that it may be that *p*, or not the case that it may not be that *p*, then it would be the case that it must be that *p*;
If it were not the case that it may be that *q*, or not the case that it may not be that *q*, then it would be the case that it must be that *q*.

If it were not the case that it may be that *p* and *q*, or not the case that it may not be that *p* and *q*, then it would be the case that it must be that *p* and *q*. (?)

To see that this is so, imagine an example in which *v* is the closest world to *w*, and the following propositions are true at each world.

- At *w*:
    - It must be that *p*:
        - It may be that *p*;
        - It is not the case that it may not be that *p*;
    - It may be that *q*;
    - It may not be that *q*;
    - It may be that *p* and *q*;
    - It may not be that *p* and *q*; and
- At *v*:
    - It must be that *q*:
        - It may be that *q*;
        - It is not the case that it may not be that *q*; and
    - It is not the case that it may be that *p* and *q*;
    - It is not the case that it must be that *p* and *q*.

In this case, at *w*, it ought to be that *p*, since at the closest world at which it is not the case that it may be that *p* or not the case that it may not be that *p*—namely, *w* itself—it must be that *p*. It ought to be that *q*, since at the closest world at which it is not the case that it may be that *q* or not the case that it may not be that *q*—namely, *v*—it must be that *q*. But it is not the case that it must be that *p* and *q*, since the closest world at which it is not the case that it may be that *p* and *q* or not the case that it may not be that *p* and *q*—namely, *v*—it is not the case that it must be that *p* and *q*.

Similar reasoning shows that this account does not vindicate the validity of Axiom K: OUGHT either, even if Agglomeration: Axiom K: MUST is valid.

Axiom K: OUGHT

If it were not the case that it may be that ¬*p* or *q*, or not the case that it may not be that ¬*p* or *q*, then it would be the case that it must be that ¬*p* or *q*;
If it were not the case that it may be that *p*, or not the case that it may not be that *p*, then it would be the case that it must be that *p*.

If it were the not the case that it may be that *q*, or not the case that it may not be that *q*, then it would be the case that it must be that *q*. (?)

In view of the purported counterexamples to the rational principles expressed by these schemata in the literature, some may not be troubled by the fact that this analysis fails to validate these principles. Since I would prefer an account that meets our second desideratum, however—and since I myself do not find these counterexamples particularly persuasive—in the next section I will explore a variation on this analysis that promises to avoid this problem.

1.3.2 A Strengthened Counterfactual Analysis

According to the simple counterfactual account, to judge that someone ought to do something is, roughly, to judge that they would have to do it if it were *not optional*. We have just seen that this analysis entails that Agglomeration: OUGHT and Axiom K: OUGHT are subject to straightforward counterexamples. These counterexamples all rely on there being worlds at which some things are required while other things are optional. In this section, I will explore an alternative analysis that avoids these counterexamples by taking these kinds of worlds out of the equation. According to

this analysis, to judge that someone ought to do something is, roughly, to judge that they would have to do it if *nothing* were optional.

(The Strengthened Counterfactual Analysis of OUGHT) For all *p*, to judge that it ought to be that *p just is* to judge that, if it were the case that, for all *r*, either it is not the case that it may be that *r*, or it is not the case that it may not be that *r*, then it would be the case that it must be that *p*.

Let us call a world at which nothing is optional a 'normatively determinate' world. That is, a world is normatively determinate if, at that world, for all *p*, either it is not the case that it may be that *p*, or it is not the case that it may not be that *p*. According to the strengthened counterfactual analysis, then, to judge that something ought to be so is, roughly, to judge that it must be so at all of the closest normatively determinate worlds. One worry about this analysis is that it may be that no possible worlds are normatively determinate worlds. In that case, judgements about what ought to be so would turn out to be *counterpossible* judgements rather than counterfactual judgements. This is a cost of the view, in so far as judgements about what ought to be so are entirely familiar and uncontroversial, whereas counterpossible judgements are not. Another worry about this analysis is that, even if some possible worlds are normatively determinate worlds, these worlds are presumably very distant from the actual world. This is a cost of the view in so far as forming judgements about what actually ought to be so intuitively does not seem to require thinking about such far-flung possibilities.[12] The strengthened counterfactual analysis is thus not without its problems. Nonetheless, I present it here as one possible way of modifying the simple counterfactual analysis in order to avoid straightforwardly invalidating Agglomeration and Axiom K.

1.4 Conclusion

In this chapter, I have provided two novel approaches to analysing the concepts MAY, MUST, and OUGHT. I argued that MAY and MUST can be analysed in terms of fitting-response concepts, and I showed that OUGHT can be analysed—one way or another—as a counterfactual MUST. I argued for these analyses by showing that they

[12] I am grateful to Alan Hájek for raising this objection.

satisfy several key desiderata. In an appendix, I explore a natural way of extending the counterfactual analysis of OUGHT to account for the concepts REASON (mass) and REASON (count). In the next chapter, I show that fitting-attitude concepts can themselves be analysed in terms of the concept RATIONAL.

2 Fittingness and Rationality

In the previous chapter, I argued that the concept OUGHT is analysable in terms of the concepts MAY and MUST (and in an appendix I show how this analysis can be extended to cover the concept REASON), and I showed that the concepts MAY and MUST are themselves analysable in terms of fitting-response concepts, such as PERMISSIBLE, ADVISABLE, ACCEPTABLE, REASONABLE, and DESIRABLE. In this chapter, I aim to show that these kinds of fitting-response concepts are in turn analysable in terms of the concept RATIONAL. Together, these views imply that all of the central normative concepts—MAY, MUST, OUGHT, and REASON—are ultimately analysable in terms of the concept RATIONAL. I thus call this a 'rationality-first' account of normativity.

One of the benefits of this kind of rationality-first account of normativity, I will argue, is that it promises to make sense of the fact that normative judgements are subject to certain distinctive epistemic principles. I begin this chapter by describing these principles. The first principle, which I will call 'Rationalism', says that, for any proposition about the normative facts, given sufficient information about the non-normative facts, one can know *a priori* whether or not that proposition is true. The second principle is Hume's Law. Hume's Law is naturally thought of as a basing prohibition, according to which no normative judgement may be based on a set of attitudes that does not itself include a normative judgement. Since an adequate formulation of this principle is hard to come by, however, I focus instead on a weaker version of it that says merely that there are *some* normative judgements that cannot be based on a set of attitudes that does not include any normative judgements.

After setting out these epistemic principles, I provide an account of the concept RATIONAL and show that, if normative concepts were all ultimately analysable in terms of this concept, then that would make sense of the fact that normative judgements are subject to these epistemic principles. The concept that I have in mind is one that can also be expressed using expressions like 'makes sense' or 'is intelligible', and is constitutively tied to mental-state concepts. There are two ways in which this concept is tied to mental-state concepts. First, simply by virtue of possessing a mental-state concept, one is in a position to know *a priori* the conditions under which that mental state would be rational or irrational. And, secondly, judgements about rationality

constrain mental-state ascriptions. Rationality requires that we exercise *charity* in ascribing mental states: in ascribing a mental state to someone, we must do so in a way that we take to be consistent with ascribing as much rationality to them as possible, given their behaviour and environment. This idea is encapsulated in a set of *principles of charity* that govern the ascription of mental states. I formulate these principles below using the expressions 'rationality requires' and 'rationality permits' as a shorthand for talking about whether someone's being in a mental state is or is not consistent with an interpretation of them that maximises the extent to which their mental states make sense. I then provide analyses of RATIONALITY REQUIRES and RATIONALITY PERMITS that make this idea precise.

Finally, I turn to the task of analysing fitting-response concepts in terms of the concept RATIONAL. I begin by setting out three desiderata for an analysis of fitting-response concepts. First, the analysis should vindicate the rational principles that govern how judgements involving fitting-response concepts relate to each other; secondly, the analysis should make sense of the connection between fitting-response concepts and the concept of rationality; and, thirdly, the analysis should make sense of the distinction between perspective-dependent and perspective-independent fitting-response concepts. I then provide an analysis of fitting-response concepts in terms of the concept RATIONALITY PERMITS, that is inspired by Rabinowicz' (2008; 2009; 2012; 2017) accounts of value and probability relations, and show that it satisfies these three desiderata. I close by acknowledging that this account is in an important respect incomplete, since it fails to account for comparisons in the fittingness of responses at different worlds and from different perspectives. This is something to be addressed in future work.

2.1 Epistemic Principles

In this section, I present two distinctive epistemic principles to which normative judgements are subject. Both of these principles are intuitively plausible, so I will not spend much time arguing for them. I begin with a principle that I will call 'Rationalism'. This principle captures the intuitive thought that, for any normative proposition, given sufficient information about the relevant non-normative facts, one can know *a priori* whether or not that proposition is true. For example, although I cannot know *a priori* whether Sam is admirable, I can know *a priori* whether or not

Sam is admirable, given sufficient information about their character, actions, achievements, and so on. Similarly, although I cannot know *a priori* whether I ought to donate money to some charity, I can know *a priori* whether or not I ought to donate to that charity, given sufficient information about that charity, how it compares with other charities, the state of my financial situation, and so on. The following principle makes this intuitive idea precise.

(Rationalism) For all p, and q:

- If:
    - The judgement that p is a normative judgement;
    - The judgement that q does not involve any normative concepts; and
    - It is not *a priori* that $\neg q$;
- Then there is some r such that:
    - The judgement that r does not involve any normative concepts;
    - It is not *a priori* that $\neg r$;
    - It is *a priori* that $r \supset q$; and
    - Either it is *a priori* that $r \supset p$ or it is *a priori* that $r \supset \neg p$.

For our purposes here, let us say that, for all p, it is *a priori* that p just in case rationality permits that the judgement that p be not ultimately based on sense experience, introspection, proprioception, memory, or the like (Russell, 2020). This formulation of rationalism is neutral on the question of what rationality permits these judgement to be based on instead—for example, rational intuition, understanding of the concepts involved, or nothing at all. I will not try to resolve this more controversial question here.

This principle can be summarised roughly as follows: for any normative proposition p and consistent (partial) description of the non-normative facts, q, there is a consistent description of the non-normative facts that adds enough extra detail to q that it *a priori* entails p or *a priori* entails $\neg p$. One might wonder whether the references to both p and q are necessary in the formulation of this principle. If the reference to q were omitted, then we would be left with a principle that says that, for every normative proposition, any (partial) description of the normative facts *a priori* entails its truth or falsity. This principle is false, since a partial description of the normative

facts may not be sufficiently detailed to establish whether the relevant normative proposition is true. It the reference to *p* were omitted, then we would be left with a principle that says that, for every normative proposition, *some* consistent description of the non-normative facts *a priori* entails its truth or falsity. This principle seems correct, but too weak. What we are after instead is a principle according to which, for every normative proposition, *any* sufficiently detailed consistent description of the non-normative facts *a priori* entails its truth or falsity.

A closely related principle to the one I have provided is the principle that, for any normative proposition, any *complete* consistent description of the non-normative facts *a priori* entails its truth or falsity. This principle also seems correct, though weaker than the principle I have proposed. My reason for focusing on the stronger principle is that it applies to the kinds of judgements that we actually make. Making an *a priori* judgement about what follows, normatively speaking, from some merely partial but sufficiently detailed description of the non-normative facts is an everyday occurrence. On the other hand, since it is doubtful that anyone has ever successfully entertained a complete description of the non-normative facts (let alone a consistent one), it is highly doubtful that anyone has ever made a judgement about what normative facts are consistent with such a description. A rational principle that says that some such judgements are *a priori* is therefore of less immediate interest.[15]

One might be sceptical about Rationalism in view of the fact that it seems possible for a fully rational agent to be uncertain about some normative fact in spite of possessing all of the relevant non-normative facts.[16] For example, one might argue as follows.

1. For some normative proposition, p, and sufficiently detailed consistent description of the non-normative facts, q, it is possible for a fully rational agent to be uncertain whether $q \supset p$ or whether $q \supset \neg p$.
2. For all p, if it is *a priori* that p, then it is not possible for a fully rational agent to be uncertain whether p.

[15] I am grateful to an anonymous examiner for questioning the formulation of this principle.

[16] I am grateful to an anonymous examiner for raising this objection.

3. So there is some normative proposition, *p*, and sufficiently detailed consistent description of the non-normative facts, *q*, such that it is not *a priori* that $q \supset p$ and it is not *a priori* that $q \supset \neg p$.
4. Therefore Rationalism is false.

The problem with this argument is that the second premise seems too demanding. A fully rational agent need not believe all *a priori* truths. For example, a fully rational agent may fail to believe an *a priori* truth because rationally forming that belief requires going through a process of reasoning which the agent has not had sufficient opportunity or reason to undertake. Forming a rational belief about what the normative upshots of some complex set of non-normative facts are often does require considerable reasoning. For example, according to a highly influential account of normative reasoning, such reasoning can require making a series of adjustments to one's overall set of normative judgements until a state of reflective equilibrium is achieved (Rawls, 1971; Daniels, 1980). In some cases, it would be irrational for an agent to form a judgement about the normative upshots of some complex set of non-normative facts before they have had a chance to complete this procedure.

Another reason that one might be sceptical about Rationalism is that it seems possible for fully rational agents to disagree about the normative facts in spite of agreeing on all of the non-normative facts.[17] For example, one might reason as follows.

1. For some normative proposition, *p*, and sufficiently detailed consistent description of the non-normative facts, *q*, it is possible for two fully rational agents to disagree about whether $q \supset p$ and whether $q \supset \neg p$.
2. For all *p*, if it is *a priori* that *p*, then it is not possible for two fully rational agents to disagree about whether *p*.
3. So there is some normative proposition, *p*, and sufficiently detailed consistent description of the non-normative facts, *q*, such that it is not *a priori* that $q \supset p$ and it is not *a priori* that $q \supset \neg p$.
4. Therefore Rationalism is false.

[17] I am grateful to an anonymous examiner for raising this objection.

It is difficult to assess the first premise of this argument. While normative disagreement seems to be quite widespread, perfect rationality is extremely rare, if not non-existent. Nevertheless, supposing this premise is true, why believe the second premise? One might reason as follows. It is *a priori* that *p* just in case there is a rational way to arrive at the judgement that *p* independently of experience, and, if there is a rational way to arrive at the judgement that *p* independently of experience, then:

i. Any fully rational agent who arrives at a conclusion about whether *p* independently of experience will arrive at the judgement that *p*; and
ii. No fully rational agent will arrive at the judgement that ¬*p* based on experience.

The idea behind ii may be that, if there a way to arrive at the belief that *p* independently of experience, then a fully rational agent would not form a conclusion about whether *p* without checking that conclusion independently of experience, and given that i is true, they would thereby arrive at the judgement that *p*. i seems plausible for other *a priori* domains of inquiry such as mathematics, but one might question whether it follows that it is true for every *a priori* domain of inquiry. If reflective equilibrium is a rational way to arrive at normative judgements, then i is plausibly false. It is widely recognised that faultlessly following the reflective equilibrium procedure may not lead agents to the same normative conclusions. Rather than taking this to support i, however, some take it instead to show that reflective equilibrium is not, by itself, a rational method for arriving at normative judgements (Kelly and McGrath, 2010). Many have also been tempted by the idea that the normative domain is somehow less objective than domains like mathematics, such that the truth of a normative judgement can vary according to who is making the judgement. If this thought is correct, then it is natural to think that i is false for normative judgements. The dialectical force of point is limited, however, since this kind of relativism is highly controversial.

I have raised a few doubts about this argument, but I have not shown conclusively that it is unsound. Some readers will thus still be inclined to take it to establish that Rationalism is false. In the first two chapters of this thesis, I argue that normative judgements *just are* judgements about rationality. To anticipate, in what follows, I will

argue that the thesis that facts about rationality can be known *a priori*, given enough information about the non-rational facts, explains why Rationalism is true. My hope is that anyone who is sceptical about Rationalism will be similarly sceptical about the thesis that all truths about rationality can be known *a priori*, given sufficient information about the non-rational facts, and so can still accept my main thesis.

The second epistemic principle is Hume's Law (Hume, 2000). Hume's Law says, roughly, that one cannot validly derive an 'ought' from an 'is'. It is tempting to try to formulate this principle as a basing prohibition along the following lines.

(Hume's Law) For all A, B_1, … , and B_n, if A is a normative judgement, and B_1, … , and B_n are non-normative judgements, then rationality prohibits that A be based on B_1, … , and B_n.

While this rational principle seems plausible at first sight, it is subject to counterexamples (Prior, 1960). Consider the judgement that tea drinking is common in England or all New Zealanders ought to be shot. If this is a normative judgement, then it yields a counterexample to Hume's law as formulated above, since it can be rationally based on a single non-normative judgement—namely, the judgement that tea drinking is common in England.

Tea drinking is common in England.

Tea drinking is common in England or all New Zealanders ought to be shot.

If instead the judgement that tea drinking is common or all New Zealanders ought to be shot is not a normative judgement, then it yields a different counterexample to Hume's Law as formulated above, since it, in combination with the non-normative judgement that it is not the case that tea drinking is common in England form the basis for a normative judgement—namely, the judgement that all New Zealanders ought to be shot.

Tea drinking is common in England or all New Zealanders ought to be shot;
It is not the case that tea drinking is common in England.

All New Zealanders ought to be shot.

There are various proposals in the literature about how to reformulate Hume's Law in a way that avoids these kinds of counterexamples and specifies the precise

conditions under which the basing prohibition applies (see, for example, Pigden, 1989; 2010; Jackson, 1974; Brown, 2014). Trying to decide between these proposals would take us too far afield here. In any case, there is little need to do so, since what is arguably most interesting about Hume's Law from a meta-ethical perspective is not what the precise conditions are under which the basing prohibition applies, but rather the fact that there are conditions under which the basing prohibition applies at all—that is, that there are normative judgements that cannot rationally be based solely on judgements that do not involve any normative concepts. In light of this, the following weaker version of Hume's Law is sufficient here.

(Weak Hume's Law) There is some *A* such that *A* is a normative judgement, and, for all B_1, … , and B_n, if B_1, … , and B_n are judgements that do not involve any normative concepts, then rationality prohibits that *A* be based on attitudes B_1, … , and B_n.

Intuitively, the judgement that all New Zealanders ought to be shot is a judgement of this kind, while, as shown above, the judgement that tea drinking is common in England or all New Zealanders ought to be shot is not.

2.2 The Concepts RATIONAL and IRRATIONAL

With these epistemic principles now on the table, I next present an account of the concepts RATIONAL and IRRATIONAL such that, if all normative concepts were ultimately analysable in terms of these concepts, then that would explain why normative judgements are subject to these epistemic principles. The concepts I have in mind can also be expressed using expressions like 'makes sense', 'makes no sense', 'is intelligible', and 'is unintelligible'. I identify these concepts in what follows by describing the ways in which they are constitutively tied to mental state concepts.

2.2.1 *A Priori* Conceptual Truths about Rationality

The first identifying feature of the concepts RATIONAL and IRRATIONAL is that anyone who possesses the relevant mental-state concepts is thereby in a position to know *a priori* whether a mental state that someone is in is rational or irrational, given sufficient information about their behaviour, mental states, and environment. In other words, judgements involving the concepts RATIONAL and IRRATIONAL are subject to an analogue of Rationalism.

(Rationalism: RATIONAL/IRRATIONAL) For all *p*, *S*, and φ:

- If:
    - The judgement that *p* is a judgement about *S*'s mental states, behaviour, and/or environment that does not involve the concepts RATIONAL or IRRATIONAL; and
    - It is not *a priori* that ¬*p*;
- Then there is some *q* such that:
    - The judgement that *q* is a judgement about *S*'s mental states, behaviour, and/or environment that does not involve the concepts RATIONAL or IRRATIONAL;
    - It is not *a priori* that ¬*q*;
    - It is *a priori* that *q* ⊃ *p*; and
    - Either it is *a priori* that *q* ⊃ *S*'s φ-ing is (ir)rational, or it is *a priori* that *q* ⊃ *S*'s φ-ing is not (ir)rational.

This principle can be summarised roughly as follows: for any agent and response, for any (partial) consistent description of the non-rational facts about that agent, *p*, there is some such description that adds enough extra detail to *p* that *a priori* entails that the relevant response is (ir)rational or that it is not (ir)rational. As with the original formulation of Rationalism, one might wonder whether the references to both *p* and *q* are necessary in the formulation of this principle. If the reference to *q* were omitted, then we would be left with a principle that says that, for any agent and response, any (partial) consistent description of the non-rational facts about that agent *a priori* entails that the relevant response is (ir)rational or that it is not (ir)rational. This principle is false, since a partial description of the non-rational facts about the agent may not be sufficiently detailed to establish whether the relevant response is (ir)rational. It the reference to *p* were omitted, then we would be left with a principle that says that, for any agent and response, *some* consistent description of the non-rational facts about that agent *a priori* entails that the relevant response is (ir)rational or that it is not (ir)rational. This principle seems correct, but too weak. What we are after instead is a principle according to which, for any agent and response, *any* sufficiently detailed consistent description of the non-rational facts about that agent *a priori* entails that the relevant response is (ir)rational or that it is not (ir)rational.

A closely related principle to the one I have provided is the principle that, for any agent and response, any *complete* consistent description of the non-rational facts about that agent entails that the relevant response is (ir)rational or that it is not (ir)rational. This principle also seems correct, though weaker than the principle I have proposed. My reason for focusing on the stronger principle is again that it applies to the kinds of judgements that we actually make. Making an *a priori* judgement about whether some agential response is (ir)rational given a merely partial but sufficiently detailed description of the non-rational facts about an agent is an everyday occurrence. On the other hand, since it is doubtful that anyone has ever successfully entertained a complete description of the non-rational facts about an agent's mental states, behaviour, and environment (let alone a consistent one), it is highly doubtful that anyone has ever made a judgement about what follows from such a description as regards the rationality of the agent's responses. A rational principle that says that some such judgements are *a priori* is therefore of less immediate interest.[18]

Each of the following, for example, is an *a priori* material conditional whose antecedent is a sufficiently specific proposition regarding someone's mental states, and whose consequent has to do with whether some mental state that they are in is rational or irrational.

> If Sam simultaneously believes that it is raining and believes that it is not raining, then Sam's simultaneously believing that it is raining and believing that it is not raining is irrational.
>
> If Sam simultaneously desires that they get to work on time, believes that, if the bus were to arrive on time, then they would get to work on time, and, based on these two attitudes, desires that the bus arrive on time, then Sam's desire that the bus arrive on time is rational.[19]

---

[18] I am grateful to an anonymous examiner for questioning the formulation of this principle.

[19] The last example shows that the kind of rationality I have in mind is a kind of *local* rationality. Even if either the desire or the belief on which Sam's instrumental desire is based does not itself make sense, the fact that their instrumental desire is based on these other attitudes in the right way is enough to make it rational or intelligible, on this view.

I do not have an argument for the truth of Rationalism: RATIONAL/IRRATIONAL. It simply seems intuitively correct to me. If Rationalism: RATIONAL/IRRATIONAL is false, then knowing all of the facts about someone's mental states, behaviour, and environment (other than the facts about the rationality of those mental states and behaviours) is not, by itself, enough to put one into a position to know which of their mental states and behaviours are rational. Rather, whether some such mental state is rational, given all of these other facts, is an empirical matter, and can only be revealed though some further experience—for example, by literally *perceiving* that the mental state in question is or is not rational. In any case, my hope is that anyone who is sceptical about Rationalism: RATIONAL/IRRATIONAL will also be sceptical about Rationalism itself, and so can still accept my main thesis that normative judgements are judgements about rationality.

The best explanation for the fact that truths of this kind are *a priori* is that they are *conceptual* truths: part of what it is to possess a mental-state concept is to be in a position to know *a priori* the conditions under which it would or would not make sense for someone to be in that mental state (Zangwill, 2005; 2010; cf. Wedgwood, 2007, p. 159). For example, part of what it is to possess the concept BELIEF is to be in a position to know *a priori* that it does not make sense simultaneously to believe that *p* and to believe that ¬*p*; and part of what it is to possess the concept DESIRE is to be in a position to know *a priori* that desiring that *q* makes sense if it is based on a desire that *p* and a belief that, if *q* were the case then *p* would be the case.

This thesis might be appropriately categorised as a form of 'normativism' about the mental—the view that mental states are essentially subject to certain normative standards. Before proceeding, I should take a moment to distinguish the thesis that I am proposing from some other theses that are sometimes associated with that view (Glüer and Wikforss, 2009, p. 46). I am proposing that facts about the rationality of mental states are *a priori*, and that being in a position to know facts of this kind *a priori* is part of what it is to possess a mental-state concept. I am not presupposing here that knowledge of this kind is capable of *guiding* agents in the formation of mental states—though in the next two chapters I will defend a view about how rational agents are in fact capable of regulating their own mental states in response to their beliefs about what rationality permits or requires of them. I am not claiming

that *being in* a particular mental state requires knowing, or being in a position to know, any facts about rationality. Nor am I claiming that being in a particular mental state requires being motivated or disposed to ensure that that mental state is rational. And, finally, I am not claiming that being in a particular mental state requires that one's mental states are rational overall, for the most part. The thesis that I am proposing is thus not susceptible to Glüer and Wikforss' (2009) influential critique of versions of normativism that are committed to these theses.

As mentioned above, the concepts RATIONAL and IRRATIONAL can be expressed using expressions such as 'makes sense', 'makes no sense', 'is intelligible', and 'is unintelligible'. Sometimes when philosophers use the words 'rational' and 'irrational', they express something other than the concepts I am describing here. We can imagine cases of mental states or behaviours that *a priori* do (not) make sense or are (not) intelligible, in the sense I have in mind here, but which some philosophers might be reluctant to call '(ir)rational', given the way they are accustomed to using that term. Consider the following example.

> If Sam has a visual experience of seeing a red rectangle, and there are no red rectangles in the vicinity, then Sam's visual experience of seeing a red rectangle does not make sense.

While some philosophers might be uncomfortable the idea of calling perceptual experiences 'rational' or 'irrational' (Goldman, 2000), there certainly is a sense of the expression 'make sense', given which this sentence expresses is an *a priori* truth. To clarify the relevant sense of 'makes sense', we might say something like the following. It is somehow *in the nature* of visual experience to be responsive to the experiencer's environment in a certain way, and, when a visual experience is out of touch with the experiencer's environment in the way that it is in this example, then something has gone *wrong*—the visual experience has been produced in an *uncharacteristic* way. It is this concept of making or not making sense that I am referring to in this chapter when I use the expressions 'RATIONAL' and 'IRRATIONAL'.

One might wonder whether the concepts RATIONAL/IRRATIONAL are identical to some statistical concepts like USUAL/UNUSUAL or TYPICAL/ATYPICAL. Analytic functionalists, for example, sometimes provide examples of putative analytic truths that are quite similar to those presented above, but that are couched in terms of what is 'typically' or 'usually' the case, rather than what is rational (see, for example, Braddon-Mitchell

and Jackson, 1996, pp. 52-53, 58). For example, analytic functionalists sometimes present purported analytic truths like the following as forming the basis for their functionalist analyses.

> For all *p* and *q*, if someone desires that *p* and believes that, if *q* were the case, then *p* would be the case, then they *typically*/*usually* desire that *q*.

Although nothing in what follows turns on this issue, I am inclined to doubt that the concepts I have identified are identical to any such statistical concepts for two related reasons. First, it seems implausible that truths about whether a mental state is usual/unusual or typical/atypical are *a priori* conceptual truths. Rather, truths about what is usual/unusual or typical/atypical are generally learnt through experience. And, secondly, it seems epistemically possible for irrational responses to be usual or typical, and for rational responses to be unusual or atypical. For example, it may be that, given the evidence, it is perfectly rational to believe that people *usually* or *typically* commit the Gambler's Fallacy while simultaneously believing that their doing so is *irrational* (Zangwill, 1998; 2005).

2.2.2 The Role of Judgements about Rationality in the Ascription of Mental States

The second identifying feature of the concepts RATIONAL and IRRATIONAL is that mental-state ascriptions are rationally constrained by judgements about rationality. Given all of the evidence about someone's observable behaviour and environment, how is one to determine what their mental states are? A familiar answer is that one should apply some kind of principle of charity or rationalisation: one should interpret them in a way that, by one's own lights, maximises the extent to which their behaviours and mental states *make sense*, given their behaviour and environment (Davidson, 1984; Dennett, 1989; Lewis, 1974). In other words, one should try to interpret them in a way that maximises the extent to which their behaviours make sense in light of their mental states, their mental states make sense in light of their other mental states, and their perceptual states make sense in light of their immediate environment.

We might say, then, that charity requires that one ascribe a mental state to someone, if, and only if, one is disposed to judge that, according to the 'best interpretation' of

them that is consistent with the facts about their behaviour and environment, they are in that mental state—where the best interpretation is the one according to which the greatest proportion of their behaviours and mental states make sense. Below, I provide analyses of RATIONALITY REQUIRES and RATIONALITY PERMITS, according to which to judge that rationality requires that someone have some response is to judge that they have that response according to every interpretation that is 'best' in this way; and to judge that rationality permits that someone have some response is to judge that they have that response according to some interpretation that is 'best' in this way. I thus formulate the principles of charity below in terms of these concepts.

Although it seems to be an assumption in the literature on the role of charity in interpretation that there is (at most) a single principle of charity, I identify six principles of charity that plausibly govern the ascription of mental states. There are two *positive* principles of charity. These have to do with ascribing a mental state to someone if one judges that they are rationally required to be in that state, given their behaviour and environment. There are also four *negative* principles of charity. Two of these have to do with refraining from ascribing a mental state to someone if one is not disposed, on reflection, to judge that they are rationally permitted to be in that state, given their behaviour and environment. The other two are perhaps more controversial and have to do with refraining from ascribing a mental state to someone if one believes that they are not rationally required to be in that state, given their behaviour and environment.

My immediate purpose in presenting these principles in this chapter is merely to give an overall sense of the role that the concept RATIONAL plays in mental-state ascription in order to make it clear which concept I am talking about. For this purpose, the details about precisely how these principles are to be formulated are not that important. Since I will be relying heavily on these principles in chapter four when I take on the task of explaining why normative judgements are subject to enkratic principles, however, I will spend some time spelling them out carefully here.

2.2.2.1 Positive Principles of Charity

First, I propose that there is a rational principle that requires that, if one judges that someone is rationally required to have some response, given their behaviour and

environment—that is, if one judges that they have that response according to every interpretation of them that maximises the extent to which their responses make sense, given their behaviour and environment—then one ascribes that response to them. The idea here is that once the question of whether that response is rationally required, given the agent's behaviour and environment, has been resolved, there is no further work to be done in order to resolve the question of whether the agent has that response, and so the belief that the agent has the response should follow immediately.

(Positive Requirement of Charity) For all *S*, *T*, *t*, and φ, rationality requires that:

- If:
    - *S* believes at *t* that, given all the facts about *T*'s behaviour and environment, rationality requires of *T* that they φ;
- Then:
    - *S* believes at *t* that *T* φs.

To clarify: this principle does not say that it is never rational to ascribe an irrational response to someone. The analysis of RATIONALITY REQUIRES that I have provided does not entail that, if a response is rationally required of someone, given their behaviour and environment, then that response would not be irrational. Rather, according to the analysis of RATIONALITY REQUIRES that I have provided, if a response is rationally required of someone, given their behaviour and environment, then that response is a part of the total package of mental states that, together, would render them as rational as possible overall, given their behaviour and environment. It is possible that the total package of mental states that would make someone as rational as possible overall, given their behaviour and environment, might include some irrational mental states. Thus, the Positive Requirement of Charity does permit one to ascribe an irrational response to someone, provided one does not take their having that response to be incompatible with their being as rational as possible *overall*, given the facts about their behaviour and environment.

Next, I propose that there is a corresponding rational basing permission according to which one is permitted to judge that someone has some response based on the

belief that, given their behaviour and environment, rationality requires that they have that response.

(Positive Permission of Charity) For all *S*, *T*, and φ, rationality permits that *S*'s belief that *T* φs be based on *S*'s belief that, given all the facts about *T*'s behaviour and environment, rationality requires of *T* that they φ.

2.2.2.2 Negative Principles of Charity

Next, I present two sets of 'negative' principles of charity. The first two principles have to do with not ascribing a response to someone if one is not disposed, on reflection, to judge that, given their behaviour and environment, rationality permits that they have that response—that is, if one is not disposed to judge that they are best interpreted as having that response, given their behaviour and environment. The first is a rational requirement not to ascribe a response in this kind of situation.

(Negative Requirement of Charity I) For all *S*, *T*, *t*, and φ, rationality requires that:

- If:
    - *S* is not disposed at *t*, on reflection, to judge that, given all of the facts about *T*'s behaviour and environment, rationality permits of *T* that they φ;
- Then:
    - *S* is not disposed at *t*, on reflection, to believe that *T* φs.

The next principle permits that one's not ascribing a response to someone be based on one's not being disposed, on reflection, to judge that rationality permits that they have that response.

(Negative Permission of Charity I) For all *S*, *T*, and φ, rationality permits that *S*'s not being disposed, on reflection, to believe that *T* φs be based on *S*'s not being disposed, on reflection, to judge that, given all of the facts about *T*'s behaviour and environment, rationality permits of *T* that they φ.

The final two principles are more controversial and have to do with not ascribing to someone a response that one believes not to be rationally required of them, given their behaviour and environment. The idea here is that one should not commit to

believing that someone has some response state if one believes that there are perfectly good interpretations of them, given their environment and behaviour, that do not ascribe that response to them. The first principle is a rational requirement not to ascribe a response that one believes not to be rationally required.

(Negative Requirement of Charity II) For all *S*, *T*, *t*, and φ, rationality requires that:

- If:
    - *S* believes at *t* that it is not the case that rationality requires of *T*, given their behaviour and environment, that they φ;
- Then:
    - *S* is not disposed at *t*, on reflection, to believe that *T* φs.

The reason this principle is more controversial than the others is that one might be inclined to think that there can sometimes be reasons for favouring one interpretation over another, even if each is equally good in view of the facts about the relevant agent's behaviour and environment. Whether this is so turns on the question of whether it is possible to have decisive evidence that someone has some response that is not also decisive evidence that their behaviour and environment are such that all of the best interpretations of them, given all of the facts about their behaviour and environment, are ones that ascribe that response to them. It has been suggested to me that the fact that someone has intentionally acted in a certain way many times in the past might be a decisive reason to believe that they will do so again, but not a decisive reason to believe that this is so according to all of the most rational interpretations of them, given all of the facts about their behaviour and environment. If this is correct, then it may be rational simultaneously both to believe that they will perform the action in question again, and to believe that rationality does not require this of them, given their behaviour and environment.[20]

It seems plausible to me that the fact that the person has intentionally acted in a particular way many times in the past does in fact constitute evidence that their

20 I am grateful to an anonymous examiner for this example.

behaviour and environment are such that the most rational interpretations of them entail that they will act in that way again. The fact that they have acted in that way in the past certainly seems to provide inductive evidence that they will *behave* in that way again. And the fact that their previous behaviour counted as *intentional* seems to provide some evidence that the best interpretations of them are ones that ascribe to them the kinds of mental states that made the previous instances of the behaviour—and so perhaps also future instances of it—intentional.

In my view, a more damaging objection to this principle is the thought that that we might have direct first-person access to some of our own mental states, and thus can have reason to self-ascribe them even in the face of the belief that the facts about our behaviour and environment do not commit us to being in them. In any case, I concede that the Negative Requirement of Charity II is controversial. This principle plays a key role in my attempt to explain why judgements about what *ought* to be the case are subject to a particular kind of Enkratic principle (see §4.2.3.3). In so far as this principle is controversial, then, that casts doubt upon that enterprise.

The final principle is the counterpart to the previous one and permits that one's not being disposed to believe that someone has some response be based on one's not believing that they are rationally required to have that response, given their behaviour and environment.

(Negative Permission of Charity in Self-Ascription II) For all *S*, *T*, and φ, rationality permits that *S*'s not being disposed, on reflection, to believe that *T* φs be based on their belief that it is not the case that rationality requires of *T*, given their behaviour and environment, that they φ.

2.2.2.3 Rationality Permits and Rationality Requires

The guiding idea behind these principles of charity is that we should be disposed to ascribe a mental state to someone if, and only if, we are confident that this mental state is part of a set of mental states that makes the most sense of their behaviours and other mental states, given the facts about their behaviour and environment. In this section, I provide analyses of RATIONALITY PERMITS and RATIONALITY REQUIRES that capture this intuitive idea and thus make the principles of charity identified above come out as true.

Judgements about what rationality permits or requires, on this view, are judgements about what rationality permits or requires *of some individual*, *given some set of conditions*. In the principles of charity set out above, the relevant individual is the individual who is being interpreted, and the relevant conditions are all the facts about their behaviour and environment. The relevant conditions can be understood using the notion of a *modal base*—a function from W to $\wp(\wp(W))$ that assigns to each possible world a (possibly empty) set of propositions that specify what the relevant conditions are at that world (Kratzer, 1981). So, in the principles of charity set out above, the modal base is a function that assigns to each possible world the set of all the true propositions about the relevant agent's behaviour and environment at that world.

We define a set V of *interpretations*—maximally specific logically consistent ways for things to be that are consistent with the *a priori* truths about rationality identified above, but which may or may not be metaphysically or even deeply epistemically possible.[21] I remain neutral on the question of the metaphysical nature of these entities; it may be useful for now to think of them as sets of propositions. Let us say that, for all w, $w \in V$, if, and only if, the following conditions hold.

1. w is a way for things to be that is *maximally specific*—that is, for all p, either p or $\neg p$ is true at w;

2. w is a way for things to be that is *logically consistent*—that is, there are no p_1, … , and p_n that are true at w such that p_1 & … & p_n is a contradiction; and

3. w is a way for things to be at which all of the *a priori* truths about the conditions under which something would be rational or irrational hold.

---

[21] I do not require that these interpretations be metaphysically or deeply epistemically possible because I find it plausible that the non-mental facts might both metaphysically and epistemically necessitate the mental facts. According to the principles of charity identified above, we form judgements about what rationality requires or permits in order to help us determine what the mental facts are, given the non-mental facts. If the non-mental facts metaphysically and epistemically necessitate the mental facts, then, if judgements about what rationality permits or requires only concerned metaphysically and epistemically possible combinations of mental and non-mental facts, then they would not be suitable for this task.

For every individual, S, we define a partial order $\leq_S$ on V such that, for all w, $v \in V$, $w \leq_S v$, if, and only if, the proportion of the mental states and behaviours that S exhibits at w that are rational is at least as great as the proportion of the mental states and behaviours that S exhibits at v that are rational. In other words, one interpretation is at least as 'good' as another, with respect to some agent, just in case the extent to which that agent's behaviours and mental states are rational according to the first interpretation is at least as great as the extent to which each of those agents' behaviours and mental states are rational according to the second interpretation.

We are now in a position to provide the analyses of RATIONALITY PERMITS and RATIONALITY REQUIRES. To judge that rationality permits of S that p, given some relevant conditions, is to judge that it is the case that p at every interpretation at which those conditions hold that is 'best' according to $\leq_S$.[22] Similarly, to judge that rationality permits of S that p given some relevant set of conditions is, roughly, to judge that, at some such interpretation that is 'best' according to $\leq_S$, it is true that p.[23]

2.2.2.4 Wide- and Narrow-Scope Rational Principles

One benefit of this analysis is that it is able to make sense of both the wide-scope and narrow-scope formulations of rational requirements that are discussed in the literature. In the literature on rational requirements, there is dispute over whether such principles are best formulated as wide-scope requirements or narrow-scope requirements (Broome, 1999; 2007; Dancy, 2000b; Wallace, 2001; Schroeder, 2004; 2009; Kolodny, 2005; Brunero, 2010; 2012; Way, 2010a; 2011; Lord, 2011; 2013).

---

[22] More precisely (and to accommodate the possibility that there may be no interpretations that are 'best'): for all w, f, S, p, and U, to judge that, at w, given modal base f, rationality requires of S that p *just is* to judge that, where U is the set of all interpretations in V at which all of the conditions in $f(w)$ hold, for every $v \in U$, there is some $u \in U$ such that $u \leq_S v$ and, for all $z \in U$ such that $z \leq_S u$, p is true at z (cf. Kratzer, 1981).

[23] More precisely (and, again, to accommodate the possibility that there may be no interpretations that are 'best'): for all w, f, S, p, and U, to judge that, at w, given modal base f, rationality permits of S that p *just is* to judge that, where U is the set of all interpretations in V at which all of the conditions in $f(w)$ hold, some $u \in U$ is such that, for all $v \in U$ such that $v \leq_S u$, there is some $z \in U$ such that $z \leq_S v$ and p is true at z (cf. Kratzer, 1981).

Let us use the Instrumental Principle as an example. This is the principle that requires that one intend to do something if one believes that it is a necessary means to something else that one intends to do. This principle can be formulated in two different ways.

(Wide-Scope Instrumental Principle) For all *S*, *t*, φ, and ψ, rationality requires of *S* that, if, at *t*, *S* intends to φ and believes that ψ-ing is a necessary means to φ-ing, then, at *t*, *S* intends to ψ.

(Narrow-Scope Instrumental Principle) For all *S*, *t*, φ, and ψ, if, at *t*, *S* intends to φ and believes that ψ-ing is a necessary means to φ-ing, then rationality requires of *S* that, at *t*, *S* intends to ψ.

An analysis of the kind presented above can easily accommodate these two ways of formulating rational principles (Worsnip, 2015). The kinds of wide-scope rational requirements that are discussed in the literature are best interpreted as principles about what rationality *unconditionally* requires or permits—that is, principles about what rationality requires or permits given a modal base that assigns to each world an empty set of conditions. The kinds of narrow-scope rational requirements that are discussed in the literature, on the other hand, are best interpreted as principles about what rationality requires or permits given only the conditions provided by the antecedent of the relevant conditional—that is, principles about what rationality requires or permits given a modal base that assigns to each world an empty set of conditions, to which the antecedent of the conditional is then added (Kratzer, 1986).

2.3 Normativity and Rationality

In the previous chapter, I argued that the central normative concepts MAY, MUST, and OUGHT can ultimately be analysed in terms of fitting-response concepts. In the next section, I will argue that fitting-response concepts can themselves be analysed in terms of the concept RATIONAL. This view thus amounts to what might be called a 'rationality-first' account of normativity. A rationality-first account of normativity is an account of normativity according to which normative facts are ultimately explained in some sense by facts about rationality. The rationality-first approach encompasses certain response-dependent views in the meta-ethical literature, according to which normative facts are ultimately explained by facts about the *mental states* of fully

rational agents (see Smith, 1994, for example), as well as certain constructivist views, according to which normative facts are ultimately explained by facts about fully rational processes of *reasoning* (Korsgaard, 1996; 2008; 2009; Street, 2008; 2010). According to the view that I am defending here, normative facts are ultimately explained by facts about rationality in the sense that all normative concepts can ultimately be analysed in terms of the concept of rationality.

Rejecting the rationality-first approach to normativity commits one either to the view that facts about rationality are ultimately explained by some class of normative facts, or to the view that facts about rationality are entirely distinct from normative facts. The most prominent example of the former is the view that facts about rationality are ultimately explained by facts about *reasons* (Kiesewetter, 2017; Lord, 2018). According to accounts of this kind, for something to be required or prohibited by rationality *just is* for one's reasons to count decisively in favour of or against it. One of the main challenges for this kind of view is that it is difficult to identify the reasons that count decisively against holding certain combinations of attitudes that rationality prohibits. For example, rationality prohibits simultaneously holding contradictory intentions, but sometimes one's reasons provide sufficient support both for intending to do something and for intending not to do it. The prohibition against holding contradictory intentions thus cannot be explained in these cases by the fact that one's reasons count decisively against either one of those intentions. Advocates of the reasons-first approach face the challenge of identifying the reasons that count decisively against this combination of attitudes in spite of the fact that there are no reasons that count decisively against either individual attitude. (See Kiesewetter, 2017, ch. 10 and Lord, 2018, ch. 2 for two prominent attempts to deal with this challenge.)

The alternative to this kind of view is the view that facts about rationality are entirely distinct from normative facts: facts about rationality are simply facts about coherence, and whether coherence has any kind normative significance is a purely contingent matter. Examples of this approach can be found in the work of Broome (2005; 2007; 2013) and Kolodny (2005; 2007; 2008). The main challenge for this kind of view is to explain the intuitively plausible thought that rationality and normativity are intimately linked. The rationality-first approach avoids both of these challenges: on the

rationality-first approach, there is no need to account for the irrationality of holding contradictory intentions in terms of the substantive normative reasons not to do so, and yet there is still an intimate link between rationality and normativity, in so far as normative facts are ultimately explained by facts about rationality.

One of the main advantages of understanding normative judgements as judgements about rationality is that it explains why normative judgements are subject to Rationalism and Weak Hume's Law. Suppose that normative judgements can all be understood as judgements involving the concept RATIONAL, as described above, in conjunction with other non-normative concepts. This would explain why Rationalism is true. Recall that judgements about what is rational or irrational are themselves subject to an analogue of Rationalism. If the presence of the concept RATIONAL is what sets normative judgements apart from non-normative judgements, and any proposition about what is rational or irrational can be known *a priori* given sufficient information about relevant mental facts, behavioural facts, and environmental facts, then it is reasonable to suppose that any normative judgement can be known *a priori* given sufficient information about the non-normative facts—where these facts include mental facts, behavioural facts, environmental facts, and more, but do not include any facts that have to do with what is rational or irrational.

Similarly, if normative judgements could all be understood as judgements about rationality, then this would explain why Weak Hume's Law is true. I have just argued that judgements about what rationality permits or requires are judgements that have to do with how certain scenarios compare with respect to the extent to which some individual's behaviours and mental states are rational. It is difficult to see how one could rationally base a judgement of this kind on some other judgements without relying on some further judgement about rationality. In the next section, I will argue that judgements involving fitting-response concepts *just are* judgements about what rationality requires or permits. In so far as judgements of this kind cannot be rationally based on a set of judgements that does not include some judgement about rationality, then, it follows that, if normative judgements just are judgements about rationality, then Weak Hume's Law is vindicated.

2.4 Fitting-Response Concepts

In the remainder of this chapter, my focus will be on fitting-response concepts such as *desirable*, *admirable*, *trustworthy*, *shameful*, and *fearsome*. I will argue that judgements involving these concepts (or at least a significant subset of them) can be understood as judgements about what rationality permits or requires. I begin by setting out three desiderata for an analysis of fitting-response concepts. First, the analysis should validate the central rational principles that govern how judgements about fitting-responses relate to each other. Secondly, it should make sense of the connection between fittingness and rationality. And, thirdly, it should make sense of the distinction between perspective-independent and perspective-dependent fittingness. I then provide an analysis of fitting-response concepts in terms of the concept RATIONALITY PERMITS that is inspired by Rabinowicz's (2008; 2012; 2017) accounts of value and probability relations and satisfies these desiderata. I conclude by highlighting an important respect in which this account is incomplete.

2.4.1 Desideratum 1: Validates the Rational Principles that Govern Judgements Involving Fitting-Response Concepts

There are a number of distinctive rational principles that govern how judgements involving fitting-response concepts relate to each other. An analysis of fitting-response concepts must validate these principles. I set out a number of these principles below. In what follows, I will use '*A*-able', '*B*-able', and '*C*-able' as placeholders for fitting-response concepts.

2.4.1.1 More

First, MORE *A*-ABLE picks out a relation that appears to have the properties of a strict partial order.

Asymmetry

x is more *A*-able than *y* is *B*-able.

It is not the case that *y* is more *B*-able than *x* is *A*-able.

Transitivity

x is more *A*-able than *y* is *B*-able; *y* is more *B*-able than *z* is *C*-able.

---

x is more *A*-able than *z* is *C*-able. (?)

The rational principles expressed by the instances of Transitivity are intuitively plausible, but are controversial because some believe that they are vulnerable to counterexamples, such as those generated by spectrum arguments (Temkin, 2012). Although these arguments are typically used to motivate the view that the better-than relation is not transitive, is natural to suppose that, if they establish that the better-than relation is not transitive, then they also establish that relations such as the more-desirable-than relation are not transitive either. Consider the following case, for example. You are presented with a list of *n* painful experiences. The first experience on the list is a month of non-stop excruciating torture. Each successive painful experience on the list is slightly less intense than the one immediately preceding it, but significantly longer—indeed, sufficiently longer that it is rational to judge that its more intense predecessor is more desirable than it is. The *n*th item on the list is a painful experience that is not very intense at all—say, the pain of a hangnail—that lasts for a lifetime.

One might think that, regarding this case, it is rational simultaneously to judge, for each $1 \leq m < n$, that experience *m* is more desirable than experience *m*+1, *and* that it experience 1 is not more desirable than experience *n*, since a lifetime of mild hangnail pain seems preferable to even just a month of non-stop excruciating torture. If this combination of judgements is rationally permitted and Transitivity is valid, however, then rationality also permits holding these judgements plus the judgement that experience 1 is more desirable than experience 3. And if this expanded combination of judgements is rationality permitted, then rationality also permits holding these judgements plus the judgement that experience 1 is more desirable than experience 4.

This line of reasoning will eventually bring us to the view that rationality permits that one hold a combination of judgements that includes both the judgement that experience 1 is more desirable than experience *n and* that the judgement that this is not the case. But this cannot be so, since rationality does not permit one to be in a

state that involves holding contradictory judgements. Thus, if this line of reasoning is correct, then this apparently plausible Transitivity schema is in fact invalid.

Despite the ingenuity of this argument, denying the validity of Transitivity is a very controversial position. To appreciate just how controversial it is, it is worth noting that the invalidity of Transitivity for fitting-response concepts entails the invalidity of a very plausible general transitivity schema for the concept MORE.

Transitivity: MORE

x is more *F* than *y* is; *y* is more *F* than *z* is.

---

x is more *F* than *z* is. (?)

2.4.1.2 Equal

Next, EQUALLY *A*-ABLE picks out a relation that has the properties of an equivalence relation.

Reflexivity

-

---

x is equally as *A*-able as *x* is *A*-able.

Symmetry

x is equally as *A*-able as *y* is *B*-able.

---

y is equally as *B*-able as *x* is *A*-able.

Transitivity

x is equally as *A*-able as *y* is *B*-able; *y* is equally as *B*-able as *z* is *C*-able.

---

x is equally as *A*-able as *z* is *C*-able.

2.4.1.3 At Least

Intuitively, the concept AT LEAST AS *A*-ABLE *just is* the concept MORE *A*-ABLE OR EQUALLY AS *A*-ABLE. It follows that AT LEAST AS *A*-ABLE picks out a relation that appears to have the properties of a quasi order—reflexivity and transitivity.

Reflexivity

-

---

x is at least as *A*-able as *x* is *A*-able.

Transitivity

x is at least as *A*-able as *y* is *B*-able; *y* is at least as *B*-able as *z* is *C*-able.

---

x is at least as *A*-able as *z* is *C*-able. (?)

The rational principles expressed by the instances of this schema are *prima facie* plausible, but, again, are controversial in the case of some fitting-attitude concepts in view of the purported counterexamples generated by spectrum arguments (Temkin, 2012). Again, to appreciate just how controversial denying Transitivity is, it is worth noting that the invalidity of Transitivity for fitting-response concepts entails the invalidity of a very plausible schema for the concept AT LEAST.

Transitivity: AT LEAST

x is at least as *F* as *y* is; *y* is at least as *F* as *z* is.

---

x is at least as *F* as *z* is. (?)

2.4.1.4 On a Par

Parity is a relation that can hold between the degrees to which things are fitting objects of attitudes, and has the properties of reflexivity and symmetry.

Reflexivity

-

The degree to which *x* is *A*-able is on a par with the degree to which *x* is *A*-able.

Symmetry

The degree to which *x* is *A*-able is on a par with the degree to which *y* is *B*-able.

The degree to which *y* is *B*-able is on a par with the degree to which *x* is *A*-able.

In the recent philosophical literature on the topic, parity is often taken to be a relation that holds between two items with respect to some quantity only if they do not possess an equal amount of it and neither of them possesses more of it than the other. For example, Chang (2016, §5) writes that, with respect to their *goodness*, '[i]tems are on a par when they are comparable, but one is not better than, worse

than, or equally good as the other'. But this is not true of the relation we typically refer to by the expression 'on a par'. First, it seems incorrect to say that, if two things are on a par with respect to some quantity, then they cannot possess an equal amount of it. When it does not simply mean *equal*, 'on a par' typically means something like *roughly equal*, *comparable*, or *in the same ballpark*. Our ordinary usage of the expression 'on a par' thus suggests that, if two items possess an equal amount of some quantity, then they are trivially on a par with respect to that quantity, since 'exactly equal' implies 'roughly equal'.

EQUAL implies ON A PAR

x is equally as *A*-able as *y* is *B*-able.

The degree to which *x* is *A*-able is on a par with the degree to which *y* is *B*-able.

Secondly, it seems incorrect to say that, if two things are on a par with respect to some quantity, then it cannot be the case one of them possesses more of it than the other. It seems perfectly coherent, for example, to say something like, 'the Toyota only costs five dollars more than the Honda, so they are on a par when it comes to price'. It seems to me that the relation that philosophers have been discussing in the recent literature is not parity as such, but rather *mere* parity—the relation that two things bear to each other with respect to some quantity when the amount of that quantity that each possesses is neither more than, nor equal to, but nonetheless still on a par with, the amount that the other possesses.

2.4.2 Desideratum 2: Vindicates the Connection Between Fittingness and Rationality

Intuitively, there is a connection between something's being a fitting object of some response, on the one hand, and its being rational for a suitably situated agent to respond to that object in that way, on the other. Suppose, for example, that Sam is admirable. If Sam is admirable, then that is because of certain other properties that Sam has—Sam may be particularly kind, generous, or wise, for example. If I knew nothing about Sam, then it would not be rational for me to admire them. But, if I knew about how kind, generous, or wise Sam is, then it is natural to suppose that it would be rational for me to admire them, based on that knowledge. It is plausible, then, that the facts in virtue of which Sam is admirable are the same as the facts on the basis

of which it would make sense for me to admire Sam, if I knew them. The same point can be made in terms of degrees of admirableness and degrees of admiration. Suppose that Sam is more admirable than Tom is, in virtue of the fact that, while Sam is kind, generous, and wise, Tom is mean, stingy, and foolish. If I knew nothing about Sam or Tom, then it would not be rational for me to admire Sam more than I admire Tom or to admire Tom more than I admire Sam. If I knew all the facts about Sam and Tom, however, then I would know all about Sam's kindness, generosity, and wisdom, and Tom's meanness, stinginess, and foolishness. If I knew all of these facts about Sam and Tom, then it seems to me that I would be rationally required to admire Sam more than I admire Tom. Our second desideratum for an analysis of fitting-response concepts, then, is that it make sense of this intuitive connection between fittingness and rationality.

2.4.3 Desideratum 3: Makes Sense of the Distinction between Perspective-Independent and Perspective-Dependent Fittingness

Some fitting-response concepts are perspective-independent. That is, there are some fitting-response concepts such that it is *a priori* that whether an object is a fitting object of the relevant response depends solely on facts about that object, and not on any facts about any potential subject of that response. ADMIRABLE, for example, appears to be a perspective-independent fitting-response concept. To use the example from the last section, Sam's being admirable depends on the fact that Sam is kind, generous, or wise, and not, for example, on anything that anyone happens to believe about Sam. Other fitting-response concepts, by contrast, are perspective-dependent. DESIRABLE, for example, is plausibly a perspective-dependent fitting-response concept. We seem to be able to make judgements about what is desirable relative to some relevant body of information—for example, relative to the *facts*, or relative to some body of *evidence*—and also from a particular point of view. For example, we can make judgements about what is desirable from the point of view of morality, from the point of view of self-interest, or from the point of view of a supporter of the English cricket team. An analysis of fitting-response concepts must be able to make sense of this distinction between perspective-independent and perspective-dependent fitting-response concepts.

2.4.4 Analysis of Fitting-Response Concepts in Terms of Rationality Permits

In this section, I provide an analysis of fitting-response concepts in terms of the concept RATIONALITY PERMITS. I begin by discussing Rabinowicz's (2008; 2009; 2012; 2017) accounts of value and probability relations. I point out some problems with Rabinowicz's approach and propose some friendly amendments. I then provide an analysis of fitting-response concepts based on this amended version of Rabinowicz's view. I argue that this analysis satisfies our three desiderata. First, it validates the rational principles that govern how judgements involving fitting-response concepts relate to each other. Secondly, it makes sense of the connection between fittingness and rationality, since fitting-response concepts are analysed in terms of the concepts RATIONALITY PERMITS and RATIONALITY REQUIRES. And, finally, in so far as our account of RATIONALITY PERMITS and RATIONALITY REQUIRES is able to make sense of the idea of judgements about what rationality permits or requires, conditional on the fact that certain conditions hold, the resulting analysis of fitting-response concepts is well placed to make sense of both perspective-independent and perspective-dependent fitting-response concepts. I close by showing that the analysis is incomplete, however, and is thus at best an approximation of the correct analysis, since it cannot make sense of comparisons with regard to the fittingness of responses at different worlds, or from different perspectives. Addressing these problems is a task for future work.

Rabinowicz (2008; 2009; 2012; 2017) has provided an elegant account of value and probability relations, according to which differences in value and probability can be understood in terms of differences in the degree to which one is rationally required to favour or believe something, respectively. Let us use Rabinowicz's (2012) account of value relations as an example. Rabinowicz's informal analysis of these notions is roughly as follows. Necessarily, for all *x* and *y*:

- *x* is better than *y* is, if, and only if, favouring *x* more than *y* is rationally required;
- *x* and *y* are equally good, if, and only if, favouring *x* and *y* equally is rationally required;

- *x* is at least as good as *y* is, if, and only if, favouring *x* at least as much as *y* (that is, favouring *x* more than *y* or favouring *x* and *y* equally) is rationally required[24];
- *x* and *y* are on a par, if, and only if, favouring *x* more than *y* is rationally permitted and favouring *y* more than *x* is rationally permitted.[25]

There are three problems with this account. The first two have to do with Rabinowicz's treatment of parity. I pointed out earlier that 'on a par' typically means something like 'roughly equal', and so it is natural to think that 'equal' implies 'on a par'. On Rabinowicz's account, however, it is impossible for two things be both equally good and on a par. This seems counterintuitive. I also pointed out earlier that it seems possible for one thing to have more of some quantity than another thing does, even though they are on a par with respect to that quantity. This is plausibly so in cases in which one thing is better than another, despite the fact that both are 'good enough'. On Rabinowicz's account, however, it is impossible for one thing to be better than, but nonetheless on a par with another. The third problem with Rabinowicz's account has to do with his treatment of the relation of being at least as good. It is very natural to suppose that one thing is at least as good as another just in case either the former is better than the latter, or they are equally good. This is not the case on Rabinowicz's account, however. According to Rabinowicz, if favouring *x* more than *y* is permitted, and favouring *x* and *y* equally is permitted, but favouring *y* more than *x* is not permitted, then *x* is at least as good as *y* is, but *x* is neither better than, nor equally as good as, *y* is.

In response to these problems, I propose the following account, which retains the spirit of Rabinowicz's account, but departs from it in the details. According to this account, it is necessary that, for all *x* and *y*:

---

[24] I take this from Rabinowicz (2008); Rabinowicz (2012) does not discuss the *at least as good* relation.

[25] I set aside Rabinowicz's treatment of incomparability here. According to Rabinowicz, two items are incomparable just in case rationality requires neither favouring them equally nor favouring one more than the other.

- *x* is better than *y* is, if, and only if, favouring *x* more than *y* is rationally permitted, and favouring *x* at least as much as *y* is rationally required;
- *x* and *y* are equally good, if, and only if, favouring *x* and *y* equally is rationally required;
- *x* is at least as good as *y* is, if, and only if, favouring *x* at least as much as *y* is rationally required;
- *x* is on a par with *y*, if, and only if, favouring *x* and *y* equally is rationally permitted.

It is easy to see that, on this account, being equally good entails being on a par, since, if two things are equally good, then favouring them equally is rationally required, and thus rationally permitted, which is all it takes for them to be on a par. Similarly, on this account, one thing can be better than another, despite them being on a par. This is so when favouring the first thing at least as much as the second thing is required, favouring the first thing more than the second thing is permitted, and favouring them each equally is permitted. Finally, on this account, something is at least as good as something else just in case the first thing is better than the second thing, or they are equally good.

Based on this amended version of Rabinowicz's informal analysis of value relations, I propose that fitting-response concepts more generally can be analysed in terms of the concepts RATIONALITY PERMITS and RATIONALITY REQUIRES. Thus, for all *x*, *y*, *A*, and *B*:

- To judge that *x* is more *A*-able than *y* is *B*-able *just is* to judge that, for all *S*, rationality permits of *S* that they *A x* more than they *B y*, and rationality requires of *S* that they *A x* at least as much as they *B y*;
- To judge that *x* is equally as *A*-able as *y* is *B*-able *just is* to judge that, for all *S*, rationality requires of *S* that they *A x* equally as much as they *B y*;
- To judge that *x* is at least as *A*-able as *y* is *B*-able *just is* to judge that, for all *S* rationality requires of *S* that they *A x* at least as much as they *B y*; and
- To judge that the degree to which *x* is *A*-able is on a par with the degree to which *y* is *B*-able *just is* to judge that, for all *S*, rationality permits of *S* that they *A x* equally as much as they *B y*.

In the same way as the amended version of Rabinowicz's account presented above makes sense of the different types of value relations, this analysis validates the schemata listed above regarding how judgements involving fitting-response concepts relate to each other. One benefit of this view in this regard is that it is capable of accommodating the controversy over whether the transitivity schemata for MORE *A*-ABLE and AT LEAST AS *A*-ABLE are valid. In so far as it is controversial whether spectrum cases provide counterexamples to the transitivity of the *more desirable than* and *at least as desirable as* relations, it seems no less controversial whether rationality requires that agents' desires and preferences in these cases be transitive (cf. Rabinowicz, 2017, n. 10).

While someone who denies transitivity for a concept like MORE DESIRABLE or AT LEAST AS DESIRABLE can accept this analysis of fitting-response concepts, if they also accept the analysis of RATIONALITY REQUIRES set out above, then they are committed to denying transitivity for the concepts DESIRES MORE and DESIRES AT LEAST AS MUCH. If transitivity fails for AT LEAST AS DESIRABLE, for example, then it is possible for there to be some *x*, *y*, and *z* such that rationality permits simultaneously judging that *x* is at least as desirable as *y* is, *y* is at least as desirable as *z* is, but *x* is not at least as desirable as *z* is. If the analysis of RATIONALITY REQUIRES that I have presented is correct, then this can only be so if the following propositions are jointly consistent: <*S* desires *x* at least as much as *y*>, <*S* desires *y* at least as much as *z*>, and <*S* does not desire *x* at least as much as *z*>. But this is only possible if the desires-at-least-as-much relation is not transitive. One might think that this is not such a cost for someone who is willing to give up transitivity for AT LEAST AS DESIRABLE, since, as discussed above, they are already committed to denying transitivity for AT LEAST more generally anyway.

Nevertheless, it is true that, given the analysis of fitting-response concepts that I have provided, denying transitivity for MORE DESIRABLE and AT LEAST AS DESIRABLE while accepting transitivity for DESIRES MORE and DESIRES AT LEAST AS MUCH would require a different analysis of RATIONALITY REQUIRES and RATIONALITY PERMITS. In particular, it would require an analysis of these concepts that does not validate the following schema.

Rationality requires that *p*; rationality requires that *q*.

Rationality requires that *r*

(where *p* & *q a priori* entails *r*).

To the extent that denying transitivity for MORE DESIRABLE and AT LEAST AS DESIRABLE while accepting transitivity for DESIRES MORE and DESIRES AT LEAST AS MUCH is a reasonable position to adopt, then, this is a cost of the package of views I have presented in this chapter.

This analysis also makes sense of the connection between fittingness and rationality, in so far as fitting-response concepts are analysed in terms of the concepts RATIONALITY PERMITS and RATIONALITY REQUIRES, and so satisfies our second desideratum. This feature of the analysis means that it is also able to make sense of the distinction between perspective-independent and perspective-dependent fitting-response concepts, and thus to satisfy our third desideratum. This is in virtue of the fact that the analyses of RATIONALITY PERMITS and RATIONALITY REQUIRES provided above can make sense of judgements about what rationality permits or requires, given that certain conditions hold. Perspective-independent fitting-response concepts can thus be analysed in terms of what rationality permits of someone, given that they have full information about, and/or acquaintance with, the relevant objects. For example, the judgement that Sam is more admirable than Tom is can be understood as the judgement that for all *S*, *given that S believes all of the relevant propositions that are in fact true of Sam and Tom*, rationality permits of *S* that they admire Sam more than Tom, and rationality does not permit of *S* that they admire Tom more than Sam.

Perspective-dependent fitting-response concepts, in turn, can be analysed in terms of what rationality permits of someone, given that they occupy some relevant perspective—for example, given that they possess a certain body of evidence, given that they satisfy certain moral requirements, or given that they support the English cricket team. Thus, the judgement that, given the available evidence, it is more desirable, from the perspective of a supporter of the English cricket team, that Ben Stokes play in the upcoming Ashes series than it is that he not play, can be understood as the judgement that, for all *S*, *given that S supports the English cricket team, and possesses the available evidence*, rationality permits of *S* that *S* desire

that Stokes play more than that he not play, and rationality does not permit of *S* that *S* desire that Stokes not play more than that he play. I conclude, then, that this analysis satisfies all three of our desiderata.

I have analysed fitting-response concepts in terms of the requirements and permissions of *rationality*. One might wonder whether it makes more sense for some fitting-response concepts to be analysed in terms of some other source of requirements and permissions instead. For example, perhaps the concept MORALLY ADMIRABLE should be analysed not in terms of what *rationality* requires, but rather in terms of what *morality* requires.[26] According to this analysis, then, one person is at least as morally admirable as another just in case *morality* requires that any suitably situated agent admire the former at least as much as the latter.

This is a plausible suggestion, and modifying the analysis to incorporate different sources of requirements should not present any major problems. I do not think that the analysis I have presented requires this kind of modification, however, given that it already has the resources to make sense of the extent to which a response is fitting *from a particular perspective*. For example, the analysis I have presented can make sense of judgements about how admirable something is *from the perspective of morality*. It is natural to think that judgements about the extent to which someone is morally admirable *just are* judgements about the extent to which they are admirable from the perspective of morality. According to the analysis presented here, the judgement that one person is at least as admirable as another, from the perspective of morality, amounts to the judgement that, given that they are fully acquainted with both people *and they satisfy certain moral constraints*, any agent is rationally required to admire the former at least as much as the latter. That is, the former is at least as admirable as the latter, from the perspective of morality, just in case one is committed to admiring the former at least as much as the latter in so far as one occupies the moral point of view (cf. Street, 2010, pp. 367-368; Southwood, 2018, §3). In fact, this seems to me to be a promising analysis of the concept MORALITY REQUIRES: for all *S* and φ, to judge that morality requires that *S* φ *just is* to judge that, given that *S* satisfies relevant moral constraints, rationality requires that *S* φ.

---

26 I am grateful to an anonymous examiner for raising this objection.

In chapter one, I drew attention to the difference between the deontic MUST, which can be thought of as expressing a kind of *requirement*, and OUGHT, which can be thought of as expressing a kind of *recommendation*. In this chapter I have analysed fitting-response concepts in terms of what rationality permits or *requires*. One might wonder whether these analyses could be improved by couching them in terms of what rationality *recommends* rather than what rationality *requires*. For example, according to my analysis, one person is at least as admirable as another just in case rationality *requires* that any suitably situated agent admire the former at least as much as the latter. One might think that this is too strong, and that for one person to be at least as admirable as another it suffices that rationality merely *recommends* that any suitably situated agent admire the former at least as much as the latter, even if rationality *permits* that they admire the latter more.[27]

My reason for analysing fitting-response concepts in terms of what rationality permits and *requires* is that this allows the analysis to satisfy our second desideratum. The second desideratum is that the analysis explain the fact that, if one person is, for example, at least as admirable than another, then it would be irrational for anyone who knew enough about them to admire the latter more than the former. In other words, such an agent would be rationally *required* to admire the former at least as much as the latter. But one might wonder whether this proposed connection between fitting attitudes and rationality is too strong. Perhaps, if one person is at least as admirable as another, then, while it would not be *irrational* for someone who knows all about their admirability-making features to admire the latter more than the former, rationally nonetheless *recommends* that they not.

I admit that I personally find the stronger thesis plausible. If, knowing everything about two people, I can have some rational basis for admiring one rather than the other, it seems strange to insist that the latter is nonetheless at least as admirable as the former is. Nevertheless, if the weaker thesis turns out to be true, then recasting the analysis in terms of what rationality recommends, rather than what rationality requires, should not pose any great problems. This would call for an account of the

27 I am grateful to an anonymous examiner for this suggestion.

difference between the requirements and recommendations of rationality. I will not pursue this task here.

This analysis goes beyond Rabinowicz's by allowing for comparisons between different types of fitting-attitude properties. One of the advantages of this analysis is that it can make sense of when such comparisons make sense and when they do not. For example, while the judgement that Sam is more admirable than she is execrable seems to make sense, it is not clear that the judgement that it is more permissible than it is credible that the Earth is over four billion years old makes sense. The analysis I have presented can account for this difference. On this analysis, the former judgement amounts to a comparison between the degree of admiration and the degree of execration that it is rational to direct toward Sam. Thus, to the extent that it makes sense to compare degrees of admiration and degrees of execration, this judgement is intelligible. By contrast, on this analysis, the latter judgement amounts to a comparison between the degree of permission and the degree of credence that it is rational to direct toward the proposition that the Earth is over four billion years old. It is not clear that such a comparison is intelligible, since permission seems to be measured on a scale that has only two degrees—permitted and not permitted—whereas credence is measured on a different scale whose degrees fall anywhere between zero and one. The fact that it is difficult to make sense of comparisons between degrees of permission and degrees of credence thus explains why it is difficult to make sense of comparisons between degrees of permissibility and degrees of credibility.

2.4.4.1 Directions for Future Research

Although this analysis of fitting-response concepts is attractive, it is incomplete. I have only provided an account of judgements that compare the fittingness of responses at a single possible world and from a single perspective. For example, the analysis only applies to judgements like the judgement that Sam is at least as admirable as Tom is deplorable, from the perspective of morality. According to this account, to judge that this is so is, roughly, to judge that anyone who occupies the moral point of view and knows all of the relevant facts about the actual world is rationally required to admire Sam at least as much as they deplore Tom. I have not provided an account of judgements that involve comparisons between the fittingness

of responses at different worlds. For example, I have not provided an account of judgements like the judgement that Sam is at least as admirable as they would have been had they donated less money to charity. This judgement compares the degree to which Sam is admirable at the actual world with the degree to which Sam is admirable at other nearby possible worlds. Nor have I provided an account of judgements that involve comparisons between the fittingness of responses from different points of view. For example, I have not provided an account of judgements like the judgement that donating to charity is more desirable, from the point of view of morality, than it is from the point of view of pure self-interest.

I have not provided an account of such judgements because doing so would require a significantly more complex analysis. Modifying the account I have provided to cover such judgements would plausibly require an analysis that involves comparisons between the degrees of the relevant responses that are rationally permitted of someone who occupies each perspective. For example, an account of the judgement that Sam is more admirable than they would have been had they donated less to charity would plausibly treat it as a comparison between the degrees of admiration for Sam that are rationally permitted of someone who believes all of the relevant propositions that are true of Sam at the actual world, and the degrees of admiration for Sam that are rationally permitted of someone who believes all of the relevant propositions that are true of Sam at the closest worlds at which Sam donates less to charity. And an account of the judgement that donating money to charity is more desirable from the point of view of morality than it is from the point of view of pure self interest would plausibly treat it as a comparison between the degrees of desire that are rationally permitted of someone who occupies a moral perspective and the degrees of desire than are rationally permitted of someone who occupies a purely self-interested perspective.

A natural first thought about how to implement this idea is to opt for some kind of interval model. For example, to use AT LEAST as an example, following Gert (2004), one might propose something like the following analysis. For all *x*, *y*, *A*, *B*, α, and β, to judge that *x* is at least as *A*-able from perspective α as *y* is *B*-able from perspective β *just is* to judge that the lower bound of the set of degrees to which rationality permits someone who occupies α to *A* *x* is at least as great as the upper bound of the

set of degrees to which rationality permits someone who occupies β to *B y*. This analysis plausibly fails to satisfy our second desideratum, however. This is because it is possible, for example, for rationality to require that anyone who occupies the moral point of view and knows all of the relevant facts about Sam and Tom admire Sam at least as much as they admire Tom, even though there is significant overlap between the sets of degrees to which it is rationally permissible for someone who occupies that perspective to admire Sam and to admire Tom. In this case, it is natural to think that Sam is at least as admirable as Tom is, from the perspective of morality, since anyone who occupies the moral point of view and knows all of the relevant facts about Sam and Tom is rationally committed to admiring Sam at least as much as they admire Tom, but, according to this model, this is false.

In light of this problem, one might try weakening the model. For example, one might propose that, for all *x*, *y*, *A*, *B*, α, and β, to judge that *x* is at least as *A*-able from perspective α as *y* is *B*-able from perspective β *just is* to judge that the upper and lower bounds of the set of degrees to which rationality permits someone who occupies α to *A x* are at least as great as the upper and lower bounds, respectively, of the set degrees to which rationality permits someone who occupies β to *B y* (cf. Rabinowicz, 2008, p. 34). This analysis also fails to satisfy our second desideratum, however. This is because it is possible, for example, for the upper and lower bounds of the set of degrees to which rationality permits that anyone who occupies the moral perspective and knows all of the facts about Sam and Tom to admire Sam to be at least as great as the upper and lower bounds, respectively, of the set of degrees to which rationality permits that anyone who occupies that perspective to admire Tom, even though rationality permits that someone who occupies that perspective admire Tom more than they admire Sam (cf. Rabinowicz, 2008, p. 35). In this case, it is natural to deny that Sam is at least as *A*-able as Tom is, from the perspective of morality, since rationality permits that someone who occupies the moral point of view and knows all of the facts about Sam and Tom admire Tom more than Sam, but, according to the model, this is false.

In order to satisfy our second desideratum, then, a more complex analysis is required. Rather than delving into the details of this more complex analysis here, I will simply acknowledge that the analysis I have presented is incomplete and issue a

promissory note that there is a complete analysis to be had of which the analysis presented here is a special case. Developing this more complex analysis is a task for future research.[28] In the rest of this thesis, I will treat the account I have provided here as a good enough approximation of the full analysis for the purposes for which it will be called into to use in later chapters.

2.5 Conclusion

In this chapter, I have provided an analysis of fitting-response concepts in terms of the concepts RATIONALITY PERMITS and RATIONALITY REQUIRES. Although this analysis is incomplete, it strikes me as a good approximation of the correct analysis, in so far as it otherwise satisfies our three desiderata for an analysis of these concepts. I will therefore freely draw upon on this analysis in chapter four, when explaining why normative judgements are subject to enkratic principles. Before I take on that task, however, in the next chapter I will distinguish a number of different types of enkratic principles to which normative judgements are subject, and I will present and argue for

---

28 Here is a brief sketch of how such an analysis might go. According to the overall picture I have presented in this chapter, judgements that compare degrees of fitting-response properties are judgements that compare the degrees to which an agent has the relevant responses at some or all of the *best interpretations* of that agent. Let us now introduce the idea of a 'hyper-interpretation' of an agent. A hyper-interpretation of an agent is a function that assigns to each possible perspective an interpretation at which that agent occupies that perspective. For every agent *S*, we can then define a partial order $\succcurlyeq_S$ on the set of hyper-interpretations of *S* such that for any hyper-interpretations of *S*, *f* and *g*, $f \succcurlyeq_S g$, if, and only if, $f(\alpha) \geq_S g(\alpha)$, for every perspective α. Judgements that compare degrees of fitting-response properties can then be understood as judgements that compare the degrees to which an agent has the relevant responses at some or all of the *best hyper-interpretations* of that agent. So, to use AT LEAST as an example again, for all *x*, *y*, *A*, *B*, α, and β, to judge that *x* is at least as *A*-able relative to perspective α as *y* is *B*-able relative to β *just is* to judge that, for every agent *S*, at every hyper-interpretation of *S*, *f*, that is ranked highest by $\succcurlyeq_S$, the degree to which *S A*s *x* at $f(\alpha)$ is at least as great as the degree to which *S B*s *y* at $f(\beta)$. For judgements that compare the degrees of fitting-response properties relative to a single perspective, this analysis should turn out to be equivalent to the original analysis in terms of best interpretations. I leave the task of showing that this is so to another time.

an account of reflective agency that will also play a role in chapter four in explaining why normative judgements are subject to such principles.

3 Cognitivism about Reflective Endorsement

In this chapter, I present and defend a thesis that I will call ‘cognitivism about reflective endorsement’. According to an intuitively appealing picture of self-governing agency, agents have the capacity to regulate their own mental states by ‘endorsing’ them. Cognitivism about endorsement is the view that to endorse a mental state *just is* to believe that one is, or will be, in that mental state. Cognitivism thus entails that agents regulate their mental states by self-ascribing them. My aim in this chapter is to show that cognitivism about endorsement is a position that is worthy of being taken seriously. In the next chapter, I demonstrate that combining cognitivism about endorsement with the analyses of normative concepts and the account of rationality presented in previous chapters provides a satisfying explanation of the various ‘enkratic’ principles that apply to normative judgements. Before turning to this account of reflective endorsement, I provide some motivation for developing an account of this kind by looking carefully at these different types of enkratic principles. Enkratic principles are rational principles that apply to the ways in which normative judgements relate to other mental states. Some of these principles, I argue, have to do with the ways in which self-governing agents regulate their own mental states. This will lead us directly to the cognitivist account that I will be proposing.

3.1 Types of Enkratic Principles

The principle known in the literature as ‘Enkrasia’ requires, roughly, that someone intend to do something if they believe that they ought to do it and that is up to them then whether they do it. I argue that there are a number of enkratic principles, which can be categorised along three dimensions. First, enkratic principles can be categorised according to the type of normative judgement and other attitude that are involved. In the next chapter, I argue that, while the traditional formulation of Enkrasia, which links judgements about what one *ought* to do to intentions, is too strong, there are other enkratic principles that link different types of normative judgement to other types of attitude. For example, there are enkratic principles that link judgements about what one *must* do to intentions, and judgements about what is *desirable* (from one’s own perspective) to desires.

Secondly, enkratic principles can be categorised according to whether they are requirements, permissions, or prohibitions. Enkratic requirements simply require that the relevant normative judgement be accompanied at a time by its counterpart attitude. Enkratic permissions permit that the counterpart attitude be held *because* the relevant normative judgement is held. Enkratic prohibitions prohibit that the relevant normative judgement be held or not held *because* the counterpart attitude is held or not held. Broome (2013) holds that enkratic permissions and prohibitions are *basing* permissions and prohibitions, allowing or forbidding that the one mental state be *based on* the other. I argue that this is the wrong way of capturing the kind of dependence at play in these principles. I argue instead that enkratic permissions and prohibitions are principles that constrain the *regulation* of attitudes. According to this view, enkratic permissions allow that the active *formation* or *maintenance* of the counterpart attitude—as opposed to the attitude itself—be based on the relevant normative judgement, while enkratic prohibitions prohibit that the active *formation*, *maintenance*, *giving up* or *withholding* of the normative judgement be based on the presence or absence of its counterpart attitude.

Thirdly, enkratic principles can be divided into what I will call 'positive' and 'negative' enkratic principles. The enkratic principles that have been discussed so far are positive enkratic principles, which concern the rational connection between holding a normative judgement and holding some other attitude. Negative enkratic principles concern the rational connection between regarding some normative question as open and withholding some other attitude. For example, it seems plausible that rationality requires that, if someone is in a state of doubt about whether it is permissible for them to do something, then they not intend to do it. As with positive enkratic principles, negative enkratic principles may be requirements, permissions, or prohibitions.

3.1.1 Enkratic Requirements, Permissions, and Prohibitions

The literature on enkratic principles focuses mainly on a principle called 'Enkrasia', which requires that one intend to do what one judges that one ought to do. In the next chapter, I will argue that this principle is false, but for now I will use it to demonstrate the different types of enkratic principles.

(Enkrasia) For all *S*, *t*, and φ, rationality requires that:

- If:
    - *S* judges at *t* that they themselves ought to φ; and
    - *S* believes at *t* that it is up to them themselves then whether they themselves φ;
- Then:
    - *S* intends at *t* to φ.

Enkrasia is an enkratic requirement. Enkratic requirements require that some normative judgement be accompanied at a time by some counterpart attitude. Enkratic permissions, on the other hand, permit that the counterpart attitude be held *because* the relevant normative judgement is held. For example, it is natural to think that, if Enkrasia is true, then there is also a rational principle that permits that one intend to do something because one judges that one ought to do it.

(Enkratic Permission) For all *S* and φ, rationality permits that *S*'s intention to φ be held *because S* judges that they themselves ought to φ.

Enkratic prohibitions, on the other hand, prohibit that one hold or not hold the relevant normative judgement *because* one holds or does not hold the counterpart attitude. For example, it is natural to think that rationality prohibits that one not judge that one ought to do something simply because one does not intend to do it.

(Enkratic Prohibition) For all *S* and φ, rationality prohibits that *S* judge, or not judge, that they themselves ought to φ *because S* intends, or does not intend, to φ.

3.1.1.1 Broome's Enkratic Basing Permissions and Prohibitions

Broome holds that the type of dependence at play in these enkratic permissions and prohibitions is *basing*: enkratic permissions and prohibitions are basing permissions and prohibitions, which allow or forbid that the (absence of) one attitude be *based on* (the absence of) the other. Thus, Broome's Enkratic Basing Permission is a principle that permits that someone's intending to do something be based on their judging that they ought to do it and believing that it is up to them whether they do it.

(Enkratic Basing Permission) For all *S* and φ, rationality permits that *S*'s intending to φ be based on:

- *S*'s judging that they themselves ought to φ; and
- *S*'s believing that it is up to them themselves whether they themselves φ.

And Broome's Enkratic Basing Prohibition prohibits that someone's not believing that they ought to do something be based on their not intending to do it.

(Enkratic Basing Prohibition) For all *S* and φ, rationality prohibits that *S*'s not believing that they themselves ought to φ be based on their not intending to φ.

Southwood (2016b) argues that Broome's Enkratic Basing Permission is false. His argument starts with the plausible premise that it is rational to base an intention to φ on the belief that one ought to φ (and the belief that it is up to one oneself whether one φs) only if it would be 'apt' to cite those considerations in answer to a question of the form 'why are you φ-ing?' or 'why are you going to φ?'. To put it another way, those considerations must be considerations that, if true, would 'bear positively on the correctness of' that intention (pp. 3420-3421). His second premise is that 'I ought to φ (and it is up to me myself whether I φ)' is not an apt answer to a question of this form. It follows that it is not rational to base an intention to φ on the belief that one ought to φ (and that it is up to one oneself whether one φs); Broome's Basing Permission is therefore false. Broome's (2016) response to this argument is to deny Southwood's second premise. He maintains that 'I ought to φ (and it is up to me myself whether I φ)' is a perfectly apt answer to the question 'why are you φ-ing?'. After all, what could bear more positively on the correctness of intending to do something than the fact that one ought to do it?

I think that Southwood is right to conclude that Broome's Enkratic Basing Permission is false: it is not the case that an intention to do something can be rationally based solely on the belief that one ought to do it and the belief that it is up to one oneself whether one does it. Note that this does not entail that an intention of this kind could not rationally be based on these beliefs in combination with some other mental state(s). For example, it seems plausible enough that rationality permits that someone's intending to do something be based on the belief that they themselves ought to do it, the belief that it is up to them themselves whether they do it, and a

conditional intention to do it if they themselves ought to do it and it is up to them themselves whether they do it. My claim, then, is only that rationality prohibits that an intention of this kind be based *solely* on beliefs of this kind.

Whether Southwood's second premise is true depends on what it means for a consideration to be an 'apt' answer to a question of the form, 'why are you φ-ing?' or 'why are you going to φ?'. I believe that there is a sense of 'apt' according to which this argument is sound. It would be 'apt', in this sense, for someone to cite a consideration in answer to a question of the form 'why are you φ-ing?' or 'why are you going to φ?', if, and only if, that consideration is the kind of consideration that could be a *normative reason* for them to φ. I thus propose the following variation on Southwood's argument.

1. Rationality permits that someone intend to do something solely on the basis of the belief that they themselves ought to do it and the belief that it is up to them themselves whether they do it only if it is possible for the fact that they ought to do it and that it is up to them whether they do it to be a normative reason for them to do it.
2. It is not possible for the fact that someone ought to do something and that it is up to them whether they do it to be a normative reason for them to do it.
3. Therefore, it is not the case that rationality permits that someone intend to do something solely on the basis of the belief that they themselves ought to do it and the belief that it is up to them themselves whether they do it.

The first premise of this argument is an instance of what seems to me to be a plausible general principle regarding the connection between basing and normative reasons: it is rational for someone to base an intention to φ solely on the belief that *p* only if it is possible for the fact that *p* to be a normative reason for them to φ (cf. Dancy, 2000b, p. 103). In other words, it is rational for someone to treat a consideration as counting in favour of some action—by basing an intention to perform that action solely on the belief that that consideration obtains—only if that consideration is the kind of thing that could in fact count in favour of performing that action, normatively speaking. Note that this is quite a weak principle. All it requires is that there be *some* possible circumstances under which the relevant consideration would be a normative reason. It does not require that that consideration be *in fact* a

normative reason, nor that one *believe* that it is a normative reason, nor that one even believe that it is *possible* for it to be a normative reason.

The second premise of the argument is true because normative reasons for someone to do something are facts partly *in virtue of which* they ought to do it (Markovits, 2010; Coates, 2013; Broome, 2013). Suppose, for example, that I ought to go to bed now, and that one of the reasons for me to go to bed now is that, if I didn't go to bed now, then I would be tired tomorrow. It seems to follow that the fact that I ought to go to bed now holds partly *in virtue of* the fact that, if I did not go to bed now, then I would be tired tomorrow. If normative reasons to do something are facts in virtue of which one ought to it, then, since it cannot be the case that the fact that one ought to φ holds (even partly) in virtue of itself, the fact that one ought to φ (and that it is up to one oneself whether one φs) cannot be a normative reason to φ. Thus, since both premises of this argument are true, I conclude, with Southwood, that Broome's Enkratic Basing Permission is false.

Even if one is not persuaded by this argument that Broome's Enkratic Basing Permission is false, there is another reason to believe that there must be a way of understanding enkratic asymmetries other than in terms of simple basing permissions. Suppose that someone believes that it is up to them whether they perform some action, and they have been deliberating about whether to do it. They have identified what they take to be all of the substantive considerations for and against doing it, weighed them against each other, and come to the conclusion that they themselves ought to do it. It is natural to imagine that, in such a case, the person might rationally be moved to form an intention to perform that action by their belief that they ought to do it, but also that the resulting intention is based solely on their beliefs about the non-normative features of that action that they take to count in favour of it. If this common-sense picture is indeed possible, then the person's coming to hold the intention *because* of their belief about what they ought to do cannot be understood simply in terms of that intention's being based on that belief.

3.1.1.2 Enkratic Principles as Constraints on Attitude Regulation

If the Enkratic Permission and Prohibition are not simple basing permissions, then how are we to understand the kind of dependence that is at play in these principles? I

believe that an answer can be found in our common-sense conception of self-regulating or self-governing agency. A self-regulating agent has the capacity to regulate their own judgement-sensitive mental states, by actively *forming* them, *maintaining* them, *giving* them *up*, or *withholding* them, rather than merely passively *acquiring* them, *retaining* them, *losing* them, or *lacking* them. For example, it is part of our common-sense conception of agency that, while we do often simply passively acquire intentions, we also have a capacity to form intentions by actively *making up our minds* about what to do (Korsgaard, 1996; 2008; 2009; Moran, 2001).

I argue that enkratic permissions and prohibitions are principles that constrain this kind of active regulation of judgement-sensitive mental states: enkratic permissions allow that the counterpart attitude be held because it is actively *formed* or *maintained* based on the relevant normative judgement, while enkratic prohibitions prohibit that the relevant normative judgement not be held because it is actively *formed*, *maintained*, *given up* or *withheld* based on the presence or absence of its counterpart attitude. Thus, while Broome is correct that these principles involve basing, they do not concern the permissibility of basing the relevant mental states themselves on the relevant normative judgements, but rather the permissibility of actively *forming*, *maintaining*, *giving up*, or *withholding* those mental states on the basis of those normative judgements. In the next chapter, I will formulate various enkratic permissions and prohibitions as principles that govern the regulation of mental states in this way. In the second half of this chapter, I will give an account of the kind of mental self-regulation to which these kinds of principles apply.

3.1.2 Positive and Negative Enkratic Principles

The enkratic principles discussed so far are what I will call 'positive' enkratic principles. They are principles that assert a rational connection between holding a normative judgement and being in some other kind of mental state. Positive enkratic principles are thus principles that have to do with avoiding a kind of *weakness* of will that Aristotle calls '*propeteia*' (Aristotle, 2009, 7.7). 'Negative' enkratic principles, by contrast, concern the connections between regarding some normative question as open, or being in a state of doubt about some normative question, and *not* being some other kind of mental state. Negative enkratic principles are thus principles that have to do with avoiding a kind of *impetuosity* of will that Aristotle calls '*astheneia*'

(Aristotle, 2009, 7.7). Let us say that someone regards the question of whether some proposition is true as open just in case they are not immediately disposed, upon reflection, to judge that it is true or that it is false.

A negative enkratic requirement, then, is a rational principle that requires that one not be in some mental state if one regards some normative question as open. A negative enkratic permission is a rational principle that permits that one not be in that mental state *because* one regards that normative question as open—more specifically, that the mental state be *withheld* or *given up* based on regarding the normative question as open. And a negative enkratic prohibition is a rational principle that forbids that one regard that normative question as closed *because* one is in that mental state—more specifically, that one not *form* or *maintain* a mental state that constitutes regarding that normative question as closed based on being in that mental state.

3.1.2.1 Deliberation

Negative enkratic principles are interesting because they highlight one of the key roles that normative concepts play in deliberation. Deliberation is an activity that is aimed at resolving some *deliberative question*—for example, the question of what to do—by forming a particular type of attitude. Deliberation also typically involves an attempt to resolve some *normative question*—for example, the question of what one *ought* to do—which seems to be distinct from the relevant deliberative question. It is natural to suppose, for example, that the deliberative question of what to do is distinct from any normative question, such as the question of what one *ought* to do, since the question of what one ought to is resolved through the formation of a belief, whereas the deliberative question of what to do is resolved through the formation of an intention.

I should pause briefly here to note that *noncognitivists* deny this. According to Gibbard (2003), for example, there is no distinction between the deliberative question of what to do and the normative question of what one ought to do.

> As I put my own version of the doctrine, *ought* questions and *reason* questions are by their very nature questions of what to do. … I the chooser don't face two clear, distinct questions, the question what to do and the question what I ought to do. (2003, p. 9)

Gibbard is right to emphasise the very close connection between deliberative questions, like the question of what to do, and normative questions, like the question of what one *ought* to do. There is room to doubt, however, whether the connection between these types of question is so tight as to constitute an identity. For example, it seems possible to have successfully resolved the question of what one *ought* to do without having yet undertaken the kind of practical commitment that results from of having resolved the question of what to do.

There is much more to be said on this issue, but since refuting noncognitivism is beyond the scope of this thesis, I will simply present here what I take to be a plausible alternative to the noncognitivist's view of the relationship between normative questions and deliberative questions. According to this alternative view, rational agents seek to resolve normative questions *in order to* resolve deliberative questions. That normative and deliberative questions are related in this way is suggested by a number of philosophers:

> Deliberation just is the search for and weighing of reasons for acting in order to resolve what to do. (Darwall, 1997, p. 307)
>
> Practical deliberation, as I think of it, is *reasoning about what is best (or satisfactory) to do with a view to making up one's mind about what to do*. … One's aim in deliberation is to make a commitment to a course of action by making a judgement about what is best (or good enough) to do. (Watson, 2003, pp. 175-176)
>
> [W]e have normative concepts … because we have to figure out what to believe and what to do. (Korsgaard, 1996, p. 46)

According to this view, as long as a rational agent regards some normative question as open, they do not hold any attitude that would settle its counterpart deliberative question, and, indeed, their not holding such an attitude may be *based on* their regarding the relevant normative question as open. It is not until they have resolved the relevant normative question that a rational agent is in a position to resolve its corresponding deliberative question. I will explore this idea in more detail, and discuss some possible objections to it, in the next chapter.

3.2 Cognitivism about Reflective Endorsement

I have argued that some enkratic principles have to do with the ways in which self-governing agents regulate their own mental states. I devote the remainder of this chapter to presenting and defending a thesis about what this kind of mental self-regulation consists in that I call 'cognitivism about endorsement'. Endorsement is a mental state by means of which agents actively regulate their own mental states. Cognitivism about endorsement is the view that to endorse a mental state *just is* to believe that one is, or will be, in that mental state. Cognitivism about endorsement thus entails that agents regulate their mental states by self-ascribing them. My aim here is to show that cognitivism about endorsement is a position that is worthy of being taken seriously. In the next chapter, I demonstrate that combining cognitivism about endorsement with the analyses of normative concepts and the account of rationality presented in previous chapters provides a satisfying explanation of the various enkratic principles that apply to normative judgements.

3.2.1 Two Types of Control over Mental States

As discussed above, it is natural to suppose that a self-regulating agent has the capacity actively to *form*, *maintain*, *give up*, and *withhold* certain mental states, as opposed to merely passively *acquiring*, *retaining*, *losing*, or *lacking* them. When an agent actively forms, maintains, gives up, or withholds a mental state, they exercise a kind of control over it that is different to the kind of control that they exercise over their intentional actions. Hieronymi (2006) identifies two different types of control that an agent might exercise over their own mental states. There is 'managerial' or 'manipulative' control, which involves the indirect regulation of one's mental states through voluntary actions that are rationally responsive to one's desires or intentions. And there is 'evaluative' control, which is a kind of direct control that is rationally responsive to an agent's assessment of the normative or rational status of their mental states, rather than to any desires or intentions that bear on them. Evaluative control is the kind of control that I have in mind when I refer to an agent's actively forming, maintaining, giving up, or withholding a mental state. When a rational agent actively forms a belief, for example, their forming that belief is not a voluntary act that is based on their desires and intentions, but rather is based on their assessment of the normative or rational status of that belief.

Some are sceptical that we do in fact actively form, maintain, give up, or withhold mental states in a way that is distinct from passively acquiring, retaining, losing, or lacking them. Strawson (2003), for example, reports that, when he introspects and pays careful attention to the ways in which his mental states come and go, he finds that he is much more passive with respect to them than might be commonly supposed. He concludes that agents only exercise control over their mental states through 'prefatory' or 'catalytic' mental acts such as 'imaging key words or sentences to oneself, rehearsing inferential transitions, refreshing images of a scene', '[shepherding] or [dragooning] one's wandering mind back to the previous thought-content', '[initiating] a kind of actively receptive blanking of the mind', and 'maintaining attention'—actions that are best seen as exercises of managerial rather than evaluative control (pp. 231-232).

To deny that we exercise control over our mental states, however, is to give up an important part of our common-sense conception of rational agency. While it may be true that much of the time we are merely passive with respect to the comings and goings of our mental states, it is nonetheless a core part of our common-sense conception of rational agency that we do sometimes actively regulate our mental states on the basis of our acknowledgement that they are supported by reasons, and thereby exercise a kind of control over them. The sceptic's mistake is to assume that the only kind of control that an agent could have over their mental states is voluntary control. That is why, when the sceptic introspects in order to look for instances of control, managerial control is all they can find. But as long as an alternative notion of evaluative control is intelligible, this sceptical position can be resisted. Making sense of evaluative control is my aim in the rest of this chapter.

3.2.2 Endorsement and the Regulation of Mental States

In what follows, I will give an account of a kind of mental state that I will by stipulation call 'endorsement'. I use this term to refer to whatever it is that plays the regulating role when an agent exercises evaluative control over their own judgement-sensitive mental states. More specifically, endorsement is a state of mind of which the following propositions are true.

For someone to *form* a mental state *just is* for them to come to be in that mental state because they *endorsed* their coming to be in that mental state.

For someone to *maintain* a mental state *just is* for them to continue to be in that mental state because they *endorsed* their continuing to be in that mental state.

For someone to *give up* a mental state *just is* for them to cease to be in that mental state because they did not *endorse* their continuing to be in that mental state.

For someone to *withhold* a mental state *just is* for them not to come to be in that mental state because they did not *endorse* their coming to be in that mental state.

For someone's forming, maintaining, giving up, or withholding a mental state to be based on some set of mental states *just is* for the endorsement or lack of endorsement that partly constitutes their forming, maintaining, giving up, or withholding that mental state to be based on that set of mental states.

I take endorsement to be subject to the following intuitively plausible rational principles. First, rationality requires that we only be in judgement-sensitive mental states that we would endorse, were we to reflect on them.

(Reflective Endorsement Requirement) Rationality requires that someone be in a judgement-sensitive mental state, if, and only if, they are disposed, on reflection, to endorse their being in that mental state then.

Next, rationality permits that one's mental states be sensitive to one's endorsement or lack of endorsement of them. In other words, forming or maintaining a mental state by endorsing it, or giving up or withholding a mental state by not endorsing it, is a rationally permissible way to regulate one's mental states.

(Positive Endorsement Permission) Rationality permits that someone's coming/continuing to be in a judgement-sensitive mental state be *explained by* their endorsing their coming/continuing to be in that mental state.

(Negative Endorsement Permission) Rationality permits that someone's not coming/continuing to be in a judgement-sensitive mental state be *explained by* their not endorsing their coming/continuing to be in that mental state.

I remain neutral here on the particular type of explanatory relation that these principles invoke. For example, I will leave it open whether when someone forms or maintains a mental state by endorsing it, their endorsing the mental state *causes* them to be in that mental state, or whether instead their endorsing the mental state somehow *constitutes* their being in that mental state. This relation should *not* be understood as *basing*, however. When someone forms or maintains an intention, for example, the intention they form should be based in the usual way on their beliefs about the attractive substantive features of the intended action, not on their endorsement of the intention itself.

Although I am officially using 'endorsement' here as a technical term for whatever it is that plays this central role in the active regulation of one's own mental states, I have chosen this word because there does seem to be an intuitive sense in which maintaining, forming, giving up, or withholding a mental state involves endorsing it, in the sense of *ratifying*, or *signing off on* it, in the same way that one might endorse, ratify, or sign off on a proposal. Korsgaard (1996; 2008) uses this term to refer to a closely related phenomenon. She paints a picture of agency according to which a self-governing agent is moved to act by a desire only if they endorse being moved to act by that desire. In other words, she adopts the view, which she attributes to Socrates, that '[t]he soul does not act directly from appetite, but from something that endorses the appetite and says yes to it.' (2008, p. 104). Similarly, on her view, a reflective agent is moved to believe by a perceptual experience only if they endorse being moved to believe by that perceptual experience (2008, p. 4).

I admit that this choice of terminology is not perfectly satisfactory. In the next section, I will argue that actively forming, maintaining, giving up, or withholding a mental state need not involve approving of it, in the sense of desiring that one be in it, or holding a positive normative assessment of it. It is natural, however, to hear 'endorsement' as connoting some kind of approval or positive normative assessment. This need not be so, however. We can separate the question of whether someone endorses, ratifies, or signs off on a proposal, for example, from whether they approve of the proposal, or judge that it meets certain normative standards. Thus, it is possible to imagine someone (perhaps grudgingly or irrationally) endorsing, ratifying, or signing off on a proposal that they do not approve of or of which they do not hold a positive normative

assessment. Separating endorsement from approval is particularly important for the view that I will be defending, since, to anticipate, according to that view, endorsement just is *belief*. In so far as one takes 'endorsement' to connote approval, then, this will make this view sound very implausible, since one needn't approve of all of the things that one believes to be true.

3.2.3 Alternative Accounts of Endorsement

Having identified the mental state in question, before presenting my positive account of endorsement, I will consider two alternative accounts of endorsement that I will call 'rationalism' and 'voluntarism', respectively. According to rationalism about endorsement, to endorse a mental state *just is* to hold some positive normative assessment of it—for example, a judgement that one is rationally committed to being in it, or that one has conclusive reason (of the right kind) to be in it (see Raz, 1999, chs. 1 and 2, for example). The idea here is that when an agent actively forms, maintains, gives up, or withholds a mental state, that mental state is directly based on their normative assessment of it, without the assistance of any additional intermediary state of endorsement.

One reason to reject this view is that, as discussed above regarding the disagreement between Southwood (2016b) and Broome (2016) about Enkratic basing permissions, it is plausible to suppose that it is an agent's *regulating* their mental state—for example, their *forming* or *maintaining* that state—rather than the mental state itself, that is rationally sensitive to their normative judgements. So, with respect to the case under dispute between Southwood and Broome, while it makes sense for an agent's *forming* or *maintaining* an intention to do something to be based on a belief that one ought to do it, it does not make sense for the intention itself to be based on that belief. This suggests that, when an agent is moved to form or maintain a mental state by their positive normative assessment of it or its object, they endorse that mental state *on the basis of* their normative judgement, in which case the endorsement cannot be identical to the normative judgement.

The second problem with the view that endorsement *just is* normative judgement is that it denies the intuitive possibility that an agent might actively form or maintain a mental state in the absence of a positive evaluation of it (Frankfurt, 2004). For

example, it seems possible—though irrational—for an agent who does not judge that preferring *x* to *y* is appropriate or justified—and perhaps even judges that, from their perspective, *y* is preferable to *x*—nonetheless to form a preference for *x* over *y*. If this is in indeed possible, then, since the agent endorses the preference without holding a relevant normative judgement, the endorsement cannot itself be a normative judgement. There are thus good reasons to think that endorsement is not a kind of normative judgement.

According to voluntarism about endorsement, by contrast, to endorse a mental state *just is* to be in a state that is, or is partly constituted by, a non-cognitive state. There are accounts in the literature, for example, according to which to 'endorse' a mental state *just is* to hold some kind of second-order desire (Frankfurt, 1988; 1999; Dworkin, 1988) or intention (Bratman, 1996; 2007) or belief-desire pair (Shoemaker, 1988). While accounts of this kind may well serve some other philosophical purposes, they are not promising as accounts of the kind of endorsement that is involved in the direct evaluative control of mental states, since it does not seem to be a part of our common-sense conception of rational agency that a rational agent's mental states are directly responsive to their non-cognitive states in this way. It is much more plausible that the kind of control that an agent would exercise through this kind of non-cognitive endorsement is managerial control, which, in so far as it involves voluntary action, is the kind of control that is rationally sensitive to these kinds of non-cognitive states.

3.2.4 A Cognitivist Account of Endorsement

We have seen that neither rationalism nor voluntarism is particularly promising as an account of endorsement. In this section, I provide an alternative account of endorsement that I call 'cognitivism', according to which endorsement is simply belief.

(Cognitivism about Endorsement) For all *p*, for someone to *endorse* its being the case that *p just is* for them to believe that *p*.

It follows that to endorse a mental state is simply to *self-ascribe* it: for someone to endorse their then being in some mental state *just is* for them to believe that they themselves are then in that mental state; and for someone to endorse their

coming/continuing to be in some mental state *just is* for them to believe that they themselves will come/continue to be in that mental state. It also follows that an agent's expectations about their mental states play the regulating role identified above in their forming, maintaining, giving up, and withholding mental states.

For someone to *form* a mental state *just is* for them to come to be in that mental state because they *believed* that they would come to be in it.

For someone to *maintain* a mental state *just is* for them to continue to be in that mental state because they *believed* that they would continue to be in it.

For someone to *give up* a mental state *just is* for them to cease to be in that mental state because they did not *believe* that they would continue to be in it.

For someone to *withhold* a mental state *just is* for them not to come to be in that mental state because they did not *believe* that they would come to be in it.

For someone's forming, maintaining, giving up, or withholding a mental state to be based on some set of mental states *just is* for the belief or lack of belief that partly constitutes their forming or maintaining that mental state to be based on that set of mental states.

What is there to be said in favour of cognitivism about endorsement? In the next chapter, I will show that combining cognitivism about endorsement with the analyses of normative concepts and the account of rationality presented in previous chapters provides an explanation of the fact that normative judgements are subject to enkratic principles. This provides the basis for an abductive argument in favour of this combination of views. In this section, I will mount an independent argument for cognitivism about endorsement by showing that self-ascription is subject to rational principles that, if cognitivism about endorsement is true, are equivalent to the three rational principles identified above to which endorsement is subject—the Reflective Endorsement Requirement, the Positive Endorsement Permission, and the Negative Endorsement Permission.

3.2.4.1 Reflective Self-Ascription Requirement

A number philosophers have been attracted to the idea that the requirements of rationality include some kind of self-knowledge requirement (Shoemaker, 1995; Burge, 1996; Smithies, 2016; Fernandez, 2013; Moran, 2001). The following rational principle, which requires that one's own judgement-sensitive mental states not be a mystery or surprise to oneself, strikes me as intuitively plausible, and, if cognitivism about endorsement were true, would be equivalent to the Reflective Endorsement Requirement.

(Reflective Self-Ascription Requirement) Rationality requires that someone be in a judgement-sensitive mental state if, and only if, they are disposed, on reflection, to believe that they themselves are in that mental state then.

One reason to believe that some principle like this is true is that such a principle promises to explain the irrationality of believing Moorean conjunctions. 'Moore's paradox', in its original form, refers to the fact that there are certain propositions that are absurd to assert, despite not containing any contradiction (Wittgenstein, 1953; Moore, 1959). For example, it would be absurd for me to assert either of the following so-called 'Moorean conjunctions'.

> <It is raining and I do not believe that it is raining.>
>
> <It is raining and I believe that it is not raining.>

While the initial discussions of Moore's paradox centred on the absurdity of *asserting* Moorean conjunctions, subsequent discussions have also highlighted the absurdity of *believing* Moorean conjunctions (Shoemaker, 1995). Many are thus now convinced that believing Moorean conjunctions is rationally prohibited.

If a rational principle like the Reflective Self-Ascription Requirement were true, then that would explain why believing Moorean conjunctions is rationally prohibited. For example, suppose, for some *p*, that I believe <*p* and I do not believe that *p*>. Since rationality requires that I believe a conjunction only if I believe each of its conjuncts, rationality requires that, if I believe this Moorean conjunction, then I believe that *p* and I believe that I do not believe that *p*. If we assume that rationality requires that I believe that something is not the case, only if I am not disposed to believe, on

reflection, that it is the case, then it follows that rationality requires that, if I believe the Moorean conjunction, then I believe that *p* and I am not disposed, on reflection, to believe that I believe that *p*. Since the Reflective Self-Ascription Requirement entails that the consequent of this conditional is rationally prohibited, it follows that believing the Moorean conjunction is also rationally prohibited.

While this line of reasoning, if correct, may give us some reason to believe that the Reflective Self-Ascription Requirement is true, it does not favour this principle over a weaker principle that requires that we merely be disposed to self-ascribe our *beliefs*, rather than all of our judgement-sensitive mental states. Is there anything more that can be said in favour of the stronger Reflective Self-Ascription Requirement?

There is another line of reasoning that supports this stronger requirement that is based on the connection between charity and rationality discussed in the previous chapter. It begins with the observation that charitable interpretation seems to involve a *defeasible* presumption against interpreting people as being out of touch with their own judgement-sensitive attitudes—that is, a defeasible presumption against ascribing to people judgement-sensitive mental states that they are not disposed, on reflection, to self-ascribe, or dispositions to self-ascribe judgement-sensitive mental states that they are not in. That is, charity appears to prohibit ascribing such states *unless* doing so would result in a more coherent overall picture of a person's mental states. Given the connection between charity and rationality discussed in the previous chapter, it seems plausible that charity defeasibly prohibits ascribing a mental state to someone just in case rationality requires that they not be in that state. It thus follows that rationality requires that someone be in a judgement-sensitive mental state if, and only if, they are disposed, on reflection, to believe that they themselves are in that mental state then.

One objection that might be raised against the Reflective Self-Ascription Requirement is that it is overly intellectualist. It seems possible to be in a mental state without even possessing the concept of that mental state: some animals, for example, have beliefs and desires without themselves possessing the concepts BELIEF or DESIRE. There are three main responses that an advocate of the Reflective Self-Ascription Requirement could make to this objection. The first is that the Self-Ascription Requirement does not say that it is *impossible* to be in a mental state and

not be disposed to self-ascribe that mental state on reflection—rather that it is merely *irrational*. So the animals that are not disposed to self-ascribe their mental states can still qualify as having mental states—they are simply less than fully rational.

The second response to make is that the Self-Ascription Requirement only requires that the relevant agent be disposed to self-ascribe the mental state should they consider the question of whether they are in that mental state. So it says that, if the relevant agent *were* to consider that question (in which case they would possess the relevant concepts), then they *would* answer it in the affirmative. This is compatible with its being the case that the relevant creature *in fact* lacks the concepts required in order to engage in this kind of reflection. The third possible response to this objection would be simply to modify the principle so that it only applies to agents that possess the relevant concepts (Shoemaker, 1995).

Although I have provided some reasons to believe that the Reflective Self-Ascription Requirement is true and defended it against some objections, I concede that it is a demanding and controversial principle (see Barnett, 2021 for a recent critical perspective). In the next chapter, I will show that, if this principle is true, then this helps to explain why normative judgements are subject to enkratic requirements. These enkratic requirements establish rational connections between judgements involving the concepts MUST, OUGHT, and MAY and intentions, and between judgements involving certain types of fitting-response concepts and their counterpart responses. The idea that there are enkratic requirements that establish a link between normative judgements and intentions is widely recognised. The idea that there are enkratic requirements that establish a link between fitting-response concepts and those responses themselves is less widely recognised, and, as we will see, these principles, if true, only apply to a small subset of fitting-response concepts. It is worth noting that, in order to explain why judgements involving MUST, OUGHT, and MAY are subject to enkratic principles, a weaker version of the Reflective Self-Ascription Requirement, which only applies to the self-ascription of one's own *intentions*, would suffice.

(Intention Self-Ascription Requirement) Rationality requires that someone intend to do something if, and only if, they are disposed, on reflection, to believe that they themselves then intend to do it.

Wallace (2001, pp. 21-22) and Setiya (2007b, pp. 669-671) each defend something very close to this principle (though cf. critical responses from Bratman, 2009, §4; and Brunero, 2008, §2). In so far as one has doubts about the purported enkratic principles that link fitting-response concepts to their counterpart responses, then, one can rely on these weaker—though still controversial—principles instead to make sense of the enkratic requirements that link judgements involving the concepts MUST, OUGHT, and MAY to intentions.

3.2.4.2 Self-Ascription Permissions

Next are two rational principles that permit that one's expectations about one's own mental states constrain which judgement-sensitive mental states one ends up in.

(Positive Self-Ascription Permission) Rationality permits that someone's being in a judgement-sensitive mental state be *explained by* their believing that they themselves will be in that mental state.

(Negative Self-Ascription Permission) Rationality permits that someone's not being in a judgement-sensitive mental state be *explained by* their not believing that they will be in that mental state.

According to these principles, a rational agent's self-ascriptions constitute a kind of 'practical knowledge'—knowledge that is 'the cause of what it understands' (Anscombe, 1957, p. 87). As above, I remain neutral here on the question of which particular type of explanatory relation these principles invoke—for example, whether it is a causal or a constitutive relation. The relation should *not* be understood as *basing*, however. If it were the case that rationality permits, say, one's intention to do something to be based on the belief that one will intend to do it, then rationality would permit the fact that one will intend to do it to be one's *reason* for intending to do it. But it seems irrational for the fact that one will intend to do something to be one's reason for intending to do it; rather, one's reasons for intending to do something should be supposed facts about the attractive substantive features of the intended action.

If cognitivism about endorsement were true, then the first of these principles would be equivalent to the Positive Endorsement Permission, while the second would be

equivalent to the Negative Endorsement Permission. Although these principles may appear surprising, there are two main reasons to think that they are true. The first is that they help to explain why the Reflective Self-Ascription Requirement is true. According to the Reflective Self-Ascription Requirement, fully rational agents are always disposed to self-ascribe their mental states correctly. This raises the question of how rational agents are in a position to self-ascribe their own mental states so reliably. The Self-Ascription Permissions provide an answer: at least part of the reason that rational agents' self-ascriptions are so reliable is that rational agents are able to adapt their mental states to fit their self-conception.

The second reason to endorse these principles is that they capture the intuitively appealing thought that rational agents have a capacity to exercise *authorship* over their own mental states. It is part of our common-sense conception of rational agency that agents have a kind of agency with respect to their own mental states that consists in an ability to make it the case that they are in a particular mental state simply by 'saying so'. These principles capture this intuitive picture nicely: rational agents have a capacity to make it the case that they will come to be in a mental state simply by believing that they will. Moran (2001) identifies two different perspectives that one might take in considering the question of whether one is, or will be, in some mental state. When one considers the question from the theoretical perspective, one sees the fact of the matter as being independent of whatever answer one comes to affirm. When one considers the question from the deliberative perspective, on the other hand, one sees the fact of the matter as dependent on the answer that one comes to affirm. For example, suppose I consider the question of whether I would prefer to have pasta or stir-fry for dinner. When I consider this question from the deliberative perspective, it seems to me that, if I were to conclude that I would prefer to have pasta, then because of that, I would come to prefer pasta, and similarly, if I were to conclude that I would prefer to have stir-fry, then because of that, I would prefer to have stir-fry (Baker, 2017; cf. Velleman, 2005, pp. 227-228; 2009, pp. 129-130). These Self-Ascription Permissions thus spell out an intuitively attractive picture of the kind of authorship or agency that rational agents have over some of their own mental states.

One might object that this picture attributes to rational agents a kind of freedom with respect to their own mental states that is too radical. According to this picture, a rational agent can make it the case that they are in some mental state—no matter how preposterous or unlikely—simply by 'saying so'. For example, it implies that, in so far as I am rational, I could make it the case that I believe that I am the Queen of England, simply by believing that I will adopt such a belief, which is absurd.

While, strictly speaking, it is true that this is a consequence of these Self-Ascription Permissions, this objection ignores the fact that these self-fulfilling self-ascriptions are themselves subject to rational norms. While I may, if I am rational, have the capacity to make myself believe that I am the Queen of England simply by believing that I will adopt this belief, in so far as I am rational, I am prevented from self ascribing this belief by the fact that I do not see any good reason to hold it. Just as I am rationally required not to ascribe mental states to others that I think would not make any sense for them to be in, I am rationally required not to ascribe such mental states to myself. So, while this picture does imply that rational agents have a radical kind of freedom with respect to their own mental states, it does not imply that rational agents may exercise that freedom arbitrarily.

3.2.5 Potential Objections to Cognitivism about Endorsement

In the last section, I provided an argument for cognitivism about endorsement by showing that self-ascription is subject to rational principles that, if cognitivism were true, would be equivalent to the rational principles to which endorsement is subject. In the next chapter, I provide more reason to believe that cognitivism about endorsement is true, by showing that, in combination with the analyses of normative concepts and account of rationality presented in earlier chapters, it explains the various Enkratic principles to which normative judgements are subject. Before that, however, I will consider some potential objections to cognitivism about endorsement.

3.2.5.1 Purported Counterexamples

It might seem that cognitivism about endorsement is vulnerable to counterexamples: one might think that it is possible to think of hypothetical scenarios in which someone endorses a mental state while failing to believe that they are in that mental state, or believes that they are in a mental state while failing to endorse their being in that

mental state. For example, I might endorse my preferring to work on my thesis over watching the cricket, while not believing that I have that preference; or I might believe that I prefer to watch the cricket over working on my thesis, without endorsing that preference. If these are in fact possibilities, then cognitivism about endorsement is false.

The first line of defence against counterexamples of this kind is to draw attention to the distinction between the endorsement of a mental state, on the one hand, and the positive assessment of that mental state, on the basis of which one might endorse it, on the other. Even if it is not possible for me to endorse my preferring to work on my thesis over watching the cricket without thereby believing that I have that preference, it is certainly possible for me to believe that I am required to prefer working on my thesis over watching the cricket, without believing that I have that preference. And even if it is not possible for me to believe that I prefer to watch the cricket over working on my thesis without thereby endorsing that preference, it is certainly possible for me to believe that I prefer to watch the cricket without believing that this is an appropriate preference for me to have.

The second line of defence against counterexamples of this kind is to draw attention to the distinction between the endorsement of a mental state and a desire or preference that one be in that mental state. Thus, even if it is not possible for me to endorse my preferring to work on my thesis over watching the cricket without thereby believing that I have that preference, it is certainly possible for me to desire or prefer that I prefer working on my thesis over watching the cricket, without believing that I have that preference. And even if it is not possible for me to believe that I prefer to watch the cricket over working on my thesis without thereby endorsing that preference, it is certainly possible for me to believe that I prefer to watch the cricket without desiring or preferring that I have this preference. Once endorsement is distinguished from normative judgement, desire, and preference, these counterexamples thus lose much of their appeal.

3.2.5.2 Purported Differences between Endorsement and Belief

Next, I will consider some purported differences between endorsement and belief that one might think show that they are distinct states. The first purported difference

is that reasons to believe that one is, or will be, in a mental state are often not reasons to endorse being in a mental state. For example, the fact that my psychologist tells me that I resent someone is a reason for me to believe that I resent them, but not necessarily a reason for me to endorse resenting them.

The response to this objection draws on the discussion of principles of charity from the previous chapter. It was argued in that chapter that there is a very close connection between beliefs about what mental states someone is in and beliefs about what mental states they are rationally required to be in, given their environment and behaviour. Given this close connection, it seems that a fact can only be a reason to believe that one is in some mental state if it is also a reason to believe that one is rationally required to be in that mental state, given one's environment and behaviour. Thus, the fact that my psychologist tells me that I resent someone is only a reason for me to believe that I resent them if it is also a reason for me to believe that, given my environment and behaviour, rationality requires that I resent them. Moreover, it is plausible that, in so far as something is a reason to believe that one is rationally required to be in a mental state, given one's environment and behaviour, it is also a reason to endorse being in that mental state. Thus, any reason to believe that one is in a mental state is also a reason to endorse being in that mental state.

Another purported difference between endorsement and belief is that, while rationality sometimes permits self-ascribing a mental state that one believes to be irrational, rationality never permits endorsing being in a mental state that one believes to be irrational. The response to this objection is that it is not clear that rationality does permit self-ascribing a mental state that one believes to be irrational. Worsnip (2018), for example, gives an account of rationality according to which it is *constitutive* of rationality that, as soon as you detect in yourself a mental state that you believe to be irrational, you consciously give up that mental state. We might also make an argument against the rationality of self-ascribing a mental state that one believes to be irrational along the following lines. If one believes that one is in an irrational mental state, then either (1) one's belief is true, in which case one is indeed in an irrational state; or (2) one is correct in believing that one is in the relevant mental state, but incorrect in believing that that mental state is irrational, in which case one is irrational because one believes an *a priori* falsehood; or (3) one is

incorrect in believing that one is in the relevant mental state, in which case one is irrational because one violates the Reflective Self-Ascription Requirement. This is objection can thus be resisted.

3.2.6 Related Views in the Literature

I will conclude my discussion of cognitivism about endorsement by comparing it to some related views in the literature. Cognitivism about endorsement bears a close resemblance to a view in the philosophy of mind called 'constitutivism' (Shoemaker, 1994; Bilgrami, 2012; Coliva, 2012; Taylor, 1976; 1985). Constitutivism is a thesis about the metaphysics of certain types of mental states, according to which part of what it is to be in a mental state is to believe that one is in that mental state then. Since, according to this view, any mental state of the relevant kind is at least partly constituted by the belief that one is then in that mental state, it is plausible to suppose that a rational agent's coming or continuing to be in such a mental state would be constrained by their expectations about whether they will come or continue to be in that mental state.

Constitutivism thus appears to entail cognitivism about endorsement. Cognitivism about endorsement does not entail constitutivism, however. Constitutivism is a thesis that, if true, would explain *how* a rational agent's self-ascriptions regulate their other mental states. Cognitivism about endorsement, by contrast, is merely the view that a rational agent's self-ascriptions *do* in fact play this regulating role. Constitutivism is a controversial position, and I do not take a stand on the question of how self-ascriptions play their regulating role in rational agents in this chapter.

Cognitivism about endorsement also bears some similarity to cognitivism about practical reason (Velleman, 1989; 2000; 2009; Setiya, 2007a; Harman, 1976; Marušić and Schwenkler, 2018). Cognitivism about practical reason is a view about the metaphysics of intentions according to which intentions are at least partly constituted by self-fulfilling beliefs about what one will do. According to cognitivism about practical reason, then, agents regulate their intentional actions by self-ascribing them. In so far as it is committed to the idea that agent's self-ascriptions play a regulating role, cognitivism about practical reason bears a close similarity to cognitivism about endorsement. The views differ in that, according to cognitivism

about practical reason, it is only an agent's intentional actions that are sensitive to their self-ascriptions, whereas according to cognitivism about endorsement, it is an agent's judgement-sensitive mental states that are sensitive to their self-ascriptions. Although they bear some similarity to each other, it is worth pointing out that cognitivism about endorsement is not committed to the controversial metaphysical thesis about the relation between intention and belief that lies at the core of cognitivism about practical reason.

3.3 Conclusion

In the first half of this chapter, I identified a several different types of enkratic principles. First, there are enkratic requirements, permissions, and prohibitions. I argued that Broome's account of these enkratic permissions and prohibitions as basing permissions and prohibitions is not plausible, and proposed that they instead be understood as principles that govern mental self regulation. I then introduced the distinction between positive enkratic principles, which have to with being in some mental state if one holds some relevant normative judgement, and negative enkratic principles, which have to do with not being in some mental state if one regards some normative question as open. In the second half of the chapter, I gave an account of the kind of self-regulation that was referred to in the first half of the chapter. According to this account, self-governing agents regulate their own mental states by endorsing them, and to endorse a mental state is simply to self-ascribe it. I provided an argument for this view by showing that self-ascriptions are subject to rational principles that are analogous to the rational principles that apply to endorsement. In the next chapter I will provide more reason to believe that cognitivism about endorsement is true by showing that this view, in combination with the other views defended in this thesis, explains why normative judgements are subject to enkratic principles.

4 Explaining Enkratic Principles

This chapter is devoted to formulating and explaining the various enkratic principles that apply to normative judgements. These principles include the kinds of positive and negative enkratic requirements, permissions, and prohibitions discussed in the previous chapter. I begin by discussing the enkratic principles that apply to judgements involving the concepts MUST, MAY, and OUGHT. The principle known in the literature as 'Enkrasia' is a principle that requires that one intend to do what one believes one ought to do. I argue that this principle is too demanding as a thesis about the rational relation between judgements about what one ought to do and intentions, and is instead better construed as a principle that concerns judgements about what one *must* do. I then present and discuss a number of enkratic principles that apply to judgements involving fitting-response concepts, such as DESIRABLE and PREFERABLE. I argue that only some judgements involving fitting-response concepts are subject to enkratic principles, namely judgements about what is fitting *from one's own perspective*.

My task in the second half of the chapter is to show that the analyses of normative concepts presented in chapters one and two, the account of rationality presented in chapter two, and the cognitivist account of reflective endorsement presented in chapter three, together explain why normative judgements are subject to these enkratic principles. I begin by clarifying just which types of judgements I will be showing to be subject to these principles. The judgements I have in mind can all ultimately be understood as judgements about what rationality permits or requires of someone whose behaviour and environment are identical to one's own. Then, for each enkratic principle, I employ a three-step strategy to explain why it holds. The 'analysis step' involves showing that there is a connection between the relevant normative judgement and some judgement about what rationality requires or permits. The 'charity step' involves showing that there is a connection between that judgement about what rationality requires or permits and the self-ascription of some relevant mental state. Finally, the 'reflective-endorsement step' involves showing that there is a connection between self-ascribing that mental state and being in that mental state.

4.1 Enkratic Principles

4.1.1 Must

Enkrasia is generally presented as a principle that establishes a rational connection between judgements about what one *ought* to do and intentions. In my view, a principle of this kind is more promising as an account of the rational connection between judgements about what one *must* do and what one intends to do. Replacing 'ought' with 'must' in our formulation of Enkrasia from the previous chapter yields the following.

For all *S*, *t*, and φ, rationality requires that:

- If:
    - *S* judges at *t* that they themselves must φ; and
    - *S* believes at *t* that it is up to them themselves then whether they φ;
- Then:
    - *S* intends at *t* to φ.

Let us call the clause in this principle about the agent's believing that it is 'up to' them themselves whether they perform the relevant action, the 'control condition'. Broome (2013) formulates the control condition as follows.

- *S* believes at *t* that, if they themselves were then to intend to φ, then because of that they would φ; and
- *S* believes at *t* that, if they themselves were not then to intend to φ, then because of that they would not φ.

Broome argues that this condition is needed because otherwise the principle would be vulnerable to counterexamples. For example, if the agent were instead to believe that it is *not* then up to them whether the relevant outcome obtains since the outcome would obtain even if they did not then intend it to, then their failing to intend to do it need not be irrational. Broome provides the following example to illustrate the point.

> Suppose you have moved to Fiji, and consequently believe you [must] learn about Fijian culture. But suppose you believe this will happen anyway, since you are living in Fiji. Then you may be rational even if you do not intend to learn about Fijian culture.[1] (p. 171)

Similarly, Broome argues that, if the agent were to believe that it is not then up to them then whether the relevant outcome obtains since that outcome would not obtain even if they intended it to, then their failing to intend to do it would not be irrational. He provides the following example to illustrate the point.

> Suppose you believe you [must] believe in God, but you do not believe that intending to believe in God would bring you do so. You do not believe you have that sort of control over this belief. Then you might be rational even if you do not intend to believe in God. (p. 171)

It is worth pointing out that Broome's formulation of the control condition is stronger than is required in order to deal with the examples he discusses (Southwood, 2016b). In each of these examples, the agent believes that it is *not* up to them then whether they perform the relevant action. These examples are thus compatible with the following variation on the original enkratic principle.

For all *S*, *t*, and φ, rationality requires that:

- If:
    - *S* judges at *t* that they themselves must φ;
    - *S* does *not* believe at *t* that it is *not* the case that, if they themselves were then to intend to φ, then because of that they would φ; and
    - *S* does *not* believe at *t* that it is *not* the case that, if they themselves were not then to intend to φ, then because of that they would not φ;
- Then:
    - *S* intends at *t* to φ.

This principle seems too demanding, however. Suppose that, in the Fiji example, you do not believe outright that you would learn about Fijian culture even if you didn't intend to, but you are nonetheless fairly confident that you would. It is not obvious to me that failing to intend to learn about Fijian culture is necessarily irrational in this case. Similarly, suppose that, in the God example, you do not believe outright that you would not believe in God even if you intended to, but you are nonetheless fairly confident that this is so. It is at least questionable whether failing to intend to believe in God would necessarily be irrational in this case. The following principle therefore seems to me to be a more plausible strengthening of the original principle.

For all *S*, *t*, and φ, rationality requires that:

- If:
    - *S* judges at *t* that they themselves must φ; and
    - *S* is *sufficiently confident* at *t* that, if they themselves were then to intend to φ, then because of that they would φ; and
    - *S* is *sufficiently confident* at *t* that, if they themselves were not then to intend to φ, then because of that they would not φ;
- Then:
    - *S* intends at *t* to φ.

While some principle along these lines may be true, in what follows I will focus on the original less demanding version of the principle, since it is less controversial.

The concept Broome is spelling out in his analysis of the control condition is the concept of *causal dependence* (Lewis, 1973). For someone to judge that it is 'up to' them themselves at some time whether they do something *just is* for them to judge that whether they themselves will do it causally depends on whether they themselves intend at that time to do it. It thus seems plausible that the control condition might be equally well analysed in terms of Lewisian 'non-backtracking' counterfactuals (Lewis, 1979). But since there is no need to take a position here on how best to analyse the concept of causal dependence, I will formulate the principle as follows.

(Positive Enkratic Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
    - *S* judges at *t* that they themselves must φ; and
    - *S* believes at *t* that whether they themselves will φ causally depends on whether they themselves then intend to φ;
- Then:
    - *S* intends at *t* to φ.

This principle is plausible; there certainly seems to be something irrational about failing to intend to do something when you believe that you must do it and that whether you will do it causally depends on whether you then intend to do it. It is also plausible that there is an enkratic permission that corresponds to this enkratic requirement. That is, it seems rational to *form* or *maintain* an intention to do

something, based on the judgement that one must do it and the belief that it is up to one oneself whether one does it.

(Positive Enkratic Permission) For all *S* and φ, rationality permits that *S*'s forming or maintaining the intention to φ be based on:

- Their believing that they themselves must φ; and
- Their believing that whether they themselves will φ causally depends on whether they themselves intend to φ.

And finally, it is plausible that corresponding to this enkratic permission there is an enkratic prohibition. While it can be rational to form an intention to do something based on the belief that one must do it, it is irrational to form, maintain, withhold or give up the belief that one must do something simply because one does or does not intend to do it.

(Positive Enkratic Prohibition) For all *S* and φ, rationality prohibits that *S*'s forming, maintaining, withholding or giving up the judgement that they themselves must φ be based on their intending, or not intending, to φ.

4.1.2 May

We have just looked at three plausible *positive* enkratic principles that establish a rational connection between judging that one must do something and intending to do it. In this section, I outline three *negative* enkratic principles that establish a rational connection between *not* being disposed to judge that one may do something and *not* intending to do it. The first principle captures the intuitive thought that it does not make sense to intend to do something if one is in doubt about whether one may do it. For example, if I am not sure whether I may mention my sister-in-law's pregnancy to a mutual friend today—because I am uncertain whether she has decided share the news with anyone outside of her close family yet—then it would not make sense for me to intend to mention it.

(Negative Enkratic Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
    - At *t*, *S* is not disposed, on reflection, to judge that they themselves may φ;

- Then:
    - At *t*, *S* does not intend to φ.

Note that this principle does not contain the condition that *S* believe that it is up to them themselves whether they φ. That condition does not seem to be required in this case. If *S* were not sufficiently confident at *t* that it is up to them (at some point) whether they φ, then it would not be rational for them to intend to φ, regardless of what they thought about its permissibility. This is because it only makes sense for someone to intend to do something if they are sufficiently confident that it is up to them whether they do it. Let us call this the 'Self-Efficacy Requirement'.

(Self-Efficacy Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
    - *S* is not sufficiently confident at *t* that whether they themselves will φ causally depends on whether they themselves intend (at some point) to φ;
- Then:
    - *S* does not intend at *t* to φ.

It is also the case, however, that, even if *S* were sufficiently confident that it is up to them whether they φ, it would not be rational for them to intend to φ if they were also in doubt about whether they may φ. That is, if *S* were sufficiently confident that it is up to them whether they φ, but they were in a state of doubt about whether they may φ, then it would not make sense for them to intend to φ. Thus, for all *S*, *t*, and φ, rationality requires that:

- If:
    - At *t*, *S* is not disposed, on reflection, to judge that they themselves may φ; and
    - At *t*, *S* is sufficiently confident that whether they themselves will φ causally depends on whether they themselves intend to φ;
- Then:
    - *S* does not intend at *t* to φ.

Thus, the Negative Enkratic Requirement for MAY holds regardless of how confident someone is about whether the relevant action is up to them.

One might be sceptical about the Negative Enkratic Requirement in view of the fact that one can sometimes find oneself in a situation in which must make a decision about what to do before one is in a position to resolve the question of what one may do. In this case, one might think that it can be rational to form an intention to do something despite being in doubt about whether one may do it. If this is true, then the Negative Enkratic Requirement for MAY is false. To assess whether this is true, the sense in which one 'must' make a decision needs to be clarified.

Suppose that one 'must' make a decision in the sense that it is somehow impossible for one not to make a decision. One might think, for example, that there is a sense in which any deliberating agent ultimately cannot help but make a decision. Korsgaard (2009) claims that human beings are 'condemned to choice and action'—that 'choosing and acting … is our plight' (pp. 1-2).

> Human beings are condemned to choice and action. Maybe you think you can avoid it, by resolutely standing still, refusing to act, refusing to move. But it's no use, for that will be something you have chosen to do, and then you will have acted after all. Choosing not to act makes not acting a kind of action, makes it something that you do. … You have no choice but to choose, and to act on your choice. (Korsgaard, 2009, p. 1)

Taking inspiration from these suggestive remarks, one might reason as follows.

1. If a deliberating agent were to conclude their deliberation without making a decision, then they would be acquiescing in what they believe to be the consequences of failing to make that decision.
2. But acquiescing in these believed consequences in this way amounts to deciding that those very consequences shall obtain.
3. Therefore, it is impossible for a deliberating agent to conclude their deliberation without making a decision.

Perhaps there is a weak sense of 'decide' according to which this argument is sound, but as long as deciding that something shall be so involves intending that it be so, then the second premise of the argument is false. There is a difference between accepting that something will occur and intending that it occur. Intending that something occur implies a commitment to ensuring that it occurs that merely accepting that it will occur does not (cf. Bratman, 1987, ch. 10). The appeal of this line of thinking may be due to confusing the idea that someone who brings their

deliberation to a close without making a decision would be *responsible* for the foreseeable consequences of their failure to make a decision with the idea that they *intended* that those consequences should come about.

The idea that it is impossible for any deliberating agent not to make a decision is thus implausible. Nevertheless, suppose that failing to make a decision is impossible for some other reason. For example, perhaps an devious neuroscientist has planted a chip in someone's brain that prevents them from not making a decision on some matter (cf. Frankfurt, 1969). One might think that is can be rational in this case for the person to decide to do something in spite of being in doubt about whether they may do, since this is *the best they can do*, rationally speaking. This is false, however, since the fact that being in a certain state is the best a person can do, rationally speaking, does not entail that being in that state is rational. In this case, it seems to me that the neuroscientist's interference has prevented the person from being fully rational.

One might object that there is something wrong with saying that a perfectly conscientious person violates a rational requirement, given that they have done all they can do to avoid irrationality. The first line of response to this objection is to highlight the distinction between *conditional* and *unconditional* requirements of rationality. The person in this situation may be perfectly conscientious in the sense that they have done everything that rationality requires of them, *conditional on their being in the unusual circumstances in which they find themselves*. Their being conscientious in this sense is compatible with their violating some *unconditional* requirement of rationality. The Negative Enkratic Principle for MAY is an *unconditional* requirement of rationality. It says, roughly, that no maximally rational agent would intend to do something while being in a state of doubt about whether they may do it. The corresponding conditional requirement says, roughly, that no agent who is maximally rational, given that they have been manipulated in this way by a devious neuroscientist, would intend to do something while being in a state of doubt about whether they may do it. It is plausible that this principle is false. The second line of response to this objection is to point to the distinction between violating an unconditional rational requirement and being *blameworthy* or *criticizable* for violating that requirement. In this case, it is plausible that, although the person violates an

unconditional requirement of rationality, the fact that they have been manipulated by the devious neuroscientist excuses them from being blamed or criticised for doing so.

The idea that the fact that failing to make a decision is impossible can make it rational to intend to do something despite being in doubt about whether one may do it thus seems implausible. Perhaps the idea behind this objection is instead that this can be rational in cases in which one 'must' make a decision in the sense that one holds some other attitudes that rationally commit one to making a decision. For example, suppose that one intends some end, and believes that one would fail to achieve that end unless one were immediately to decide on the means to that end. Broome (2013, p. 170) argues that being in a situation like this rationally commits one to deciding immediately on some means to one's end. He thus proposes the following rational requirement.

(Generalised Instrumental Requirement) For all *S*, *t*, φ, and ψ_1, … , and ψ_n, rationality requires that:

- If:
    - *S* intends at *t* to φ;
    - *S* believes at *t* that, if it were not the case that they themselves ψ_1, … , or ψ_n, then because of that they themselves would not φ; and
    - *S* believes at *t* that, if they themselves were not then to intend to ψ_1, then because of that they themselves would not ψ_1, … , and if they themselves were not then to intend to ψ_n, then because of that they themselves would not ψ_n;
- Then *S* intends at *t* to ψ_1, or … , or *S* intends at *t* to ψ_n.

This principle seems plausible; any fully rational agent who, at some time, intends an end and believes that their achieving that end requires that they then intend one of the means to that end should, at that time, intend what they take to be one of the means to that end. It follows that, if it is not irrational to be in this situation while also failing to be disposed, on reflection, to judge, of any of those means, that one may take them, then the Negative Enkratic Requirement for MAY is false.

It seems to me that it is in fact irrational to be in this kind of state. It seems irrational to continue to intend an end when one believes that achieving that end requires

immediately deciding on the means to it but one is in a state of doubt, for each of those means, about whether one may take it. If one's beliefs in this situation are rational, and one's doubts about the permissibility of the means are rational, then it seems to me that the rational thing to do is to give up the intention to achieve the end. Continuing to intend the end in this situation seems to exhibit precisely the kind of irrational impetuosity of will that the Negative Enkratic Requirement for MAY prohibits.

One might object, however, that there are clear examples in which being in a state of this kind is rational. For example, suppose that Sam is faced with a choice between saving her husband's life and saving 100 strangers' lives. She believes that, if she were not to make a decision now, then because of that 1,000,000 people, including her husband and the 100 strangers, would die. She judges that she must not let the 1,000,000 die, so she intends not to let that happen. She is, however, in a state of doubt about whether she may save her husband—since she is not sure whether the loss of the lives of 100 strangers is an acceptable cost of doing so. And she is in a state of doubt about whether she may save the 100 strangers—since she believes that she has a special duty to protect her husband. Nonetheless, she intends to save her husband. There needn't be anything irrational about this combination of attitudes.

It still seems to me that, if Sam truly cannot make up her mind about whether she may save her husband or save the 100 strangers, then, in so far as she is rational, she will—regrettably—fail to make up her mind about whether to save him, and so she will—regrettably—be forced to abandon her intention not to let the 1,000,000 others die. One might argue that the fact that Sam judges that she *must* not let the 1,000,000 die makes it rational to hold onto her intention to do so, and thus compel her to decide on a means. This is false, however, since, if Sam is in doubt, for each of the believed means to her end, about whether she may take that means, then she is *a fortiori* not persuaded that the importance of achieving that end justifies taking that means. In my opinion, we are only tempted to think that Sam's judgement that she must not let the 1,000,000 die makes a difference in this case to the extent that we find it obvious that the importance of that end is sufficient to justify either of the means.

I think that the reason that Sam's intending to save her husband in spite of being in doubt about whether she may do so appears rational in this case is that it is natural to think that it would be rational for Sam to *bring it about* that she has this intention, if she can. After all, it is natural to describe this case as one in which, in spite of being in doubt about what to do, Sam *forces* herself to make a decision. This suggests that Sam has a way of *intentionally bringing it about* that she intends to save her husband, based on her intention not to let him and the 100,000 die, and her belief that intending to save her husband would achieve that end. Her intentionally bringing it about that she has this intention seems rational, but that does mean that her holding this intention, while being in doubt about what she may do, is rational. In this case, Sam rationally brings it about that she is in an irrational state. It is thus a case of what Parfit (1984, ch. 1, §5) calls 'rational irrationality'.

Another possible objection to the Negative Enkratic Requirement for MAY is that it is overly intellectualist. It seems possible for someone to intend to do something despite not even possessing the concept MAY. Perhaps some animals, for example, can intend to do things despite lacking the capacity for normative judgement. The first line of response to this objection is to point out that the Negative Enkratic Requirement for MAY does not imply that it is impossible for someone to intend to do something despite lacking the concept MAY. Rather, it merely says that this it is *irrational* to intend to do something without being disposed, on reflection, to judge that one may do so. The second line of response to this objection is to point out that the Negative Enkratic Requirement for MAY only requires that someone who intends to do something be disposed to judge that they may do it, *should they consider the question of whether they may do it*. So it says that, were they to consider that question—in which case they would possess the relevant concept—then, they would answer it in the affirmative. Nevertheless, one might still maintain that if someone lacks the concept MAY, then it would not be irrational for them to intend to do something without being disposed to judge, on reflection, that they may do so. The simplest response to this objection would be to add a clause to the antecedent of the conditional that the Negative Enkratic Requirement for MAY says is required that specifies that the relevant agent possesses the concept MAY.

I conclude, then, that these objections to the Negative Enkratic Requirement do not succeed. It is plausible that there is also a corresponding Negative Enkratic Permission that permits that one give up or withhold an intention to do something on the basis of not being disposed to judge that one may do it.

(Negative Enkratic Permission) For all *S* and φ, rationality permits that *S*'s withholding or giving up the intention to φ be based on:

- Their not being disposed, on reflection, to judge that they may φ; and
- Their belief that whether they themselves φ causally depends on whether they themselves intend to φ.

The belief that the action is up to oneself is required in this case. It does not make sense for someone's not intending to do something to be based on their being in doubt about whether they may do it unless they believe that it is something that is within their control. If they lacked the belief that they have control over the relevant action, then it would make sense for their not intending to perform that action to be based on that, rather than on their being in doubt about whether they may perform it. For example, it would not make sense for my not intending to mention my sister-in-law's pregnancy to our mutual friend today to be based on my doubt about whether I may do so if I were uncertain whether it was even up to me whether I mention it to them—for example, if I thought that they might have left this morning on a long-haul flight with no telephone or internet reception. It would make more sense in this case for my not intending to mention the news to be based on my uncertainty about whether it is even up to me whether I do so, rather than my doubt about whether I may do so. Thus, I am not convinced that a stronger principle than the Negative Enkratic Permission for MAY, that does not include the 'up to' clause, is plausible.

Finally, there is an enkratic prohibition that is related to the two principles just discussed. While it can be rational to withhold or give up an intention to do something based on not being disposed, on reflection, to judge that one may do it, it is irrational to form, maintain, give up, or withhold the judgement that one may do something just because one intends, or does not intend, to do it.

(Negative Enkratic Prohibition) For all *S* and φ, rationality prohibits that *S*'s forming, maintaining, giving up, or withholding the judgement that they themselves may φ be based on their intending, or not intending, to φ.

4.1.3 Ought

The principle of Enkrasia, as it is typically formulated in the literature, is too demanding. Enkrasia is typically thought to be a principle that requires that one intend to do what one believes one ought to do.

(Enkrasia) For all *S*, *t*, and φ, rationality requires that:

- If:
    - *S* judges at *t* that they themselves ought to φ; and
    - *S* believes at *t* that it is up to them themselves then whether they themselves φ;
- Then:
    - *S* intends at *t* to φ.

To see that Enkrasia is too demanding, consider the following scenario. Sam has been exercising for 20 minutes. She judges that, while she *ought* to exercise for a whole hour, she doesn't *have to*—anything over 20 minutes would be *ok*. Sam also believes that it is up to her then whether she exercises for a whole hour: if she were then to intend to push on for the full hour, then because of that she would; but if she were not then to intend to do so, then because of that she would quit. In fact, Sam does not intend to exercise for the full 60 minutes and her motivation to keep going quickly dissipates.

In this case, Sam judges that she *ought* to exercise for a whole hour and believes that it is up to her then whether she does, but she also judges that she *may* exercise for less than a full hour. In other words, she views her exercising for a full hour as *recommended*, but *optional*. Given that in this case Sam sees exercising for the full hour as merely optional, her failure to commit to doing so does not seem to me to be irrational. We are not rationally required to commit to courses of action that we believe to be merely optional. I thus conclude that the principle of Enkrasia is false.

This conclusion may seem unsatisfying. Since Sam believes that she ought to exercise for the full hour rather than quitting after 20 minutes, she clearly regards the latter as a *worse option* than the former. Surely this normative assessment should be reflected in her motivational states somehow. Although I think that Enkrasia is false, I nonetheless think that Sam's normative assessment of these two options does place a rational constraint on her intentions. Although in this example rationality does permit that Sam *not intend* to continue for the whole hour, it seems to me that rationality does not permit that Sam *intend not* to continue for the full hour. To intend to undertake a course of action is to *commit* to that course of action in a way that involves *ruling out* from consideration any available options that one takes to be incompatible with that course of action (Bratman, 1987). It seems to me that, as long as someone believes that they ought to do something, it does not make sense for them to rule it out as an option in this way, by intending to do something else. Thus, in this example, I maintain that, although Sam is not required to commit to continuing for the full hour, she is not permitted to commit to stopping before the full hour is up either. The following principle thus strikes me as a satisfying weakening of Enkrasia.

(Negative Enkratic Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that they themselves ought *not* to φ; and
  - *S* believes at *t* that whether they themselves will φ causally depends on whether they themselves then intend to φ;
- Then:
  - *S* does *not* intend at *t* to φ.

It is plausible that this principle also has a corresponding enkratic permission. Not only does rationality require that one not intend to do something that one judges one ought not to do, rationality permits that one actively *give up* or *withhold* an intention to do something based on the judgement that one ought not to do it.

(Negative Enkratic Permission) For all *S* and φ, rationality permits that *S*'s giving up or withholding the intention to φ be based on:

- Their judgement that they themselves ought not to φ; and

- Their belief that whether they themselves will φ causally depends on whether they themselves intend to φ.

Finally, as with MUST and MAY, it is plausible that there is a counterpart enkratic prohibition for OUGHT. While it may be rational to give up or withhold an intention to do something based on the judgement that one ought not to do it, it is irrational to form, maintain, give up, or withhold the judgement that one ought not to do something just because one does or does not intend to do it.

(Enkratic Prohibition) For all *S* and φ, rationality prohibits that *S*'s forming, maintaining, withholding or giving up the judgement that they themselves ought not to φ be based on their intending, or not intending, to φ.

4.1.4 Fitting-Response Concepts

So far, I have been discussing enkratic principles that apply to judgements involving the concepts MUST, MAY, and OUGHT. In this section, I will show that some judgements involving fitting-response concepts are also subject to enkratic principles. More specifically, I will show that judgements about the degree to which something is fitting *from one's own perspective* are subject to enkratic principles. In chapter two, I pointed out that, while some fitting-response concepts are perspective-independent, such that whether they apply to an object depends only on the properties of that object and not on anyone's attitudes toward it, other fitting-response concepts are perspective-dependent. My focus here will thus be on a subset of these perspective-dependent concepts. DESIRABLE (FROM MY PERSPECTIVE) and PREFERABLE (FROM MY PERSPECTIVE) are good examples of the kinds of concept I have in mind. One might be somewhat sceptical about these concepts. They do seem rather artificial, as witnessed by the parenthetic text needed to make clear their perspective-dependence.[29] Nonetheless, they seem to me to be the best candidates when it comes to fitting-response concepts that are subject to enkratic principles. Below I will argue that other fitting-response concepts are not subject to enkratic principles. In so far as one is sceptical about these concepts, then, the main lesson to be drawn from

[29] I am grateful to an anonymous examiner for raising this objection.

this section is that ordinary fitting-response concepts are not subject to enkratic principles.

4.1.4.1 At Least and Equal

Let us begin with judgements to the effect that the degree to which it is fitting, from one's own current perspective, to hold some attitude toward some object is *at least as great as* the degree to which it is fitting, from one's own current perspective, to hold some other attitude toward some other object. An example is the judgement that the degree to which it is desirable, from my perspective, that Australia wins the Ashes is at least as great as the degree to which it is desirable, from my perspective, that England wins the Ashes. The first principle I will focus on captures the intuitive idea that, if I hold this judgement, then I am rationally committed to desiring that Australia wins the Ashes at least as much as I desire that England wins the Ashes.

(Positive Enkratic Requirement) For all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:
  - *S* judges at *t* that the degree to which *x* is then *A*-able (from their own perspective) is at least as great as the degree to which *y* is then *B*-able (from their own perspective);
- Then:
  - *S A*s *x* at least as much as they *B y* at *t*.

This is a plausible principle. It is also plausible that rationality permits that one regulate one's mental states in order to bring them into line with these kinds of judgements. For example, it is plausible that rationality permits that I form or maintain a desire that Australia win the Ashes that is at least as strong as my desire that England win the Ashes based on my belief that it will continue to be the case that the former is at least as desirable as the latter is, from my own perspective.

(Positive Enkratic Permission) For all *S*, *x*, *y*, *A*, and *B*, rationality permits that *S*'s forming or maintaining the mental state of *A*-ing *x* at least as much as they *B y* be based on their judging that it will continue to be the case that the degree to which *x* is *A*-able (from their own perspective) is at least as great as the degree to which *y* is *B*-able (from their own perspective).

Finally, there is an asymmetry here, in so far as it does not seem rational for me to form, maintain, give up, or withhold the judgement that the degree to which it is desirable that Australia wins the Ashes is at least as great as the degree to which it is desirable that England wins the Ashes just because I do or do not desire that Australia wins the Ashes at least as much as I desire that England win the Ashes.

(Positive Enkratic Prohibition) For all *S*, *x*, *y*, *A*, and *B*, rationality prohibits that *S*'s forming, maintaining, giving up, or withholding the judgement that the degree to which *x* is *A*-able (from their own perspective) is at least as great as the degree to which *y* is *B*-able (from their own perspective) be based on their *A*-ing *x* at least as much as they *B y*, or their not doing so.

It thus seems plausible that there are enkratic principles that govern judgements to the effect that the degree to which it is fitting, from one's own current perspective, to hold some attitude toward some object is at least as great as the degree to which it is fitting, from one's own current perspective, to hold some other attitude toward some other object. Given the close relationship between the concepts EQUAL and AT LEAST AS, it is plausible that there are analogous enkratic principles that govern judgements that the degree to which it is fitting, from one's own current perspective, to hold some attitude toward some object is *equal to* the degree to which it is fitting, from one's own current perspective, to hold some other attitude toward some other object. For the sake of conciseness, I will not outline these here.

4.1.4.2 Greater

It is tempting to think that there might also be enkratic principles, analogous to the principles discussed in the previous section, that govern judgements to the effect that the degree to which it is fitting, from one's own perspective, to hold some attitude toward some object is *greater than* the degree to which it is fitting, from one's own perspective, to hold some other attitude toward some other object. For example, one might think that there is a positive enkratic requirement that entails that rationality requires that, if I believe that it is more desirable, from my perspective, that the Ashes series end in a tie than it is that England win the Ashes, then I desire that the Ashes series end in a tie more than I desire that England win the Ashes. This is false, however, since I may believe that, although it is more desirable, from my perspective,

that the Ashes end in a tie than it is that England wins, these outcomes are nonetheless *on a par*. As long as I regard these outcomes as being on a par, it seems to me that rationally permits that I be indifferent between them, and thus desire them equally. I am thus not rationally required to desire the former more than I desire the latter.

It does seem plausible, however, that there are enkratic principles that apply to judgements to the effect that the degree to which it is fitting, from one's own perspective, to hold some attitude toward some object is *greater than and not on a par with*, the degree to which it is fitting, from one's own perspective, to hold some other attitude toward some other object. For example, it seems plausible that, if I judge that the degree to which it is desirable, from my perspective, that Australia wins the Ashes is greater than, and not on a par with, the degree to which it is desirable, from my perspective, that England wins the Ashes, then I am indeed rationally committed to desiring that Australia win the Ashes more than I desire that England win the Ashes.

(Positive Enkratic Requirement) For all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:
  - *S* judges at *t* that the degree to which *x* is then *A*-able (from their own perspective) is greater than and not on a par with the degree to which *y* is then *B*-able (from their own perspective);
- Then:
  - *S A*s *x* at *t* more than they *B y* at *t*.

Similarly, it is plausible that there is a principle that permits that I form or maintain a desire that Australia win the Ashes that is stronger than my desire that England win the Ashes based on my believing that it will continue to be the case that the degree to which the former is desirable, from my perspective, is greater than, and not on a par with, the degree to which the latter is desirable, from my perspective.

(Positive Enkratic Permission) For all *S*, *x*, *y*, *A*, and *B*, rationality permits that *S*'s forming or maintaining the mental state of *A*-ing *x* more than they *B y* be based on their judging that it will be the case that the degree to which *x* is *A*-able (from their

own perspective) is greater than and not on a par with the degree to which *y* is *B*-able (from their own perspective).

Finally, it does not seem rational for me to form ,maintain, give up, or withhold the judgement that the degree to which it is desirable, from my perspective, that Australia wins the Ashes is greater than, and not on a par with, the degree to which it is desirable, from my perspective, that England wins the Ashes simply because I do or do not desire that Australia wins the Ashes more than I desire that England win the Ashes.

(Positive Enkratic Prohibition) For all *S*, *x*, *y*, *A*, and *B*, rationality prohibits that *S*'s forming, maintaining, giving up, or withholding the judgement that the degree to which *x* is *A*-able (from their own perspective) is greater than and not on a par with the degree to which *y* is *B*-able (from their own perspective) be based on their *A*-ing *x* more than they *B y*, or their not doing so

4.1.4.3 On a Par

The principles we have been considering so far are *positive* enkratic principles, which have to do with the relation between holding some judgement about fitting attitudes and being in some other mental state. In this section I will focus on some *negative* enkratic principles involving the concept ON A PAR. It seems plausible that rationality requires that, if I am in doubt about whether the degree to which it is desirable, from my own current perspective, that the Ashes end in a tie is *on a par* with the degree to which it is desirable, from my own current perspective, that England wins the Ashes, then I do not desire these outcomes equally. In other words, it only makes sense for me to desire these outcomes equally if I am prepared to judge that they are at least on a par. This is because to judge that two things are on a par is to judge that they are 'roughly equal', and if one were in doubt about whether two things are even roughly equal in desirability, then it would not make sense to commit oneself to desiring them equally.

(Negative Enkratic Requirement) For all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:

- *S* is not disposed at *t*, on reflection, to judge that the degree to which *x* is then *A*-able (from their own perspective) is on a par with the degree to which *y* is then *B*-able (from their own perspective);

- Then:
  - At *t*, *S* does not *A x* equally as much as they *B y*.

There is plausibly also an enkratic permission according to which rationality permits that I give up or withhold the mental state of not desiring that the Ashes end in a tie equally as much as I desire that England win the Ashes, based on my regarding it as an open question whether these outcomes will be on a par at the time that I would otherwise end up holding those attitudes.

(Negative Enkratic Permission) For all *S*, *x*, *y*, *A*, and *B*, rationality permits that *S*'s giving up or withholding the mental state of *A*-ing *x* equally as much as they *B y* be based on their not being disposed, on reflection, to judge that it will be the case that the degree to which *x* is *A*-able (from their own perspective) is on a par with the degree to which *y* is *B*-able (from their own perspective).

Finally, there is plausibly a Negative Enkratic Prohibition that entails that rationality prohibits that I form, maintain, give up, or withhold an attitude that would settle the question of whether the degree to which it is desirable, from my perspective, that the Ashes end in a tie is on a par with the degree to which it is desirable from my perspective, that England win the Ashes just because I do or do not desire those outcomes equally.

(Negative Enkratic Prohibition) For all *S*, *x*, *y*, *A*, and *B*, rationality prohibits that *S*'s forming, maintaining, giving up, or withholding the judgement that the degree to which *x* is *A*-able then (from their own perspective) is on a par with the degree to which *y* is *B*-able then (from their own perspective) be based on their *A*-ing *x* equally as much as they *B y*, or their not doing so.

4.1.4.4 Perspective-Independent Fittingness

So far, I have been focusing on judgements featuring fitting-response concepts that involve a particular kind of perspective-dependent fittingness—judgements about what is fitting from one's own perspective. It seems plausible that judgements of this

kind are subject to enkratic principles. Many fitting-response concepts are not perspective-dependent in this way, however. For example, judgements about whether someone is *admirable* cannot plausibly be interpreted as judgements about the degree to which it is fitting, from one's own perspective, to admire them. Rather, as discussed in chapter two, they are plausibly interpreted as judgements about the degree to which it is fitting from some objective perspective—say a perspective that involves perfect knowledge about and acquaintance with that person—to admire them. Are judgements involving such perspective-independent fitting-attitude concepts also subject to enkratic principles?

There is good reason to believe that judgements involving such perspective-independent fitting-attitude concepts are not subject to enkratic principles. Let us use the concept ADMIRABLE as an example. Suppose that I have been told by someone whom I believe to be extremely reliable that the degree to which Sam is admirable is much greater than (and thus not on a par with) the degree to which Tom is admirable, and I therefore come to believe that this is true. Does it follow that I am thereby rationally committed to admiring Sam more than I admire Tom? It does not. Suppose that I know nothing else about Sam or Tom. In that case, it would be strange for me to admire either of them more than the other, since it only makes sense to admire someone on the basis of what I believe about their character, actions, achievements, and so on. It would thus only be rational for me to admire Sam more than Tom if I had some further beliefs about how they differ with regard to their character, actions, and so on (Whiting, 2015; Lord, 2016; cf. Wodak, 2021).

I therefore conclude that there is no enkratic requirement that entails that judging that Sam is more admirable than Tom is rationally commits one to admiring Sam more than Tom. The same kind of reasoning applies for putative enkratic requirements that apply to judgements about the comparative perspective-independent fittingness of attitudes. Neither do there appear to be any related enkratic permissions that apply to these kinds of judgements. For example, it would not be rational for my admiring Sam more than I admire Tom to be based on my belief that Sam is much more admirable than Tom is, because rationality requires that my doing so be based on my beliefs about their character, actions, and achievements instead. Thus, rationality

does not permit that my admiring Sam more than I admire Tom be based on my believing that Sam is much more admirable than Tom is.

Even if one grants this point, however, one might wonder whether there is an enkratic principle that entails that rationality requires that, if I believe that Sam is much more admirable than Tom is, *in virtue of* certain facts about them—in virtue, say, of the fact that Sam donates more money to charity than Tom does, for example—then I admire Sam more than I admire Tom. That is, one might wonder whether there are enkratic principles like the following.

(Positive Enkratic Requirement) For all *S*, *t*, *x*, *y*, *A*, *B*, and *p*, rationality requires that:

- If:
    - *S* believes at *t* that the degree to which *x* is then (perspective-independently) *A*-able is greater than, and not on a par with, the degree to which *y* is then (perspective-independently) *B*-able, *in virtue of the fact that p*;
- Then:
    - *S A*s *x* more than they *B y* at *t*.

This principle is false. Suppose you are at the pet store deciding whether to buy a cat or a dog. The salesperson tells you that they have one cat—Felix—and one dog—Fido—and that, while Felix and Fido are both loveable little creatures, Fido is more loveable than Felix is, because he is much more affectionate. Nevertheless, since you believe that cats are easier to take care of than dogs are, you take Felix home. Naturally, over time you grow to love the somewhat aloof Felix, and not the affectionate Fido. Thus, you come to love Felix more than you love Fido, even though you still believe that Fido is objectively more loveable in virtue of his more affectionate nature. This does not seem to be irrational. Even though you know what it is about Fido that makes him more loveable than Felix is, it does not make sense for you to love Fido at all, since you do not have any kind of relationship with him. Thus, I conclude that judgements involving perspective-independent fitting-attitude concepts are not subject to enkratic principles.

4.2 Explaining Enkratic Principles

Having established how these enkratic principles are best formulated, I turn now to the task of showing how the analyses of these concepts from chapters one and two, the account of rationality from chapter two, and the cognitivist account of endorsement from chapter three together make sense of the fact that these enkratic principles hold. I begin by clarifying just which kinds of normative judgements I will be showing to be subject to enkratic principles. I then review some principles that will play a role in the argument to follow. With that background in place, I then set about showing why these enkratic principles hold. I will explain the enkratic principles that apply to judgements involving the concepts MAY, MUST, and OUGHT, and, with regard to judgements involving fitting-response concepts, I will focus, by way of example, on the enkratic principles that apply to judgements involving the concepts AT LEAST and ON A PAR. I employ roughly the same three-step strategy for each enkratic principle. The 'analysis step' establishes a connection between the relevant normative judgement and some judgement about what rationality requires or permits. The 'charity step' establishes a connection between that judgement about what rationality requires or permits and the self-ascription of the relevant mental state. Finally, the 'reflective-endorsement step' establishes a connection between self-ascribing that mental state and being in that mental state.

4.2.1 Fittingness from One's Own Perspective

In chapter one, I argued that the concepts MUST, MAY, and OUGHT can be analysed in terms of fitting-response concepts such as PERMISSIBLE, CONCEIVABLE, or DESIRABLE. The types of judgement I will be concerned with here are judgements about a particular kind of *permissibility*, where the relevant notion of fittingness is the notion of fittingness *from one's own perspective* discussed above. Thus, all of the judgements I will be focusing on in this section can ultimately be understood as judgements about what is fitting from one's own perspective.

In chapter two, I argued that judgements involving fitting-response concepts can be understood as judgements about what rationality requires or permits. Here I will be understanding judgements about what is fitting from one's own perspective as judgements about what rationality requires or permits of someone given that their

behaviour and environment are identical to one's own. Thus, for me to judge that something is rationally required or permitted in this sense is for me to judge that it is rationally required or permitted of anyone who behaves exactly in the ways that I behave, and is related to their environment in exactly the ways in which I am related to my environment. In what follows, however, for the sake of readability, I will omit this detail, and simply refer to judgements about what rationality requires or permits.

4.2.1.1 Allowing

As mentioned above, I will be understanding judgements involving the concepts MUST, MAY, and OUGHT here as judgements about what is *permissible*, where to judge that something is permissible is to judge that *rationality permits* that it be *permitted*. To avoid confusion, I will use the term 'allow' rather than 'permit' to refer to this kind of agential response from now on. In chapter one, I identified a distinctively *practical* type of allowing. One allows that something be so, in this sense, if, say, one fails to prevent it from being so when it is in one's power to do so, or if one simply intentionally brings it about. Thus, if I gain weight because I fail to prevent myself from doing so when I could have, then I thereby allow myself to gain weight. Similarly, if I gain weight intentionally, then I thereby allow myself to gain weight. I will not provide a precise analysis of this concept here. For current purposes, all that is required is to identify two *a priori* truths about allowing.

(Failing to Prevent) For all *S* and *p*, if its being the case that *p* causally depends on *S*'s intending that *p*, and *S* does not intend that *p*, then *S* allows that ¬*p*.

(Bringing About) For all *S* and *p*, if its being the case that *p* causally depends on *S*'s intending that *p*, and *S* intends that *p*, then *S* allows that *p*.

In explaining why normative judgements are subject to enkratic principles below, I will draw on some rational principles regarding allowing and intending that hold in virtue of the fact that these truths are *a priori*. It will be useful to summarise these here. The first two principles are plausible in view of Failing to Prevent. The first is a rational requirement according to which, if one believes that whether one will do something depends on whether one then intends to do it, and one believes that one will not allow oneself not to do it, then rationality requires that one be disposed, on reflection, to believe that one then intends to do it. The second principle is a basing permission

according to which it is rational to base a belief that one intends to do something on the belief that whether one will do it causally depends on whether one intends to do it, and the belief that one will not allow oneself not to do it.

The second two principles are plausible in view of Bringing About. The first of these is a rational requirement according to which, if you believe that whether you will do something depends on whether you then intend to do it, then you are disposed to believe that you then intend to do it, only if you are also disposed to believe that you will allow yourself to do it. The second principle is a basing permission according to which it is rational for one's not believing that one will intend to do something to be based on the belief that whether one will do it depends on whether one intends to do it, and one's not being disposed, on reflection, to believe that one will allow oneself to do it.

4.2.1.2 Principles of Charity in Self-Ascription

In chapter two, I set out several principles of charity that connect judgements about what rationality permits or requires to mental-state ascriptions. Below, I will rely on variants of each of these principles. These variants differ from the original principles in two minor ways. First, they pertain only to the *self*-ascription of mental states, and, secondly, the relevant judgements about what rationality permits or requires are subtly different from those that feature in the original principles. The original principles concerned judgements about what rationality permits or requires *of a particular person*, given the facts about their behaviour and environment. The principles I draw on here concern judgements about what rationality permits of *anyone* whose behaviour and environment are identical to those of a particular person. Since the difference between these two judgements is so subtle, I take it that substituting one for the other in these principles does not affect their truth. My reason for making this substitution is that, given the analyses of fitting-response concepts in chapter two, judgements involving these concepts are judgements of this form. As mentioned above, however, for the sake of readability in what follows, I will omit this detail and simply refer to judgements about what rationality requires or permits.

The first of these principles is derived from the general requirement to ascribe a mental state to someone if one believes that, given their behaviour and environment, rationality requires of them that they be in that mental state.

(Positive Requirement of Charity in Self-Ascription) For all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* judges at *t* that rationality requires φ-ing;
- Then:
  - *S* believes at *t* that they themselves φ.

The next principle is a rational permission that is a counterpart of the first, and is derived from the general principle that permits that one's ascribing a mental state to someone be based on the belief that they are rationally required to be in that mental state, given their environment and behaviour.

(Positive Permission of Charity in Self-Ascription) For all *S* and φ, rationality permits that *S*'s belief that they themselves φ be based on their judgement that rationality requires φ-ing.

The next two principles have to do with not being disposed to ascribe a mental state to oneself if one is in a state of doubt about whether rationality permits that one be in that mental state. The first is a requirement not to ascribe a mental state to oneself in these circumstances.

(Negative Requirement of Charity in Self-Ascription I) For all *S*, *t*, and φ, rationality requires that:

- If:
  - At *t*, *S* is not disposed, on reflection, to judge that rationality permits φ-ing;
- Then:
  - At *t*, *S* is not disposed, on reflection, to believe that they themselves φ.

The next principle is a counterpart to the previous one, and permits that one not be disposed to ascribe a mental state to oneself based on being in a state of doubt about whether rationality permits being in that mental state.

(Negative Permission of Charity in Self-Ascription I) For all *S* and φ, rationality permits that *S*'s not being disposed, on reflection, to believe that they themselves φ be based on *S*'s not being disposed, on reflection, to believe that rationality permits φ-ing.

The final two principles have to do with not ascribing to oneself a mental state that one believes not to be rationally required of oneself. The idea here is that one should not commit to believing that one is in some mental state if one believes that there are perfectly good interpretations of oneself, given one's environment and behaviour, that do not ascribe that mental state.

(Negative Requirement of Charity in Self-Ascription II) For all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that it is not the case that rationality requires φ-ing;
- Then:
  - *S* is not disposed at *t*, on reflection, to believe that they themselves φ.

The final principle is the counterpart to the previous one and permits that one's not being disposed to believe that one is in a mental state be based on one's not believing that one is rationally required to be in that mental state, given one's behaviour and environment.

(Negative Permission of Charity in Self-Ascription II) For all *S* and φ, rationality permits that *S*'s not being disposed, on reflection, to believe that they themselves φ be based on their belief that it is not the case that rationality requires φ-ing.

4.2.1.3 Reflective Endorsement

In what follows, I rely the truth of cognitivism about endorsement, which I defended in chapter three.

(Cognitivism about Endorsement) For all *p*, for someone to *endorse* its being the case that *p just is* for them to believe that *p*.

Thus, I will assume that for someone to *form* or *maintain* a mental state *just is* for them to be in that mental state because they *believed* that they would be in it; for someone to *give up* or *withhold* a mental state *just is* for them not to be in that mental state because they did not *believe* that they would be in it; and for someone's forming, maintaining, giving up, or withholding a mental state to be based on some other mental state *just is* for the belief that partly constitutes their forming, maintaining, giving up, or withholding that mental state to be based on that attitude. I will also rely on the Reflective Self-Ascription Requirement, defended in chapter three.

(Reflective Self-Ascription Requirement) Rationality requires that someone be in a judgement-sensitive mental state if, and only if, they believe, or are disposed, on reflection, to believe, that they themselves are in that mental state then.

4.2.1.4 The Logic of Rationality Requires

The last two background assumptions that are required before moving on have to do with the logical features of RATIONALLY REQUIRES. These are intuitively plausible principles that follow from the account of rationality presented in the previous chapter.

(Syllogism) For all p, q, r, and s, if rationality requires that $p \supset q$, and rationality requires that $q \,\&\, r \supset s$, then rationality requires that $p \,\&\, r \supset s$.

(Exhaustion) For all p and q, if rationality requires that $p \supset q$, and rationality requires that $\neg p \supset q$, then rationality requires that q.

4.2.2 Enkratic Permissions

Let us begin with the Enkratic permissions that govern normative judgements. These are the principles that permit that an agent form, maintain, give up, or withhold some mental state on the basis of some relevant normative judgement.

4.2.2.1 Must

Recall the Positive Enkratic Permission for MUST presented above. This is the principle that requires that you intend to do what you believe you must do.

(Positive Enkratic Permission) For all *S* and φ, rationality permits that *S*'s forming or maintaining the intention to φ be based on:

- Their belief that they must φ; and
- Their belief that whether they will φ causally depends on whether they intend to φ.

In what follows, I will show that, given the views developed in this thesis, one can reason one's way, by a few short steps, from the belief that one must do something, and the belief that whether one will do it causally depends on whether one intends to do it, to the formation or maintenance of an intention to do it. Before I delve into the details, however, allow me provide a brief overview of the account. Given cognitivism about endorsement, to show that this enkratic permission holds, it suffices to show that one can reason one's way from the premise beliefs to the *belief* that one will intend to do what one believes one must do. The demonstration of this is as follows. First, from the judgement that one must do something, given the analysis of MUST discussed above, one can infer that rationality requires one not to allow oneself not to do it. Secondly, from the belief that rationally requires one not to allow oneself not to do something, charity in interpretation allows one to infer that one will not in fact allow oneself not to do it. And finally, from the belief that one will not allow oneself to do something and the belief that whether one will do it causally depends on whether one intends to do it, one can infer that one will intend to do it—since failing to do so would amount to allowing oneself not to do it. Thus one's belief that one must do something, in conjunction with one's belief that whether one will do it causally depends on whether one intends to do it, can lead one to believe that one will intend to do it. It follows that the Positive Enkratic Permission for MUST is true.

Let us now go through this in more detail. Given cognitivism about endorsement and the above analysis of MUST, the Positive Enkratic Permission for MUST is equivalent to the following principle.

(Positive Enkratic Permission) For all *S* and φ, rationality permits that *S*'s *believing that they will intend* to φ be based on:

- Their belief that *rationality permits allowing oneself to φ, and rationality does not permit allowing oneself not to φ*; and

- Their belief that whether they will φ causally depends on whether they intend to φ.

For ease of reading in what follows, I will express this principle and others like it using the following self-explanatory shorthand.

B: $\Diamond P\varphi$ & $\neg\Diamond P\neg\varphi$
B: $D(\varphi, I\varphi)$

B: $I\varphi$

Our first step is to show that any judgement about what must be the case is closely connected to some judgement about what rationality requires. The following principle is based on the intuitively plausible idea that, if rationality permits that you allow yourself to do something and rationality does not permit that you allow yourself not to do it, then rationality requires that you not allow yourself to do it. This principle is very plausible; indeed, if RATIONALITY PERMITS is a dual of RATIONALITY REQUIRES, then the second conjunct of the premise belief is equivalent to the proposition that features in the conclusion belief.

B: $\Diamond P\varphi$ & $\neg\Diamond P\neg\varphi$

B: $\Box\neg P\neg\varphi$

The next principle is an instance of the Positive Permission of Charity in Self-Ascription presented above. It says that it is rational for you to base the belief that you will not allow yourself not to do something on the belief that rationality requires this of you.

B: $\Box\neg P\neg\varphi$

B: $\neg P\neg\varphi$

The final principle is one of the principles discussed above that follows from the practical account of allowing. It says that rationality permits that you believe that you will intend to do something based on the belief that you will not allow yourself not to do it and the belief that whether you will do it causally depends on whether you intend to do it. As discussed above, this is plausible in so far as, when whether you

do something depends on whether you intend to do it, failing to intend to do it *a priori* entails allowing yourself not to do it.

B: ¬*P*¬φ
B: *D*(φ,*I*φ)

B: *I*φ

Given these three principles, it follows that, from the belief that rationality permits allowing oneself to do something and rationality does not permit allowing oneself not to do it, and the belief that whether you will do it causally depends on whether you intend to do it, you can reason your way, by a few short steps, to the belief that you will intend to do it. Since the route from the former beliefs to the latter belief is so short, I assume that rationality permits that one skip the intervening steps and base the latter belief directly on the former beliefs. So the Positive Enkratic Permission for MUST holds.

B: ◇*P*φ & ¬◇*P*¬φ
B: *D*(φ,*I*φ)

B: *I*φ

Thus, together, the cognitivist account of mental self-regulation from chapter three, the analysis of MUST from chapter one, and the account of rationality from chapter two entail the validity of the Positive Enkratic Permission for MUST.

4.2.2.2 May

Next, consider the Negative Enkratic Permission for MAY, according the which rationality permits that someone give up or withhold an intention to do something based on being in a state of doubt about whether they may do it, along with the belief that it is up to them whether they do it.

(Negative Enkratic Permission) For all *S* and φ, rationality permits that *S*'s withholding or giving up the intention to φ be based on:

- Their not being disposed, on reflection, to believe that they may φ; and
- Their belief that whether they will φ causally depends on whether they intend to φ.

In what follows, I will show that, given the views developed in this thesis, one can reason one's way, by a few short steps, from being in doubt about whether one may do something, and believing that whether one will do it causally depends on whether one intends to do it, to withholding or giving up the intention to do it. Again, however, before I delve into the details, allow me to provide a brief overview. Given cognitivism about endorsement, to show that this enkratic prohibition holds, it suffices to show that one can reason one's way from these mental states to *not believing* that one will intend to do what one is unsure one may do. The demonstration of this is as follows. First, given the analysis of MAY discussed above, to be in doubt about whether one may do something *just is* to be in doubt about whether one is rationally permitted to allow oneself to do it. Secondly, charity in interpretation allows that one be in doubt about whether one will allow oneself to do something, based on being in doubt about whether rationality permits one to do so. And finally, it makes sense not to believe that one will intend to do something based on being in doubt about whether one will allow oneself to do it and believing that whether one will do it causally depends on whether one intends to do it—since intending to do it is obviously a way of allowing oneself to do it. Thus, being in doubt about whether one may do something, in conjunction with the belief that whether one will do it causally depends on whether one intends to do it, can lead one not to believe that one will intend to do it. It follows that the Negative Enkratic Permission for MAY is true.

Now let us go through this in more detail. Given cognitivism about endorsement and the above analysis of MAY, the Negative Enkratic Permission for MAY is equivalent to the following principle.

(Negative Enkratic Permission) For all *S* and φ, rationality permits that *S*'s *not believing that they will intend* to φ be based on:

- Their not being disposed, on reflection, to believe that *rationality permits allowing oneself to φ*; and
- Their belief that whether they will φ causally depends on whether they intend to φ.

Again, we can express this using shorthand for ease of reading.

$\neg DB\colon \Diamond P\varphi$

$B\colon D(\varphi, I\varphi)$

$\neg B\colon I\varphi$

The following principle is an instance of the Negative Permission of Charity in Self-Ascription presented above. It says that it is rational not to be disposed to believe that you will allow yourself to do something based on not being disposed to believe that rationality permits that you do so.

$\neg DB\colon \Diamond P\varphi$

$\neg DB\colon P\varphi$

The next principle is one of the principles discussed above that follows from the practical account of allowing. It says that it is rational for your not believing that you intend to do something to be based on your believing that whether you will do it depends on whether you intend to do it and your not being disposed to believe that you will allow yourself to do it. As discussed above, this is plausible in so far as, when whether you do something depends on whether you intend to do it, your intending to do it *a priori* entails your allowing yourself to do it.

$\neg DB\colon P\varphi$

$B\colon D(\varphi, I\varphi)$

$\neg B\colon I\varphi$

It follows from these two principles that, in so far as you are rational, if you are not disposed to believe that rationality permits that you allow yourself to do something, and you believe that whether you will do it depends on whether you intend to do it, then this may lead you, by a couple of short steps, not to believe that you will intend to do it. Since the route from the former mental states to the latter state of non-belief is so short, I assume that rationality permits that one skip the intervening steps and base the latter state of non-belief on the former mental states. In other words, the above formulation of the Negative Enkratic Permission for MAY holds.

$\neg DB$: $\Diamond P\varphi$

B: $D(\varphi, I\varphi)$

$\neg B$: $I\varphi$

Thus, together, the cognitivist account of mental self-regulation from chapter three, the analysis of MAY from chapter one, and the account of rationality from chapter two entail the validity of the Negative Enkratic Permission for MAY.

4.2.2.3 Ought

Now consider the Negative Enkratic Permission for OUGHT, according to which rationality permits that one give up or withhold an intention to do something based on the belief that one ought not to do it.

(Negative Enkratic Permission) For all *S* and φ, rationality permits that *S*'s giving up or withholding the intention to φ be based on:

- Their belief that they ought not to φ; and
- Their belief that whether they will φ causally depends on whether they intend to φ.

In what follows, I will show that the views developed in this thesis together explain why this principle holds. Again, before delving into the details, allow me to provide a brief overview. Given cognitivism about endorsement, in order to show that this principle is true, it suffices to show that, from the belief that one ought not to do something, and the belief that whether one will do it causally depends on whether one intends to do it, one can be led, by a few short steps, *not to believe* that one will intend to do it. The demonstration of this is as follows. First, given the simple analysis of OUGHT, from the belief that one ought not to do something, one can infer that rationality does not require that one allow oneself to do it. Secondly, charity in interpretation permits that one be in doubt about whether one will allow oneself to do something based on believing that rationality does not require that one allow oneself to do it. And, finally, one's doubt about whether one will allow oneself to do something, and one's belief that whether one will do it causally depends on whether one intends to do it, may lead one to be in doubt about whether one will intend to do it—since intending to do it would obviously involve allowing oneself to do it. Thus, the

belief that one ought not to do something, in conjunction with the belief that whether one will do it causally depends on whether one intends to do it, can lead one not to believe that one will intend to do it. It follows that the Negative Enkratic Permission for OUGHT is true.

Let us now look at this account in more detail. Given cognitivism about endorsement and the simple analysis of OUGHT, the Negative Enkratic Permission for OUGHT is equivalent to the following principle.

(Negative Enkratic Permission) For all *S* and φ, rationality permits that *S*'s *not believing that they will intend* to φ be based on:

- Their belief that, *if rationality did not permit allowing oneself not to φ, or did not permit allowing oneself to φ, then rationality would permit allowing oneself not to φ and would not permit allowing oneself to φ*; and
- Their belief that whether they will φ causally depends on whether they intend to φ.

In shorthand, this is equivalent to the following.

B: $(\neg\Diamond P\neg\varphi \vee \neg\Diamond P\neg\varphi) > (\Diamond P\neg\varphi \,\&\, \neg\Diamond P\varphi)$

B: $D(\varphi, I\varphi)$

$\neg B$: $I\varphi$

We begin by establishing a link between judgements about what ought to be the case and judgements about what rationality requires. The following principle holds in virtue of the fact that allowing yourself to do something *a priori* entails not allowing yourself not to do it. Suppose that you ought not do something. Then, since OUGHT implies MAY, you may not do it, and so, given the above analysis of MAY, rationality permits allowing yourself not to do it. Since allowing yourself not to do something entails not allowing yourself to do it, it follows that it is not the case that rationality requires that you allow yourself to do it.

B: $(\neg\Diamond P\neg\varphi \vee \neg\Diamond P\neg\varphi) > (\Diamond P\neg\varphi \,\&\, \neg\Diamond P\varphi)$

B: $\neg\Box P\varphi$

The next principle follows from the Negative Permission of Charity in Self-Ascription II. It says that rationality permits that your not being disposed to believe that you will allow yourself to do something be based on the belief that rationality does not require that you do so.

$$\frac{B\text{: } \neg\Box P\varphi}{\neg DB\text{: } P\varphi}$$

The next principle follows from the account of practical allowing. It is the same principle that we relied on in the previous section to vindicate the Negative Enkratic Permission for MAY.

$$\frac{\neg DB\text{: } P\varphi \qquad B\text{: } D(\varphi, I\varphi)}{\neg B\text{: } I\varphi}$$

Given these principles, it follows that, in so far as you are rational, if you believe that you ought not do something, and you believe that whether you will do it depends on whether you intend to do it, then this may lead you, by a few short steps, not to believe that you will intend to do it. Since the route from the former beliefs to the latter state of non-belief is so short, I assume that rationality permits that one skip the intervening steps and base the latter directly on the former. So the above formulation of the Negative Enkratic Permission for OUGHT is true.

$$\frac{B\text{: } (\neg\Diamond P\neg\varphi \vee \neg\Diamond P\neg\varphi) > (\Diamond P\neg\varphi \,\&\, \neg\Diamond P\varphi) \qquad B\text{: } D(\varphi, I\varphi)}{\neg B\text{: } I\varphi}$$

Thus, together, the cognitivist account of mental self-regulation from chapter three, the analysis of OUGHT from chapter one, together with the account of rationality from chapter two, vindicate the Negative Enkratic Permission for OUGHT.

4.2.2.4 Fitting-Response Concepts

4.2.2.5 At Least

Recall the Positive Enkratic Permission for AT LEAST AS *A*-ABLE presented above.

(Positive Enkratic Permission) For all *S*, *x*, *y*, *A*, and *B*, rationality permits that *S*'s forming or maintaining the mental state of *A*-ing *x* at least as much as they *B y* be based on their judging that it will be the case that the degree to which *x* is *A*-able (from their own perspective) is at least as great as the degree to which *y* is *B*-able (from their own perspective).

Given cognitivism about endorsement, and the analysis of fitting-response concepts discussed above, this is equivalent to the following.

(Positive Enkratic Permission) For all *S*, *x*, *y*, *A*, and *B*, rationality permits that *S*'s *believing that they will A x at least as much as they B y* be based on their judging that *rationality requires that one will A x at least as much as one Bs y*.

The truth of this principle follows directly from the Positive Permission of Charity outlined above. Thus, together, the analysis of fitting-response concepts and the account of rationality from chapter two, and the cognitivist account of endorsement from chapter three explain why the Positive Enkratic Permission for AT LEAST AS *A*-ABLE holds.

4.2.2.6 On a Par

Finally, recall the Negative Enkratic Permission for ON A PAR presented above.

(Negative Enkratic Permission) For all *S*, *x*, *y*, *A*, and *B*, rationality permits that *S*'s giving up or withholding the mental state of *A*-ing *x* equally as much as they *B y* be based on their not being disposed, on reflection, to judge that it will be the case that the degree to which *x* is *A*-able (from their own perspective) is on a par with the degree to which *y* is *B*-able (from their own perspective).

Given cognitivism about endorsement, and the analysis of fitting-response concepts discussed above, this is equivalent to the following.

(Negative Enkratic Permission) For all *S*, *x*, *y*, *A*, and *B*, rationality permits that *S*'s *not believing that they will A x equally as much as they B y* be based on their not being disposed, on reflection, to judge that *rationality permits that one will A x equally as much as one Bs y*.

The truth of this principle follows directly from the Positive Permission of Charity outlined above. Thus, together, the analysis of fitting-response concepts and the account of rationality from chapter two, and the cognitivist account of endorsement from chapter three explain why that Negative Enkratic Permission for ON A PAR holds.

4.2.3 Enkratic Requirements

We have now seen how cognitivism about endorsement, together with the analyses of normative concepts outlined above and the account of rationality from chapter two, explain the enkratic permissions to which normative judgements are subject. Next, I will show that these views also explain the enkratic requirements to which normative judgements are subject.

4.2.3.1 Must

First, consider the Positive Enkratic Requirement for MUST, according to which rationality requires that, if someone believes that they must do something, and believes that it is up to them then whether they do it, then they intend to do it.

(Positive Enkratic Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that they must φ; and
  - *S* believes at *t* that whether they will φ causally depends on whether they then intend to φ;
- Then:
  - *S* intends at *t* to φ.

In what follows, I will show that the views developed in this thesis together explain why the Positive Enkratic Requirement for MUST holds. Before delving into the details, however, let me provide a brief overview of the account. First, given the analysis of MUST discussed above, if you believe that you must do something, then you are committed to believing that you are rationality required not to allow yourself not to do it. Secondly, charity in interpretation requires that, if you believe this, then you believe that you will not in fact allow yourself not to do it. Thirdly, this belief, in conjunction with the belief that whether you will do it depends on whether you then intend to do it,

commits you to a disposition to believe that you then intend to do it—since not then intending to do it would amount to allowing oneself not to do it. And, finally, given cognitivism about endorsement, being disposed to self-ascribe this intention commits you to in fact having the intention. Thus the belief that you must do something, in conjunction with the belief that whether you will do it depends on whether you then intend to do it, rationally commits you to intending to do it.

Now let us go through this demonstration in more detail. Given the above analysis of MUST, the Positive Enkratic Requirement for MUST is equivalent to the following principle.

(Positive Enkratic Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that *rationality permits allowing oneself to φ and rationality does not permit allowing oneself not to φ*; and
  - *S* believes at *t* that whether they will φ causally depends on whether they then intend to φ;
- Then:
  - *S* intends at *t* to φ.

In order to show that the views defended in this thesis together vindicate the Positive Enkratic Requirement for MUST, I will present four rational requirements, each of which is entailed by one or other of these views or is otherwise independently plausible, which together entail this principle. We begin with an intuitively plausible principle that requires that, if you believe that rationality permits allowing something to be so and rationality does not permit allowing it not to be so, then you believe that rationality requires not allowing it not to be so. That there is a close connection between these two types of judgements is very plausible; indeed, as mentioned above, if RATIONALITY PERMITS and RATIONALITY REQUIRES are duals, then the second conjunct of the first proposition is equivalent to the second proposition. Thus, for all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that rationality permits allowing oneself to φ and rationality does not permit allowing oneself not to φ;

- Then:
  - *S* believes at *t* that rationality requires not allowing oneself not to φ.

Next, we rely on a principle that is an instance of the Positive Requirement of Charity in Self-Ascription presented above. It says that rationality requires that, if you believe that rationality requires that you not allow yourself not to do something, then you believe that you will not allow yourself not to do it. Thus, for all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that rationality requires not allowing oneself not to φ;
- Then:
  - *S* believes at *t* that they will not allow themselves not to φ.

The next principle is one of the requirements discussed above that follows from the practical account of allowing. It says that, in so far as you are rational, if you believe that you will not allow yourself not to do something, and you believe that whether you will do it depends on whether you then intend to do it, then you are disposed, on reflection, to believe that you then intend to do it. That is, for all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that they will not allow themselves not to φ; and
  - *S* believes at *t* that whether they will φ causally depends on whether they then intend to φ;
- Then:
  - At *t*, *S* is disposed, on reflection, to believe that they themselves then intend to φ.

Finally, we draw on an instance of the Reflective Self-Ascription Requirement. This principle says that rationally requires that, if you are disposed at some time to believe that you then intend to do something, then you intend to do it.

For all *S*, *t*, and φ, rationality requires that:

- If:

- At *t*, *S* is disposed, on reflection, to believe that they themselves then intend to φ;
- Then:
  - At *t*, *S* intends to φ.

These four rational requirements, together (by Syllogism) entail the Positive Enkratic Requirement for MUST. Let me summarise the argument again. First, believing that you must do something commits you to believing that rationality requires not allowing yourself not to do it. This in turn commits you to believing that you will in fact not allow yourself not to do it. This belief, in conjunction with the belief that whether you will do it depends on whether you then intend to do it, commits you to a disposition to believe that you then intend to do it. And, finally, being disposed to self-ascribe this intention commits you to in fact having the intention. Thus the belief that you must do something, in conjunction with the belief that whether you will do it depends on whether you then intend to do it, commits you to intending to do it.

4.2.3.2 May

Next consider the Negative Enkratic Requirement for MAY, which requires that you intend to do something only if you are disposed to judge that you may do it.

(Negative Enkratic Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
  - At *t*, *S* is not disposed, on reflection, to judge that they may φ;
- Then:
  - *S* does not intend at *t* to φ.

In what follows, I will show that the views developed in this thesis explain why this principle holds. Before delving into the details, however, allow me to provide a brief overview of this account. First, given the analysis of MAY discussed above, to be in doubt about whether you may do something *just is* to be in doubt about whether rationality permits that you allow yourself to do it. Secondly, charity in interpretation requires that, if you are in doubt about whether rationality permits that you allow yourself to do something, then you are in doubt about whether you will in fact allow yourself to do it. Now, suppose that you are in doubt about whether you will allow

yourself to do something. Suppose that you are also confident that whether you will do it causally depends on whether you intend to do it. Then you are required not to be disposed to believe that you now intend to do it—since that would obviously involve allowing yourself to do it. Given cognitivism about endorsement, this in turn commits you to not intending to do it. Now suppose instead that you are not confident that whether you will do it causally depends on whether you intend to do it. Then, since intending to do something requires being confident that whether you will do it depends on whether you intend to do it, you are committed to not intending to do it. So, either way, if you are in doubt about whether you may do something, then you are committed to not intending to do it. Thus, the Negative Enkratic Requirement for MAY holds.

Let us now go through this in more detail. The first step is to point out that, given the above analysis of MAY, the Negative Enkratic Requirement for MAY is equivalent to the following principle.

(Negative Enkratic Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
  - At *t*, *S* is not disposed, on reflection, to believe that *rationality permits allowing oneself to φ*;
- Then:
  - *S* does not intend at *t* to φ.

We begin with a principle that is an instance of the Negative Requirement of Charity I presented above. It says that, if you are not disposed to believe that rationality permits that you allow yourself to do something, then you are not disposed to believe that you will allow yourself to do it. Thus, for all *S*, *t*, and φ, rationality requires that:

- If:
  - At *t*, *S* is not disposed, on reflection, to believe that rationality permits allowing oneself to φ;
- Then:
  - At *t*, *S* is not disposed, on reflection, to believe that they will allow themselves to φ.

The next principle follows from the practical account of allowing. It says that, if you are not disposed to believe that you will allow yourself to do something, and you are sufficiently confident that whether you will do it depends on whether you intend to do it, then you are not disposed to believe that you then intend to do it. This is based on the idea that, when whether you do something depends on whether you intend to do it, your then intending to do it *a priori* entails your allowing yourself to do it. So if you believe that whether you will do something depends on whether you intend to do it, then rationality requires that you be disposed to believe that you then intend to do it only if you are disposed to believe that you will allow yourself to do it. Thus, for all *S*, *t*, and φ, rationality requires that:

- If:
  - At *t*, *S* is not disposed, on reflection, to believe that they will allow themselves to φ; and
  - At *t*, *S* is sufficiently confident that whether they will φ causally depends on whether they intend to φ;
- Then:
  - At *t*, *S* is not disposed, on reflection, to believe that they then intend to φ.

Finally, the following principle is an instance of the Reflective Self-Ascription Requirement. It says that you are rationally required not to intend to do something, if you are not disposed to believe that you intend to do it. Thus, for all *S*, *t*, and φ, rationality requires that:

- If:
  - At *t*, *S* is not disposed, on reflection, to believe that they themselves then intend to φ;
- Then:
  - At *t*, *S* does not intend to φ.

These three rational requirements entail the following (by Syllogism). For all *S*, *t*, and φ, rationality requires that:

- If:
  - At *t*, *S* is not disposed, on reflection, to believe that rationality permits allowing oneself to φ; and

    - At *t*, *S* is sufficiently confident that whether they will φ causally depends on whether they intend to φ;
- Then:
    - At *t*, *S* does not intend to φ.

Let me summarise the argument so far. Suppose you are not disposed to believe that you may do something. Then you are not disposed to believe that rationality permits that you allow yourself to do it. This commits you to being not disposed to believe that you will allow yourself to do it, which, in conjunction with being confident that whether you will do it depends on whether you then intend to do it, in turn commits you to being not disposed to believe that you then intend to do it. Finally, being not disposed to self-ascribe an intention commits you to not holding that intention. Thus, being not disposed to believe that you may do something, in conjunction with being confident that whether you will do it depends on whether you then intend to do it, commits you to not intending to do it.

In the discussion of the Negative Enkratic Requirement for MAY above, I presented an independently plausible rational requirement on intention that I called the 'Self-Efficacy Requirement'.

(Self-Efficacy Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
    - *S* is not sufficiently confident at *t* that whether they themselves will φ causally depends on whether they themselves intend to φ;
- Then:
    - *S* does not intend at *t* to φ.

Given that the Self-Efficacy Requirement holds, it follows that, if someone is not disposed to judge that rationality permits that they allow themselves to do something, then, whether or not they are confident that whether they will do it depends on whether they intend to do so, rationality requires that they not intend then to do it. Thus, the Negative Enkratic Requirement for MAY follows (by Exhaustion) from the previous two rational requirements.

4.2.3.3 Ought

Next, consider the Negative Enkratic Requirement for OUGHT, which requires that you not intend to do something if you believe that you ought not do it and you believe that whether you will do it causally depends on whether you intend to do it.

(Negative Enkratic Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that they ought not to φ; and
  - *S* believes at *t* that whether they will φ causally depends on whether they then intend to φ;
- Then:
  - *S* does not intend at *t* to φ.

In what follows, I will show that the views developed in this thesis together explain why this principle holds. Before delving into the details of this account, however, let me provide a brief overview of it. First, given the simple analysis of OUGHT discussed above, believing that you ought not to do something commits you to believing that rationality does not require that you allow yourself to do it. Charity in interpretation requires that, if you believe this, then you are not disposed to believe that you will in fact allow yourself to do it. If you believe this, and you believe that whether you will do it causally depends on whether you then intend to do it, then you are committed to not being disposed to believe that you then intend to do it—since doing so would obviously involve allowing yourself to do it. Finally, given cognitivism about endorsement, not being disposed to self-ascribe an intention to do something commits one to not intending to do it. Thus, judging that you ought not do something, and believing that whether you will do it causally depends on whether you now intend to do it, rationally commits you to not intending to do it.

Let us now go through this demonstration in more detail. First, given the analysis of OUGHT above, the Negative Enkratic Requirement for OUGHT is equivalent to the following principle.

(Negative Enkratic Requirement) For all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that *if rationality did not permit allowing oneself not to φ, or did not permit allowing oneself to φ, then rationality would permit allowing oneself not to φ and would not permit allowing oneself to φ*; and
  - *S* believes at *t* that whether they will φ causally depends on whether they then intend to φ;
- Then:
  - *S* does not intend at *t* to φ.

The following principle holds in virtue of the fact that allowing oneself to do something *a priori* entails not allowing oneself not to do it. Suppose you believe that that you ought not do something. This commits you to believing that you may not do it, and so, given the above analysis of MAY, to believing that rationality permits allowing yourself not to do it. Since allowing yourself not to do something entails not allowing yourself to do it, it follows that you are committed to believing that it is not the case that rationality requires that you allow yourself to do it. Thus, for all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that, if rationality did not permit allowing oneself not to φ, or did not permit allowing oneself to φ, then rationality would permit allowing oneself not to φ and would not permit allowing oneself to φ;
- Then:
  - *S* believes at *t* that it is not the case that rationality requires allowing oneself to φ.

The next principle is an instance of the Negative Requirement of Charity II presented above. It says that, if you believe that it is not the case that rationality requires that you allow yourself to do something, then you are rationally required not to be disposed to believe that you will allow yourself to do it. Thus, for all *S*, *t*, and φ, rationality requires that:

- If:
  - *S* believes at *t* that it is not the case that rationality requires allowing oneself to φ;

- Then:
  - At *t*, *S* is not disposed to believe that they will allow themselves to φ.

The next principle is one of the principles that follows from the practical account of allowing. It says that, if you are not disposed to believe that you will allow yourself to do something, and you believe that whether you will do it depends on whether you then intend to do it, then you are not disposed to believe that you then intend to do it. Thus, for all *S*, *t*, and φ, rationality requires that:

- If:
  - At *t*, *S* is not disposed, on reflection, to believe that they will allow themselves to φ; and
  - At *t*, *S* believes that whether they will φ causally depends on whether they intend to φ;
- Then:
  - At *t*, *S* is not disposed, on reflection, to believe that they then intend to φ.

The final principle was used in the previous section and follows from the Reflective Self-Ascription Requirement: for all *S*, *t*, and φ, rationality requires that:

- If:
  - At *t*, *S* is not disposed, on reflection, to believe that they themselves then intend to φ;
- Then:
  - At *t*, *S* does not intend to φ.

These four rational requirements together (by Syllogism) entail the Negative Enkratic Requirement for OUGHT. Once again, let me summarise the argument. Suppose that you believe that you ought not do something. This commits you to believing that rationality does not require that you allow yourself to do it. This belief, in turn, commits you to being not disposed to believe that you will allow yourself to do it, which, in conjunction with the belief that whether you will do it depends on whether you then intend to do it, in turn commits you to being not disposed to believe that you then intend to do it. Finally, being not disposed to self-ascribe an intention commits you to not holding that intention. Thus, believing that you ought not do something, in conjunction with the belief that whether you will do it depends on whether you then

intend to do so, commits you to not intending to do it. Thus, in this section I have shown that, taken together, the analyses of MAY, MUST, and OUGHT from chapter one, the account of rationality in chapter two, and the cognitivist account of endorsement in chapter three, are able to vindicate the Positive Enkratic Requirement for MUST, the Negative Enkratic Requirement for MAY, and the Negative Enkratic Requirement for OUGHT.

4.2.3.4 Fitting-Response Concepts

Next, I show that the views defended in this thesis also explain the enkratic requirements that apply to judgements about fitting responses that involve the concepts AT LEAST and ON A PAR.

4.2.3.5 At Least

Recall the Positive Enkratic Requirement for AT LEAST AS *A*-ABLE discussed above.

(Positive Enkratic Requirement) For all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:
  - *S* judges at *t* that the degree to which *x* is then *A*-able (from their own perspective) is at least as great as the degree to which *y* is then *B*-able (from their own perspective);
- Then:
  - *S A*s *x* at *t* at least as much as they *B y* at *t*.

Before I give a detailed account of how the views developed in this thesis explain why this principle holds, allow me to provide a brief overview of this account. First, given the analysis of fitting-response concepts discussed above, to judge that, say, one thing is at least as desirable as another, from one's own perspective *just is* to judge that one is rationally required to desire the first thing at least as much as the second. Secondly, charitable interpretation requires that, if one judges that one is rationally required to desire one thing at least as much as another, then one believes that this is so. And, finally, given cognitivism about endorsement, believing that one desires one thing at least as much as another rationally commits one to in fact desiring the former at least as much as the latter. Thus, judging that one thing is at

least as desirable as another thing is, from one's own perspective, rationally commits one to desiring the former at least as much as the latter. It follows that the Positive Enkratic Requirement for AT LEAST AS *A*-ABLE holds.

Let us now go through this demonstration in more detail. First, given the analysis of fitting-response concepts discussed above, the Positive Enkratic Requirement for AT LEAST AS *A*-ABLE is equivalent to the following.

(Positive Enkratic Requirement) For all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:
  - *S* judges at *t* that *rationality requires A-ing x at least as much as one Bs y then*;
- Then:
  - *S As x* at *t* at least as much as they *B y* at *t*.

In order to explain why this principle holds, we begin with a principle that is an instance of the Positive Requirement of Charity outlined above. It says that, if you believe that you are rationally required, say, to desire one thing more than another, then you believe that you desire the first thing more than the second thing. More specifically, for all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:
  - *S* judges at *t* that rationality requires *A-ing x* at least as much as one *Bs y* then;
- Then:
  - *S* believes at *t* that they *A x* at least as much as they *B y* then.

The next principle is an instance of the Self-Ascription Requirement. It says that rationality requires that, if you believe that, say, you desire one thing more than another, then this is in fact so. That is, for all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:
  - *S* believes at *t* that they *A x* at least as much as they *B y* then;
- Then:

- *S A*s *x* at least as much as they *B y* at *t*.

From these two rational requirements, it follows (by Syllogism) that the Positive Enkratic Requirement for AT LEAST AS *A*-ABLE holds. In other words, judging that, say, one thing is at least as desirable as another commits you to judging that you desire the former at least as much as the latter, which in turn commits you to in fact desiring the former at least as much as the latter.

4.2.3.6 On a Par

Finally, recall the Negative Enkratic Requirement for ON A PAR.

(Negative Enkratic Requirement) For all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:
  - *S* is not disposed at *t*, on reflection, to judge that the degree to which *x* is then *A*-able (from their own perspective) is on a par with the degree to which *y* is then *B*-able (from their own perspective);
- Then:
  - At *t*, *S* does not *A x* equally as much as they *B y*.

Before I provide a detailed account of how the views developed in this thesis explain why this principle holds, allow me to provide a brief overview of it. First, given the analysis of ON A PAR, to be in doubt about whether the degrees to which two things are, say, desirable, are on a par, from one's own perspective, *just is* to be in doubt about whether one is rationally permitted to desire them equally. Secondly, charitable interpretation requires that, if one is in doubt about whether one is rationally permitted to desire two thing equally, then one is not disposed to believe that one desires them equally. And, finally, given cognitivism about endorsement, if one is not disposed to believe that one desires two things equally, then one is required not to desire them equally. Thus, if one is in doubt about whether the degrees to which two things are desirable are on a par, from one's own perspective, then one is required not to desire them equally, and so the Negative Enkratic Requirement for ON A PAR is true.

Now let us go through this demonstration in detail. First, given the analysis of fitting-response concepts discussed above, the Negative Enkratic Requirement for ON A PAR is equivalent to the following principle.

(Negative Enkratic Requirement) For all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:
  - *S* is not disposed at *t*, on reflection, to judge that *rationality permits A-ing x equally as much as one Bs y then*;
- Then:
  - At *t*, *S* does not *A x* equally as much as they *B y*.

In order to explain why this principle holds, we begin with a principle that is an instance of the Negative Permission of Charity I outlined above. This principle says that, if you are not disposed to believe that rationality permits that you, say, desire two things equally, then you not be disposed to believe that you desire them equally. That is, for all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:
  - *S* is not disposed at *t*, on reflection, to judge that rationality permits *A*-ing *x* equally as much as one *B*s *y* then;
- Then:
  - *S* is not disposed at *t*, on reflection, to believe that they *A x* equally as much as they *B y*.

The next principle is an instance of the Reflective Self-Ascription Principle. It says that rationality requires that, if you are not disposed to self ascribe the mental state of, say, desiring two things equally, then you do not desire them equally. That is, for all *S*, *t*, *x*, *y*, *A*, and *B*, rationality requires that:

- If:
  - *S* is not disposed at *t*, on reflection, to believe that they *A x* equally as much as they *B y*;
- Then:
  - *S* does not *A x* equally as much as they *B y* at *t*.

From these two rational requirements, then, it follows (by Syllogism) that the Negative Enkratic Principle for ON A PAR holds. Not being disposed to judge that, say, two things are on a par with respect to their desirability commits one to being not disposed to believe that you desire them equally, which in turn commits you to not desiring them equally.

4.2.4 Enkratic Prohibitions

We have now seen that, together, the analyses of normative concepts presented in chapters one and two, the account of rationality presented in chapter two, and the cognitivist account of endorsement presented in chapter three explain why normative judgements are subject to enkratic permissions and requirements. All that remains now is to make sense of the enkratic prohibitions that apply to normative judgements. Showing that the views developed in this thesis explain why normative judgements are subject to enkratic permissions and requirements was an admittedly somewhat complex affair. Each demonstration involved multiple steps and drew on several different rational principles. By contrast, showing that the views developed in this thesis explain why normative judgements are subject to enkratic prohibitions is straightforward.

Enkratic prohibitions prohibit that one's forming, maintaining, giving up, or withholding a normative judgement be based on some other relevant mental state. For example, in the case of judgements involving the concepts MUST, MAY, and OUGHT, these principles prohibit that one form, maintain, give up, or withhold a judgement about whether one must, may, or ought to do something based on one's intending or not intending to do it. According to cognitivism about endorsement, to form, maintain, give up, or withhold a normative judgement *just is* to believe that one will or will not hold that judgement, and for one's forming, maintaining, giving up, or withholding that judgement to be based on some other attitude *just is* for the belief that partly constitutes the forming, maintaining, giving up, or withholding of that judgement to be based on that attitude. Given this cognitivist view of mental regulation, enkratic prohibitions can easily be explained by the general principle that believing or not believing something cannot be rationally based on other kinds of mental states such as (not) intending or (not) desiring.

(Basing Prohibition for Belief) For all *p* and *q*, rationality prohibits that believing or not believing that *p* be based on intending or not intending that *q* or desiring or not desiring that *q* or … .

Consider, for example, the Positive Enkratic Prohibition for MUST.

(Positive Enkratic Prohibition) For all *S* and φ, rationality prohibits that *S*'s forming, maintaining, withholding or giving up the belief that they themselves must φ be based on their intending, or not intending, to φ.

Given Cognitivism about Endorsement, this is equivalent to the following principle.

(Positive Enkratic Prohibition) For all *S* and φ, rationality prohibits that *S*'s *believing or not believing that they themselves will judge that they themselves must φ* be based on their intending, or not intending, to φ.

This principle is straightforwardly entailed by the Basing Prohibition for Belief. Thus cognitivism about endorsement, in conjunction with the Basing Prohibition for Belief, easily explains the Positive Enkratic Prohibition for MUST. Since all of the enkratic prohibitions set out above can be explained in the same way, I will not go through them all individually here.

4.3 Conclusion

I began this chapter by carefully considering how the enkratic principles that apply to normative judgements are best formulated. I have now demonstrated that these principles, so formulated, are vindicated by the following combination of views: the analyses of normative judgements from chapters one and two, the account of rationality from chapter two, and the cognitivist account of endorsement from chapter three. The fact that this combination of views is capable of vindicating these Enkratic principles provides additional reason to believe that they are correct.

5 Conclusion

In this thesis, I have argued for a combination of three main views: a rationality-first account of normativity; an account of rationality in terms of its role in interpretation; and a cognitivist account of reflective endorsement. The overarching argument for this combination of views is an abductive one. The explanandum is the fact that normative judgements are subject to various rational principles: there are rational principles that govern how normative judgements are related to each other; there are distinctive epistemic principles that apply to normative judgements; and there are enkratic principles that govern how normative judgements relate to other kinds of attitude.

Together, the three views defended in this thesis explain why normative judgements are subject to these rational principles: the rationality-first analysis of normative concepts explains why normative judgements are related to each other by the rational principles by which they are; the rationality-first analysis of normative concepts and the account of rationality in terms of its role in interpretation explains why normative judgements are subject to the epistemic principles to which they are subject; and the rationality-first analysis of normative concepts, the account of rationality in terms of its role in interpretation, and the cognitivist account of reflective endorsement together explain why normative judgements are subject to enkratic principles, as shown in chapter four.

In each chapter, I have also tried to provide some independent reasons to think that each of these main views is viable in its own right. In chapters one and two, and an appendix I developed a novel rationality-first account of normativity. I presented a fitting-response analysis of MAY and MUST, a counterfactual analysis of OUGHT, a closeness account of REASON (mass) and REASON (count), and an analysis of fitting-response concepts in terms of the concept RATIONAL. I argued for these analyses by showing that they validate the rational principles that govern how judgements involving these concepts relate to each other, and by showing that they satisfy various other desiderata.

In chapter two, I presented an account of the concept of rationality in terms of its role in interpretation by identifying two distinctive facts about the concept: first, there are *a*

priori conceptual truths about the rationality of mental states; and, secondly, there are principles of charitable interpretation according to which judgements about rationality rationally constrain the ascription of mental states. Finally, in chapter three, I presented an account of reflective endorsement according to which to endorse a mental state is simply to self-ascribe that mental state. I provided an independent argument for this position by showing that self-ascriptions are subject to rational principles that are analogous to the rational principles to which endorsement is subject, namely the Reflective Endorsement Requirement and the Positive and Negative Reflective Endorsement Permissions.

I have left a number of central meta-ethical debates unresolved in this thesis. I have not resolved the debate between meta-ethical cognitivists and non-cognitivists, although I have removed one of the main motivations for non-cognitivism: I have shown how to make sense of the close connection between normative judgements and motivation in a way that does not require taking normative judgements to be non-cognitive states. I have not provided an account of normative *language* in this thesis, since I have focused throughout on the *mental state* of normative judgement and its constituent concepts. Neither have I provided an account of normative *metaphysics*. Exploring the metaphysical implications of the account of rationality in chapter two is a worthy avenue for future research.

There is room for more work to be done on the account of fitting-response concepts in chapter two. As I noted in that chapter, the analysis that I provided there is in an important respect incomplete, since it cannot make sense of judgements that compare the fittingness of responses across different worlds and from different perspectives. There are also connections between fitting-response concepts and other normative concepts that I did not address. For example, I did not discuss how to make sense of the obvious connections between judgements involving fitting-response concepts and judgements about reason(s) of the *right kind*.

Neither did I investigate the interesting interconnections between the fitting-response concepts ADVISABLE and CRITICIZABLE, and the concept REASON (mass). ADVISABLE and CRITICIZABLE are important fitting-response concepts because judgements about advisability and criticizability seem to bear rational relations to judgements about reason that judgements involving other fitting-attitude concepts do not. Intuitively,

there is a close connection between how *advisable* it is for someone to do something and how much *reason* they have to do it, and how *criticizable* someone is for doing something and how much *reason* they had to do it. Providing an account of advice and criticism that makes sense of the connections between these concepts, as they have been treated in this thesis, is another worthy avenue for future research.

Finally, I did not investigate the interesting connections between the fitting-response concept BLAMEWORTHY and the concept of MORAL REASON (mass). Judgements about blameworthiness seem to bear rational relations to judgements about *moral reason* that judgements involving other fitting-attitude concepts do not: intuitively, there is a close connection between how *blameworthy* someone is for doing something and how much *moral reason* they had to do it. Extending the views developed in this thesis to incorporate judgements about blameworthiness and moral reason is a worthy avenue for future research, in so far is it has potential to shed some light on the question of the relation between morality and normativity.

6 Appendix: The Closeness Analysis of REASON

In chapter one, I provided an analysis of OUGHT, according to which to judge that something ought to be so is to judge that, if it were not the case that it may be so, or not the case that it may not be so, then it would be the case that it must be so. We may sum up this idea by saying that to judge that something ought to be so is to judge that it is *closer* to being required than it is to being not permitted. In this appendix, I provide an analysis of REASON (mass) and REASON (count) that builds on this idea. According this account, judgements involving the concepts REASON (mass) and REASON (count) are judgements about the relative *closeness* of worlds at which certain deontic facts hold. The closeness account can be given the following rough gloss: how much reason there is for something is a matter of how close it is to being *permitted*, and how far it is from being *not permitted*; and how strong or weighty a reason is in favour of something is a matter of how close it comes to *explaining why* it is *permitted*, and how far it is from *not explaining why* it is *permitted*. I argue for this view by showing that it satisfies several key desiderata for an analysis of this kind.

6.1 Desiderata for an Analysis of REASON (Mass)

6.1.1 Desideratum 1: Makes Sense of the Idea of Different Amounts of Amounts of Reason

The English word ‘reason’ is sometimes used as a mass noun and sometimes as a count noun. When it is used as a mass noun, the word expresses a concept that is pervasive in our normative thinking: much of our normative thinking involves comparing options with regard to *how much reason* there is in their favour. When ‘reason’ is used as a count noun, by contrast, it often refers to a *normative reason*—a consideration or fact that counts in favour of some response. While there is a substantial meta-ethical literature on normative *reasons*, *reason* has received comparatively little attention, though it has recently come under discussion by Broome (2018) and Fogal (2016).

An account of this notion must be able to make sense of the idea that there can be different *amounts* of reason for different responses. Efforts to make sense of this fact have been made by Finlay (2014), who understands amounts of reason in terms of

degrees of *goodness*; and Kearns and Star (2008; 2009; 2013), who understand amounts of reason as amounts of *evidence*. In this section, I set out a number of distinctive rational principles to which judgements that compare amounts of reason are subject. It is a desideratum for an analysis of REASON (mass) that it validate these principles.

6.1.1.1 More Reason

First, MORE REASON picks out a relation that has the properties of a strict partial order—asymmetry and transitivity.

Asymmetry

There is more reason for it to be the case that *p* than there is for it to be the case that *q*.

It is not the case that there is more reason for it to be the case that *q* than there is for it to be the case that *p*.

Transitivity

There is more reason for it to be the case that *p* than there is for it to be the case that *q*; there is more reason for it to be the case that *q* than there is for it to be the case that *r*.

There is more reason for it to be the case that *p* than there is for it to be the case that *r*. (?)

The rational principles expressed by the instances of Transitivity are intuitively plausible, but are controversial because some believe that they are vulnerable to counterexamples, such as those generated by spectrum arguments (Temkin, 2012). For example, recall the case described in chapter two in which you are presented with a list of *n* painful experiences. Item 1 on the list is a month of non-stop excruciating pain, item *n* is a lifetime of mild hangnail pain, and each successive item on the list is slightly less intense, but significantly longer than its predecessor. Suppose you are forced to choose to undergo one of these experiences. One might think, regarding this case, that, for all $1 \leq m < n$, it is rational simultaneously to judge that there is more reason to choose to undergo experience *m* than there is to choose to undergo experience $m + 1$, and that it is *not* the case that there is more reason to choose to undergo experience 1 than there is to choose to undergo experience *n*,

since a lifetime of mild hangnail pain seems preferable to even a month of non-stop excruciating torture. If this is so and Transitivity is valid, however, then, as shown above, it follows that rationality permits one to be in a state that involves holding contradictory judgements, which is false. In spite of its initial intuitive appeal, then, Transitivity is controversial.

6.1.1.2 Equal Reason

EQUAL REASON picks out a relation has the properties of an equivalence relation—reflexivity, symmetry, and transitivity.

Reflexivity

-

There is equally as much reason for it to be the case that *p* as there is for it to be the case that *p*.

Symmetry

There is equally as much reason for it to be the case that *p* as there is for it to be the case that *q*.

There is equally as much reason for it to be the case that *q* as there is for it to be the case that *p*.

Transitivity

There is equally as much reason for it to be the case that *p* as there is for it to be the case that *q*; there is equally as much reason for it to be the case that *q* as there is for it to be the case that *r*.

There is equally as much reason for it to be the case that *p* as there is for it to be the case that *r*.

6.1.1.3 At Least as Much Reason

Intuitively, the concept of AT LEAST AS MUCH REASON *just is* the concept of MORE REASON OR EQUALLY AS MUCH REASON. AT LEAST AS MUCH REASON thus appears to pick out a relation that has the properties of a quasi order—reflexivity and transitivity.

Reflexivity

-

There is at least as much reason for it to be the case that *p* as there is for it to be the case that *p*.

Transitivity

There is at least as much reason for it to be the case that *p* as there is for it to be the case that *q*; there is at least as much reason for it to be the case that *q* as there is for it to be the case that *r*.

There is at least as much reason for it to be the case that *p* as there is for it to be the case that *r*. (?)

Again, the rational principles expressed by the instances of Transitivity are intuitively plausible, but are controversial because some believe that they are vulnerable to counterexamples, such as the kinds of spectrum arguments discussed above.

6.1.1.4 On a Par

As discussed above, parity is a reflexive and symmetrical relation, and EQUAL implies ON A PAR.

Reflexivity

-

The amount of reason there is for it to be the case that *p* is on a par with the amount of reason there is for it to be the case that *p*.

Symmetry

The amount of reason there is for it to be the case that *p* is on a par with the amount of reason there is for it to be the case that *q*.

The amount of reason there is for it to be the case that *q* is on a par with the amount of reason there is for it to be the case that *p*.

6.1.1.5 EQUAL implies ON A PAR

There is equally as much reason for *p* to be the case as there is for *q* to be the case.

The amount of reason there is for *p* to be the case is on a par with the amount of reason there is for *q* to be the case.

6.1.1.6 Conclusive Reason, Sufficient Reason, and More Reason

Finally, the concepts CONCLUSIVE REASON, SUFFICIENT REASON and MORE REASON differ in strength in a way that mirrors the way that MAY, OUGHT, and MUST differ in strength.

There is conclusive reason for it to be the case that *p*.

There is sufficient reason for it to be the case that *p*.

There is conclusive reason for it to be the case that *p*.

There is insufficient reason for it to be the case that ¬*p*.

There is conclusive reason for it to be the case that *p*.

There is more reason for it to be the case that *p* than there is for it to be the case that ¬*p*.

One might think that the following schema is valid too.

There is sufficient reason for it to be the case that *p*;
there is more reason for it to be the case that *p* than there is for it to be the case that ¬*p*.

There is conclusive reason for it to be the case that *p*. (×)

This schema seems invalid to me, however. To see this, suppose that you have been offered two jobs. Each of the offers is excellent, and you would be very happy to accept either of them. Nonetheless, one of the offers is slightly better than the other. In this case, it seems to me that, although there is more reason for you to accept one of the offers than the other, there is sufficient reason for you to accept either of them, since each is well and truly good enough. But if there is sufficient reason to accept either of them, then it cannot be the case that there is conclusive reason to accept one of them. Some philosophers appear to use the expression ‘conclusive reason’ to express something like the concept MOST REASON (cf. Schroeder, 2015, §3; Parfit, 2011, p. 32). When the expression is used in this way, this schema will appear to be

valid. My problem with using the expression in this way is that, as long as cases like the one I just described are conceivable, doing so fails to capture the intuitive entailment from CONCLUSIVE REASON to INSUFFICIENT REASON NOT TO.

6.1.2 Desideratum 2: Validates the Connections between Reason (Mass) and May, Must, and Ought

Our second desideratum for an analysis of REASON (mass) is that it validate the rational connections between judgements about reason and judgements involving the concepts MAY, MUST, and OUGHT. Intuitively, something must be the case just in case there is conclusive reason for it to be the case; something may be the case just in case there is sufficient reason for it to be the case; and something ought to be the case just in case there is sufficient reason for it to be the case and more reason for it to be the case than there is for it not to be the case.

It must be that *p*.

There is conclusive reason for it to be the case that *p*.

There is conclusive reason for it to be the case that *p*.

It must be that *p*.

It may be that *p*.

There is sufficient reason for it to be the case that *p*.

There is sufficient reason for it to be the case that *p*.

It may be that *p*.

It ought to be that *p*.

There is sufficient reason for it to be the case that *p*.

It ought to be that *p*.

There is more reason for it to be the case that *p* than there is for it to be the case that ¬*p*.

There is sufficient reason for it to be the case that *p*; there is more reason for it to be the case that *p* than there is for it to be the case that ¬*p*.

It ought to be that *p*.

6.1.2.1 Ownership of Obligations and Reason

In chapter one, we looked at the distinction between 'owned' and 'un-owned' obligations, recommendations, and permissions. Recall Broome's examples.

1. Alison ought to get a sun hat.

2. Alex ought to get a severe punishment.

Following Broome, I expressed the ownership of obligations, recommendations, and permissions in the following way.

> *S* must/ought/may that *p*.

A similar distinction can be made when it comes to quantities of reason. We may talk about how much reason *there is* for something to be the case, and we may talk about how much reason someone *has* to have some response. For example:

3. Alison has some reason to get a sun hat.

4. There is some reason for Alex to get a severe punishment.

Intuitively, whether an agent 'owns' a requirement, permission, or recommendation depends on how much reason they *have* in favour of what is required, permitted, or recommended, rather than how much reason *there is* in favour of it (Broome, 2013, p. 65). When it comes to obligations and permissions that are owned in this way, then, it seems correct to say that someone must do something just in case they have conclusive reason to do it; someone may do something just in case they have sufficient reason to do it; and someone ought to do something just in case they have sufficient reason to do it and more reason to do it than to do anything else.

S must that *S* φs.

S has conclusive reason to φ.

S has conclusive reason to φ.

S must that *S* φs.

S may that *S* φs.

S has sufficient reason to φ.

S has sufficient reason to φ.

S may that *S* φs.

S ought that *S* φs.

S has sufficient reason to φ.

S ought that *S* φs.

S has more reason to φ than not to φ.

S has sufficient reason to φ; *S* has more reason to φ than not to φ.

S ought that *S* φ.

An analysis of REASON (mass) must validate the rational connections between judgements involving this concept and judgements involving the concepts MAY, MUST, and OUGHT, and explain the distinction between the amount of reason that *there is* and the amount of reason that *someone has*.

6.2 The Closeness Analysis of REASON (Mass)

In this section, I provide a new analysis of the concept REASON (mass) that is a natural companion to the counterfactual analysis of OUGHT presented in chapter one. According to this analysis, all judgements involving the concept REASON (mass) are judgements about *amounts of reason*, where the amount of reason that there is for a response is, roughly, a measure of how *close* it is to being permitted, and how *far* it is from not being permitted. On this view, if something is permitted, then the amount of reason in favour of it is, roughly, a matter of how different the world would need to be, normatively speaking, in order for it not to be permitted; and, if something is not permitted, then the amount of reason in favour of it is, roughly, a matter of how different the world would need to be, normatively speaking, in order for it to be permitted. In what follows, I will offer one way of making this idea precise.

I begin with the intuitive idea that that all judgements about reason can be understood as judgements about *amounts of reason*. Thus, for all *p*, *q*, *w*, and *v*:

- To judge that there is *more reason* at *w* for it to be the case that *p* than there is at *v* for it to be the case that *q just is* to judge that the amount of reason that there is at *w* for it to be the case that *p* is *greater than* the amount of reason that there is at *v* for it to be the case that *q*;
- To judge that there is *equally as much reason* at *w* for it to be the case that *p* as there is at *v* for it to be the case that *q just is* to judge that the amount of reason that there is at *w* for it to be the case that *p* is *equal to* the amount of reason that there is at *v* for it to be the case that *q*;
- To judge that there is *at least as much reason* at *w* for it to be the case that *p* as there is at *v* for it to be the case that *q just is* to judge that the amount of reason that there is at *w* for it to be the case that *p* is *at least as great as* the amount of reason that there is at *v* for it to be the case that *q*;
- To judge that its being the case that *p* at *w* is *on a par* with its being the case that *q* at *v*, with respect to reason *just is* to judge that the amount of reason that there is at *w* for it to be the case that *p* is *on a par with* the amount of reason that there is at *v* for it to be the case that *q*;
- To judge that there is *sufficient reason*, at *w*, for it to be the case that *p just is* to judge that the amount of reason that there is at *w* for it to be the case that *p* is at least as great as the relevant *threshold* amount of reason;
- To judge that there is *insufficient reason* at *w* for *p* to be the case *just is* to judge that the relevant *threshold* amount of reason is greater than the amount of reason that there is at *w* for it to be the case that *p*;
- To judge that there is *conclusive reason* at *w* for it to be the case that *p just is* to judge that the amount of reason that there is at *w* for it to be the case that *p* is at least as great as the relevant *threshold* amount of reason, and the relevant *threshold* amount of reason is greater than the amount of reason that there is at *w* for it to be the case that ¬*p*;
- To judge that there is *some reason*, at *w*, for it to be the case that *p just is* to judge that the amount of reason that there is at *w* for it to be the case that *p* is greater than the *minimum* amount of reason; and

- To judge that there is *no reason*, at *w*, for it to be the case that *p just is* to judge that the amount of reason that there is at *w* for it to be the case that *p* is equal to the *minimum* amount of reason.

6.2.1 Preliminaries

We start with a closeness relation between possible worlds. I will assume that this closeness relation is a similarity relation of a kind that could feature in a Lewis-Stalnaker analysis of counterfactuals (Lewis, 1973; Stalnaker, 1968). This similarity relation is the same similarity relation that featured in the analysis of OUGHT in chapter one. As I argued there, this relation is best thought of as one that prioritises—or perhaps even consists in—similarity with respect to relevant *normative* facts, such as facts about what *may* and *must* be so, facts about the degree to which it is *fitting* to hold various attitudes, and facts about what *explains* these normative facts. Intuitively, then, how close two worlds are to each other in this sense is (primarily) a matter of how alike they are normatively speaking.

We begin, then, with a set of possible worlds, W, and a partially ordered set $(D, \geq_1)$, where D is the set of all possible *degrees of similarity* between worlds, such that, for all x and $y \in D$, $x \geq_1 y$ just in case, for any worlds w, u, v, $z \in W$, if w is similar to u to degree x, and v is similar to z to degree y, then w is at least as similar to u as v is similar to z. We can then define a function d: $W \times W \rightarrow D$ such that, for any w and $v \in W$, $d(w, v)$ is the degree to which w is similar to v.

We will also want to be able to compare *sets* of degrees of similarity, based on which set contains greater degrees of similarity. So we define a quasi order of $\wp(D)$, $\geq_2$, such that, for all x and $y \in \wp(D)$, $x \geq_2 y$ just in case there is some $a \in x$ such that, for all $b \in y$ such that $b \geq_1 a$, there is some $c \in x$ such that $c \geq_1 b$. (If x and y are finite, this amounts to the condition that there is some $a \in x$ for which there is no $b \in y$ such that $b \geq_1 a$ and $\neg(a \geq_1 b)$.)

6.2.2 Amount of Reason

I provide two possible analyses of REASON (mass): a simple analysis, which is a natural companion to the simple counterfactual analysis of OUGHT in chapter one, and a strengthened analysis, which is a natural companion to the strengthened

counterfactual analysis of OUGHT in chapter one. My task here is to define a function that will play the role of assigning to each proposition an 'amount of reason' at each possible world. For the simple analysis, this function will take an ordered pair consisting of a proposition *p* and a world *w*, and assign to it an ordered pair, the first element of which represents whether or not at *w* it may be the case that *p*, and the second element of which represents how far from *w* all the possible worlds are at which the opposite is true.

Thus, for the simple analysis, we define a function r: $\wp(W) \times W \rightarrow \{0, 1\} \times \wp(D)$ such that, for all $(p, w) \in \wp(W) \times W$:

- If, at *w*, it may be that *p*, then $r(p, w) = (1, x)$, where:
  - $x = \{y \in D$ | there is some $v \in W$ such that:
    - At *u*, it is not the case that it may be that *p*; and
    - $y = d(v, w)\}$; and
- If, at *w*, it is not the case that it may be that *p*, then $r(p, w) = (0, x)$, where:
  - $x = \{y \in D$ | there is some $v \in W$ such that:
    - At *u*, it may be that *p*;
    - and $y = d(v, w)\}$.

Recall that the strengthened counterfactual analysis of OUGHT in chapter one was designed only to take 'fully determinate' worlds into account—that is, worlds at which nothing is *optional*. If we accept this strengthened analysis of OUGHT, then, in order to validate the schema presented above that connects judgements involving the concept REASON (mass) to judgements involving the concept OUGHT, we should also modify our definition of *r* so that it only takes these kinds of worlds into account. Thus, for the strengthened analysis, we define *r* such that, for all *p* and *w*:

- If, at *w*, it may be that *p*, then $r(p, w) = (1, x)$, where:
  - $x = \{y \in D$ | there is some $v \in W$ such that:
    - At *v*:
      - It is not the case that it may be that *p*; and
      - *For all r, it is not the case that it may be that r or it is not the case that it may not be that* $\neg r$; and
    - $y = d(v, w)\}$; and

- If, at w, it is not the case that it may be that p, then $r(p, w) = (0, x)$, where:
  - $x = \{y \in D \mid$ there is some $v \in W$ such that:
    - At v:
      - It may be that p; and
      - *For all r, it is not the case that it may be that r or it is not the case that it may not be that* $\neg r$;
    - and $y = d(v, w)\}$.

We can thus think of r as a function that assigns to each proposition an *amount of reason* at each world, where an amount of reason is, roughly, a measure of both the deontic status of a proposition at a world and the distances from that world of relevant worlds at which that proposition has a different deontic status.

6.2.3 At Least as Great

In order to be able to compare amounts of reason, we must define an ordering on the outputs of r. So let us define a quasi order of $\{0, 1\} \times \wp(D)$, $\geq_3$, such that, for any two amounts of reason, (x, y) and $(z, w) \in \{0, 1\} \times \wp(D)$, $(x, y) \geq_3 (z, w)$ just in case: either $x = 1$ and $z = 0$; or $x = z = 1$ and $w \geq_2 y$; or $x = z = 0$ and $y \geq_2 w$. With this quasi order in hand, we are in a position to start making sense of judgements about amounts of reason as judgements about r. Thus we can say that, for all p, q, w, and v, to judge that the amount of reason that there is at w for it to be the case that p is *at least as great as* the amount of reason that there is at v for it to be the case that q *just is* to judge that $r(p, w) \geq_3 r(q, v)$—and $r(p, w) \geq_3 r(q, v)$ just in case one of the following conditions holds:

1. p is permitted at w and q is not permitted at v; or
2. p is permitted at w and q is permitted at v, and it is *not* the case that q is more *robustly* permitted at v than p is at w: that is, roughly, it is not the case that the closest worlds to w at which p is not permitted are closer to w than the closest worlds to v at which q is not permitted are to v; or
3. p is not permitted at w and q is not permitted at v, and it is *not* the case that p is more *robustly* not permitted at v than p is at w: that is, roughly, it is not the case that the closest worlds to v at which q is permitted are closer to v than the closest worlds to w at which p is permitted are to w.

The fact that $\geq_3$ is, by definition, a quasi order makes sense of the fact that AT LEAST AS MUCH REASON picks out a relation that has the properties of a quasi order.

6.2.4 Greater Than

In order to make sense of judgements about amounts of reason that are *greater than* others, let us define a strict partial order of $\{0, 1\} \times \wp(D)$, $>_3$, that corresponds to $\geq_3$, such that, for any two amounts of reason, (x, y) and $(z, w) \in \{0, 1\} \times \wp(D)$, $(x, y) >_3 (z, w)$ just in case $(x, y) \geq_3 (z, w)$ and it is not the case that $(z, w) \geq_3 (x, y)$. Having defined this strict partial order, we can say that, for all p, q, w, and v, to judge that the amount of reason that there is at w for it to be the case that p is *greater than* the amount of reason that there is at v for it to be the case that q *just is* to judge that $r(p, w) >_3 r(q, v)$—and $r(p, w) >_3 r(q, v)$ just in case one of the following conditions holds:

1. p is permitted at w and q is not permitted at v;
2. p is permitted at w and q is permitted at v, but p is more *robustly* permitted at w than q is at v: that is, roughly, the closest worlds to v at which q is not permitted are closer to v than the closest worlds to w at which p is not permitted are to w; or
3. p is not permitted at w and q is not permitted at v, but q is more *robustly* not permitted at v than p is at w: that is, roughly, the closest worlds to w at which p is not permitted are closer to w than the closest worlds to v at which q is not permitted are to v.

The fact that $>_3$ is, by definition, a strict partial order makes sense of the fact that it is *a priori* that MORE REASON picks out a property that has the properties of a strict partial order.

6.2.5 Equal

In order to make sense of amounts of reason that are *equal*, let us define an equivalence relation on $\{0, 1\} \times \wp(D)$, $=_3$, that corresponds to $\geq_3$, such that, for any two amounts of reason, (x, y) and $(z, w) \in \{0, 1\} \times \wp(D)$, $(x, y) =_3 (z, w)$ just in case $(x, y) \geq_3 (z, w)$ and $(z, w) \geq_3 (x, y)$. Thus, for all p, q, w, and v, we can say that, to judge that the amount of reason that there is at w for it to be the case that p is *equal to* the amount of reason that there is at v for it to be the case that q *just is* to judge that $r(p,$

w) $=_3$ $r(q, v)$. The fact that $=_3$ is, by definition, an equivalence relation makes sense of the fact that it is *a priori* that EQUALLY AS MUCH REASON picks out a relation that has the properties of an equivalence relation.

6.2.6 On a Par

In order to make sense of amounts of reason that are *on a par*, let us define a reflexive and symmetric relation on $\{0, 1\} \times \wp(D)$, $\approx_3$, such that, for any two amounts of reason, (x, y) and $(z, w) \in \{0, 1\} \times \wp(D)$, $(x, y) \approx_3 (z, w)$, if, and only if $x = z$. Having defined this relation, we can say that, for all p, q, w, and v, to judge that the amount of reason that there is at w for it to be the case that p is *on a par with* the amount of reason that there is at v for it to be the case that q *just is* to judge that $r(p, w) \approx_3 r(q, v)$.

Given this definition of $\approx_3$, it follows that, for all p, q, w, and v, the amount of reason that there is at w for it to be the case that p is on a par with the amount of reason that there is at v for it to be the case that q just in case either p is permitted at w and q is permitted at v, or p is not permitted at w and q is not permitted at v. This is certainly one kind of 'rough equality' with respect to amounts of reason, but one might worry that this conception of parity is *too* coarse-grained. If this is correct, then this part of the analysis will need to be revised.

As it stands, however, the fact that $\approx_3$ is, by definition, a reflexive and symmetric relation makes sense of the fact that it is *a priori* that ON A PAR WITH RESPECT TO AMOUNT OF REASON picks out a reflexive and symmetric relation. It also follows from this definition that, for any two amounts of reason, (x, y) and $(z, w) \in \{0, 1\} \times \wp(D)$, if $(x, y) =_3 (z, w)$ then $(x, y) \approx_3 (z, w)$. This vindicates the fact that EQUALLY AS MUCH REASON implies ON A PAR WITH RESPECT TO AMOUNT OF REASON.

6.2.7 Threshold Amount

$(1, D)$ is an amount of reason that marks the threshold between being permitted and not being permitted. That is, as I will show below, for all p and w, $r(p, w) \geq_3 (1, D)$, if, and only if at w it may be that p. It is thus natural to analyse SUFFICIENT REASON and INSUFFICIENT REASON in terms of this threshold amount. Thus we may say that, for all p and w, to judge that, at w, there is sufficient reason for it to be the case that p *just is*

to judge that $r(p, w) \geq_3 (1, D)$; to judge that there is insufficient reason for it to be the case that *p just is* to judge that $(1, D) >_3 r(p, w)$; and to judge that there is conclusive reason for it to be the case that *p just is* to judge that $r(p, w) \geq_3 (1, D)$ and $(1, D) >_3 r(\neg p, w)$. The fact that, by definition, for all *p* and *w*, if $r(p, w) \geq_3 (1, D)$ and $(1, D) >_3 r(\neg p, w)$, then $r(p, w) >_3 r(\neg p, w)$ makes sense of the fact that CONCLUSIVE REASON implies MORE REASON TO … THAN NOT TO.

6.2.8 Minimum Amount

There is no $(x, y) \in \{0, 1\} \times \wp(D)$ such that $(0, \varnothing) >_3 (x, y)$. Thus, we can think of $(0, \varnothing)$ as the minimum amount of reason. It follows that, for all *p* and *w*, to judge that there is no reason at *w* for *p* to be the case *just is* to judge that $r(p, w) =_3 (0, \varnothing)$; and to judge that there is some reason at *w* for *p* to be the case *just is* to judge that $r(p, w) >_3 (0, \varnothing)$.

One possible objection to this analysis is that it implies that it is almost never the case that there is no reason for something to be so, which is implausible. For all *p* and *w*, $r(p, w) =_3 (0, \varnothing)$ just in case there is no world in the set of worlds *W* at which it may be that *p* (or there are some such worlds, but for some reason they are not assigned a degree of similarity to *w*). One might worry that this makes cases in which there is no reason for something to be so too rare, since, for almost any proposition, there is some—perhaps extremely distant—world at which it may be that that proposition holds.

There are two main responses that the proponent of the closeness analysis could make here. The first is to point out that *W* need not contain every possible world. It is not implausible that people would ignore worlds that are sufficiently dissimilar to the actual world when making judgements about how much reason there is to do things. The second response is to draw on Schroeder's (2007, pp. 92-97) observation that intuitions to the effect that there is no reason for something to be so are notoriously unreliable, and that it is in fact very rare that there is literally no reason for something to be so. For example, consider any action that one might think there is no reason for me to perform—say, blinking twenty times in the next five seconds. There is an extremely miniscule chance that my blinking 20 times in the next five seconds would somehow bring about the end of world poverty. This is of course absurd, but not

literally impossible. This seems to provide me with an extremely tiny amount of reason to blink 20 times in the next five seconds—though of course this amount of reason is nowhere near sufficient to make it even worth considering. Thus, in light of the fact that cases in which there is literally no reason for something to be so are in fact very rare, this objection to the closeness analysis can be resisted.

6.2.9 Relation to May, Must, and Ought

This analysis of REASON (mass) validates the rational principles that govern how judgements about reason relate to judgements involving the concepts MAY, MUST, and (given the counterfactual analysis from chapter one) OUGHT.

6.2.10 May and Sufficient Reason

SUFFICIENT REASON implies MAY.

$$\frac{r(p, w) \geq_3 (1, D).}{\text{At } w\text{, it may be that } p.}$$

Assume, for some arbitrary p and w, that $r(p, w) \geq_3 (1, D)$. Then there is some A such that $r(p, w) = (1, A)$. Given the definition of r, it follows that, at w, it may be that p.

MAY implies SUFFICIENT REASON.

$$\frac{\text{At } w\text{, it may be that } p.}{r(p, w) \geq_3 (1, D).}$$

Assume, for some arbitrary p and w, that, at w, it may be that p. It follows that there is some $A \subseteq D$ such that $r(p, w) = (1, A)$. Thus, given the definition of $\geq_3$, $r(p, w) \geq_3 (1, D)$ just in case $D \geq_2 A$. Assume that it is not the case that $D \geq_2 A$. Then, for all $a \in D$, there is some $b \in A$ such that $b \geq_1 a$ and there is no $c \in D$ such that $c \geq_1 b$. But this is impossible, since $A \subseteq D$. So $D \geq_2 A$ and $r(p, w) \geq_3 (1, D)$.

INSUFFICIENT REASON implies ¬MAY.

$$\frac{(1, D) >_3 r(p, w).}{\text{At } w\text{, it is not the case that it may be that } p.}$$

Assume, for some arbitrary p and w, that $(1, D) >_3 r(p, w)$. It follows that there is some x and A such that $r(p, w) = (x, A)$ and either (i) $x = 0$ or (ii) $\neg(D \geq_2 A)$. If (i) is true, then, given the definition of r, at w, it is not the case that it may be that p. Assume that (ii) is true. Then, for all $a \in D$, there is some $b \in A$ such that $b \geq_1 a$ and there is no $c \in D$ such that $c \geq_1 b$. But this is impossible, since $A \subseteq D$. Thus (ii) is false. Since (i) is true and (ii) is false, then, it is true that, at w, it is not the case that it may be that p.

¬MAY implies INSUFFICIENT REASON.

At w, it is not the case that it may be that p.

$(1, D) >_3 r(p, w)$.

Assume, for some arbitrary p and w, that, at w, it is not the case that it may be that p. Then, given the definition of r, there is some A such that $r(p, w) = (0, A)$, and so, since $(1, D) >_3 (0, A)$, it follows that $(1, D) = r(p, w)$.

6.2.11 Ought and More Reason

This analysis, when combined with the counterfactual analysis of OUGHT from chapter one, and a Lewis-Stalnaker-style closeness analysis of counterfactuals (Lewis, 1973; Stalnaker, 1968), vindicates the schemata identified above that connect judgements involving the concept REASON (mass) to judgements involving the concept OUGHT. Given their simple analyses, OUGHT implies MORE REASON THAN NOT.

At w, if it were not the case that it may be that p or not the case that it may not be that p, then it would be the case that it must be that p.

$r(p, w) >_3 r(\neg p, w)$

Assume, for some arbitrary p and w, that, at w, if it were not the case that it may be that p or not the case that it may not be that p, then it would be the case that it must be that p. As shown in chapter one, this entails that it may be that p. Either (i) at w, it is not the case that it may be that $\neg p$, or (ii) at w, it is the case that it may be that $\neg p$. If (i) is true then there is some A and B such that $r(p, w) = (1, A)$ and $r(\neg p, w) = (0, B)$, and so $r(p, w) >_3 r(\neg p, w)$. If (ii) is true, then there is some A and B such that $r(p, w) = (1, A)$ and $r(\neg p, w) = (1, B)$. Given the counterfactual whose truth we assumed above,

at all of the closest worlds to *w* at which it is not the case that it may be that *p* or not the case that it may not be that *p*, it must be that *p*, and so it may be that *p* and it is not the case that it may not be that *p*. It follows that, for every world at which it is not the case that it may be that *p*, there is a closer world at which it is not the case that it may not be that *p*, and so $\neg(A \geq_2 B)$. Moreover, we may assume that there is some world in *W* at which it is not the case that it may be that *p* or it is not the case that it may not be that *p*, since it is plausibly an *a priori* truth that nothing is optional as a matter of necessity. It follows that there is some world at which it is not the case that it may not be that *p* for which there is no closer world at which it is not the case that it may be that *p*, and so $B \geq_2 A$. Thus, since $B \geq_2 A$ and $\neg(A \geq_2 B)$, it follows that $r(p, w) = (1, A) >_3 (1, B) = r(\neg p, w)$.

The same implication holds given the strengthened analyses of these concepts.

At *w*, if it were the case that, for all *q*, it is not the case that it may be that *q*, or it is not the case that it may not be that *q*, then it would be the case that it must be that *p*.

$r(p, w) >_3 r(\neg p, w)$

Assume, for some arbitrary *p* and *w*, that, at *w*, the relevant counterfactual holds. As shown in chapter one, this entails that it may be that *p*. Either (i) at *w*, it is not the case that it may be that $\neg p$, or (ii) at *w*, it is the case that it may be that $\neg p$. If (i) is true then there is some *A* and *B* such that $r(p, w) = (1, A)$ and $r(\neg p, w) = (0, B)$, and so $r(p, w) >_3 r(\neg p, w)$. If (ii) is true, then there is some *A* and *B* such that $r(p, w) = (1, A)$ and $r(\neg p, w) = (1, B)$. Given the counterfactual whose truth we assumed above, at all of the closest normatively determinate worlds to *w*, it must be that *p*, and so it may be that *p* and it is not the case that it may not be that *p*. It follows that, for every normatively determinate world at which it is not the case that it may be that *p*, there is a closer normatively determinate world at which it is not the case that it may not be that *p*, and so $\neg(A \geq_2 B)$. Moreover, we made the assumption in chapter one that there is some normatively determinate world in *W*. It follows that there is some normatively determinate world at which it is not the case that it may not be that *p* for which there is no closer normatively determinate world at which it is not the case that it may be that *p*, and so $B \geq_2 A$. Thus, since $B \geq_2 A$ and $\neg(A \geq_2 B)$, it follows that $r(p, w) = (1, A) >_3 (1, B) = r(\neg p, w)$.

Given the simple analyses of REASON (mass) and OUGHT, SUFFICIENT REASON and MORE REASON imply OUGHT.

$$r(p, w) \geq_3 (1, D); r(p, w) >_3 r(\neg p, w).$$

At w, if it were not the case that it may be that p or not the case that it may not be that p, then it would be the case that it must be that p..

Assume that the premise holds, for some arbitrary p and w. Since $r(p, w) \geq_3 (1, D)$, there is some A such that $r(p, w) = (1, A)$, and so, given the definition of r, at w, it may be that p. Suppose that $r(\neg p, w) = (x, B)$, for some arbitrary x and B. Either (i) $x = 0$ or (ii) $x = 1$. If (i) is true, then, at w, it is not the case that it may be that $\neg p$ and so, since, at w, it may be that p, it must be that p, and thus the counterfactual in the conclusion is trivially true. If (ii) is true, then, given that $r(p, w) >_3 r(\neg p, w)$, it is not the case that $A \geq_2 B$. It follows that, for every closest world to w at which it is not the case that it may be that p, there is a closer world to w at which it is not the case that it may not be that p, and thus also the case that it must be that p. Thus, the closest worlds to w at which it is not the case that it may be that p or it is not the case that it may not be that p are worlds at which it must be that p. It follows that, at w, if it were the case that, for all q, it must be that q, or not the case that it may be that q, then it would be the case that it must be that p.

The same implication holds given the strengthened analyses of these concepts.

$$r(p, w) >_3 r(\neg p, w); r(p, w) \geq_3 (1, D).$$

If it were the case that, for all q, it is not the case that it may be that q, or it is not the case that it may not be that q, then it would be the case that it must be that p.

Assume that the premise holds, for some arbitrary p and w. Since $r(p, w) \geq_3 (1, D)$, there is some A such that $r(p, w) = (1, A)$. So, at w, it may be that p. Given that $r(p, w) >_3 r(\neg p, w)$, there is some x and B such that $r(\neg p, w) = (x, B)$. Either (i) $x = 0$ or (ii) $x = 1$. If (i) is true, then, at w, it is not the case that it may be that $\neg p$ and so, since, at w, it may be that p, it must be that p, and so, as shown in chapter one, the counterfactual in the conclusion is true. If (ii) is true, then it is not the case that $A \geq_2 B$. So, for every normatively determinate world at which it is not the case that it may be that p, there is a normatively determinate world that is at least as close to w at which it is not the case that it may not be that p, and there is no normatively determinate world that is at

least as close at which it is not the case that it may be that *p*. It follows that, for every normatively determinate world, there is a normatively determinate world that is at least at close at which it is the case that it may be that *p* and not the case that it may not be that *p*, and there is no closer normatively determinate world and this is not the case. It follows that the counterfactual is true.

6.2.12 Ownership of Reason

I have presented the closeness analysis of REASON (mass) as an account of judgements about the amount reason *there is* for something to be so. Given that, according to this analysis, REASON (mass) is analysed in terms of MAY, there is a straightforward way to extend the analysis to cover judgements about the amount of reason someone *has* to do something. Judgements about the amount of reason *there is* are judgements about the relative closeness of worlds at which certain things may be so, where the relevant permission may be thought of as *unowned*. Judgements about the amount of reason someone *has*, by contrast, can be understood as judgements about the relative closeness of worlds at which certain things may be so, where the relevant permission is taken to be *owned* by that agent. The analysis thus fully satisfies our second desideratum.

6.2.13 Summary

In this section, I presented the closeness analysis of REASON (mass) and showed that it meets both of our desiderata: it makes sense of the idea of relative amounts of reason, and it validates the connections between REASON (mass) and the concepts MAY, MUST, and OUGHT (in both their owned and unowned guises). One weakness of the analysis is that it does not make sense of the fact that it is controversial that MORE REASON and AT LEAST AS MUCH REASON pick out transitive relations: according to the analysis I have provided, these concepts straightforwardly pick out transitive relations. I will leave the substantial task of deciding whether and how to modify the analysis to accommodate Temkin's (2012) counterexamples to transitivity to another time.

6.3 Desiderata for an Analysis of REASON (Count)

In the previous section, the focus was on judgements about amounts of reason, where 'reason' is a mass noun. The focus in this section will be on judgements about normative reasons, where 'reason' is a count noun. Judgements about normative reasons are judgements about the degree to which considerations or facts count in favour of responses. In the next section, I will provide a closeness account of these judgements. Before I do that, however, I will identify several desiderata for an account of this kind.

6.3.1 Desideratum 1: Makes Sense of the Idea of Relative Strengths of Reasons

Normative reasons come in different strengths or weights. The fact that I would crash my car is a reason for me not to drive blindfolded. The fact that I do not like the song that is playing on the radio is a reason for me to change the station. The first of these reasons is much stronger than the second. In the rest of this section, I will set out a number of rational principles that apply to judgements about reasons and their strengths. Validating these principles is our first desideratum for an analysis of REASON (count). At the end of this section, I will consider some accounts of reasons in the literature that fail to account adequately for the idea that reasons can differ in strength.

6.3.1.1 Stronger Reason

First, STRONGER REASON picks out a relation that has the properties of a strict partial order—transitivity and asymmetry.

Transitivity

The fact that *p* is a stronger reason for it to be the case that *s* than the fact that *q* is for it to be the case that *t*; The fact that *q* is a stronger reason for it to be the case that *t* than the fact that *r* is for it to be the case that *u*.

The fact that *p* is a stronger reason for it to be the case that *s* than the fact that *r* is for it to be the case that *u*.

Asymmetry

The fact that *p* is a stronger reason for it to be the case that *s* than the fact that *q* is for it to be the case that *t*.

It is not the case that the fact that *q* is a stronger reason for it to be the case that *t* than the fact that *p* is for it to be the case that *s*.

6.3.1.2 Equally Strong Reason

Next, EQUALLY STRONG REASON picks out a relation that has the properties of an equivalence relation—reflexivity, symmetry, and transitivity.

Reflexivity

-

The fact that *p* is an equally strong reason for it to be the case that *s* as the fact that *p* is for it to be the case that *s*.

Symmetry

The fact that *p* is an equally strong reason for it to be the case that *s* as the fact that *q* is for it to be the case that *t*.

The fact that *q* is an equally strong reason for it to be the case that *t* as the fact that *p* is for it to be the case that *s*.

Transitivity

The fact that *p* is an equally strong reason for it to be the case that *s* as the fact that *q* is for it to be the case that *t*; The fact that *q* is an equally strong reason for it to be the case that *t* as the fact that *r* is for it to be the case that *u*.

The fact that *p* is an equally strong reason for it to be the case that *s* as the fact that *r* is for it to be the case that *u*.

6.3.1.3 At Least as Strong

The concept of AT LEAST AS STRONG A REASON *just is* the concept of STRONGER OR EQUALLY AS STRONG A REASON. AT LEAST AS STRONG A REASON thus picks out a relation that has the properties of a quasi order—reflexivity and transitivity.

Reflexivity

-

The fact that *p* is at least as strong a reason for it to be the case that *s* as the fact that *p* is for it to be the case that *s*.

Transitivity

The fact that *p* is at least as strong a reason for it to be the case that *s* as the fact that *q* is for it to be the case that .; The fact that *q* is at least as strong a reason for it to be the case that *t* as the fact that *r* is for it to be the case that *u*.

The fact that *p* is at least as strong a reason for it to be the case that *s* as the fact that *r* is for it to be the case that *u*.

6.3.1.4 On a Par

Again, the parity relation is a reflexive and symmetrical relation.

Reflexivity

-

The strength of the fact that *p* as a reason for it to be the case that *s* is on a par with the strength of the fact that *p* as a reason for it to be the case that *s*.

Symmetry

The strength of the fact that *p* as a reason for it to be the case that *s* is on a par with the strength of the fact that *q* as a reason for it to be the case that .

The strength of the fact that *q* as a reason for it to be the case that *t* is on a par with the strength of the fact that *p* as a reason for it to be the case that *s*.

And, in the same way that EQUAL REASON implies ON A PAR WITH RESPECT TO AMOUNT OF REASON, as discussed above, EQUALLY STRONG REASON implies ON A PAR WITH RESPECT TO STRENGTH OF REASON.

The fact that *p* is an equally strong reason for it to be the case that *s* as the fact that *q* is for it to be the case that .

The strength of the fact that *p* as a reason for it to be the case that *s* is on a par with the strength of the fact that *q* as a reason for it to be the case that .

6.3.1.5 Conclusive, Sufficient, and *pro tanto* Reasons

Finally, CONCLUSIVE REASON is stronger than SUFFICIENT REASON, which is stronger than *PRO TANTO* REASON.

The fact that *p* is a conclusive reason for it to be the case that *s*.

The fact that *p* is a sufficient reason for it to be the case that *s*.

The fact that *p* is a sufficient reason for it to be the case that *s*.

The fact that *p* is a *pro tanto* reason for it to be the case that *s*.

6.3.1.6 Existing Accounts of Strength of Reasons in the Literature

Some accounts of reasons in the literature do a good job of accounting for the strength of reasons, while others do not. The accounts that do a good job do so by making sense of the strength of reasons in terms of some other gradable property. For example, there is a family of views in the literature according to which a reason is a piece of *evidence*—either evidence that something *ought* to be the case (Kearns and Star, 2008; 2009; 2013), or that it is *not* the case that something ought *not* to be the case (Lee, 2020), or that something is *fitting* (Thomson, 2008). In so far as some pieces of evidence are stronger than others, such accounts appear well placed to make sense of the strength of reasons. Other accounts in the literature make sense of reasons and their strength in terms of *goodness* (Finlay, 2006; 2014; Raz, 1999; Searle, 2001; Maguire, 2016; Gregory, 2016). For example, according to Finlay (2014), a reason is an explanation why something would be *good* (in some respect and to some degree), and, according to Gregory (2016), a reason is a *good basis* for some response. Since goodness is a gradable property, these accounts also appear well placed to make sense of the strength of reasons. According to Fogal (2016; Fogal and Risberg), a reason is a fact that explains why there is a certain amount of *reason* for something to be the case. In so far as reason comes in different amounts, this view also appears well placed to make sense of the strength of reasons. The account I will offer below will make use of another gradable property to make sense of the strength of reasons—the closeness or similarity of worlds.

Not all accounts of the strengths of reasons fare so well, however. According to one prominent account of reasons, reasons are premises in good reasoning (Setiya,

2007b; 2014; Hieronymi, 2011; 2021; Kauppinen, 2015; McHugh and Way, 2016; Way, 2017). Let us use Way's (2017) version of this view as an example here.

For all *p*, *S*, and φ, for the fact that *p* to be a reason for *S* to φ *just is* for there to be a good pattern of reasoning from the belief that *p*, perhaps together with other correct/fitting attitudes which *S* has, to φ-ing (Way, 2017; McHugh and Way, 2016).

One might think that these authors could also make use of the fact that goodness is a gradable property, and make sense of the strength of reasons in terms of better and worse patterns of reasoning. This would bring this account of reasons closer to Gregory's (2016) view. In fact, however, these authors do not try to make sense of the strength of reasons in terms of the gradability of goodness. Instead, Setiya and Kauppinen account for the strength of a fact as a reason for a response in terms of the amount of *motivation*, *inclination*, or *causal influence* that the belief that that fact obtains provides toward that response under the right conditions. Making sense of the strength of reasons in terms of some kind of concept of degrees of causal influence or inclination seems like a promising route to take for proponents of these accounts. Understanding the strength of reasons in terms of the concept of motivation, by contrast, does not seem promising, in so far as there can be reasons for certain attitudes—like belief—that are not naturally thought of as objects of motivation (Gregory, 2016).

Way (2017), by contrast, tries to account for what it is for one reason to *outweigh* another in terms of defeasible patterns of reasoning. Even if Way's account of outweighing is able to account for the comparative strengths of *competing* reasons, however, it fails to account for the strengths of reasons that are not in competition with each other. For example, the fact that I would crash my car is a much stronger reason for me not to drive blindfolded than the fact that I do not like the song that is now playing on the radio is for me to change the station. Since these two actions are unrelated to each other, however, neither of these reasons outweighs the other in Way's sense, and yet they nonetheless differ in strength. Way's account of the strength of reasons is thus incomplete (see Star, 2018, p. 255` for a similar objection).

Finally, Broome (2013) provides an account of reasons that is inspired by the plausible idea that reasons help to *explain* facts about what ought to be the case. The view that Broome eventually settles on is as follows.

For all *p*, *S*, and φ, for the fact that *p* to be a reason for *S* to φ *just is* for the fact that *p* play the 'for-φ' role in a 'weighing explanation' of why *S* ought to φ, or of why *S* ought not to φ, or of why it is neither the case that *S* ought to φ nor that *S* ought not to φ.

Although Broome identifies reasons with facts that have certain 'weights' in weighing explanations, he does not give an account of what it is for the weight of one fact to be greater than, equal to, or on a par with the weight of another fact in a weighing explanation; the concept of a fact's weight in a weighing explanation thus appears to be treated as primitive on Broome's account. One might think that the idea of a fact's having a weight in a weighing explanation of a normative fact is no more illuminating than the idea of a reason's having a particular weight. Broome's account is therefore not particularly informative when it comes to accounting for the weights of reasons.

6.3.2 Desideratum 2: Validates Connections between Reason (Count) and Reason (Mass)

Our second desideratum for an analysis of REASON (count) is that it validate the rational principles that connect judgements about *reasons* and judgements about *reason*. Judgements about amounts of *reason* and judgements about *reasons* are closely related in ways one would expect.

There is a conclusive reason for it to be the case that *s*.

There is conclusive reason for it to be the case that *s*.

There is a sufficient reason for it to be the case that *s*.

There is sufficient reason for it to be the case that *s*.

There is a *pro tanto* reason for it to be the case that *s*.

There is some reason for it to be the case that *s*.

It is not obvious that there are also schemata that operate in the other direction—for example, a schema that says that judging that there is conclusive reason to do something rationally commits one to judging that there is *a* conclusive reason to do it.

This is because there seem to be examples in which there is conclusive reason for some response, but no particular reason that counts in its favour. For example, it seems to me that there is conclusive reason for me to prefer pleasure to pain. But I cannot identify any particular reason that counts in favour of this preference. It seems rather to be simply a brute normative fact.

6.3.3 Desideratum 3: Makes Sense of the Connection between Reasons and Explanation

Our third desideratum for an analysis of REASON (count) is that it make sense of the fact that there seems to be an intimate connection between *reasons* and *explanations*. It is customary to distinguish *normative* reasons, of the kind we have been discussing here, from *explanatory* reasons and *motivating* reasons. Explanatory reasons explain why something is the case; it is plausible that to judge that the reason that the refrigerator stinks is that there are old beans in there *just is* to judge that the fact that there are old beans in there explains why the refrigerator stinks. Motivating reasons are often thought to be a species of explanatory reason that have to do with explaining agents' actions and mental states in a particular way. For example, it is natural to think that to judge that my reason for throwing out the beans in the fridge was that I wanted to get rid of the smell *just is* to judge that my wanting to get rid of the smell explains, in a particular way, why I threw out the beans.

Since both the concept of an explanatory reason and the concept of a motivating reason seem to be closely connected to the concept of *explanation*, it is natural to think that the concept of a normative reason might be too. Thus, a number of philosophers have put forward accounts of reasons according to which normative reasons are (or are parts of) explanations of some sort—either explanations why something *ought* to be so (Broome, 2013), why it would be *good* if something were so (Finlay, 2006; 2014; Raz, 1999; Searle, 2001; Maguire, 2016), why some response would be *fitting* (Chappell, 2012; Howard, 2019), or why there is a certain amount of *reason* for something to be so (Fogal, 2016; Fogal and Risberg).

In so far as Gregory's (2016) account of reasons as good bases, and McHugh and Way's (2016; Way, 2017) and Setiya's (2007b; 2014) accounts of reasons as premises in good reasoning, can be understood as making sense of normative

reasons in terms of motivating reasons, these accounts arguably also secure a connection to explanation, to the extent that motivating reasons are themselves explanatory. Accounts of reasons according to which reasons are pieces of evidence (Kearns and Star, 2008; 2009; 2013; Lee, 2020; Thomson, 2008), by contrast, appear not to satisfy this desideratum. Since it does seem plausible that the concept of a normative reason is somehow connected to the concept of explanation, I will treat it as a desideratum for an analysis of this concept that it make sense of this connection.

6.3.4 Desideratum 4: Does not, by itself, Entail the Nonexistence of either the 'Right Kind' or 'Wrong Kind' of Reasons for Attitudes

Our fourth desideratum for an analysis of the concept REASON (count) is that it remain neutral on the existence of either the 'right kind' or the 'wrong kind' of reasons for attitudes. The 'right kind' of reasons for attitudes—sometimes called 'object-given' reasons (Parfit, 2001; 2011, Appendix A; cf. Rabinowicz and Rønnow-Rasmussen, 2004; Schroeder, 2012)—are, roughly, considerations that bear on the question of whether an attitude is *appropriate* or *correct*. The fact that the weather report says that it will rain tomorrow is thus a reason of the right kind for me to believe that it will rain tomorrow, since it is a consideration that bears on the question of whether it is appropriate or correct for me to believe that it will rain tomorrow.

The 'wrong kind' of reasons for attitudes—sometimes called 'state-given' reasons (Parfit, 2001; Parfit, 2011, Appendix A; cf. Rabinowicz and Rønnow-Rasmussen, 2004; Schroeder, 2012)—are, roughly, considerations that bear on the question of whether an attitude is *desirable* to hold. The fact that I have been offered an enormous financial reward to believe that it will rain tomorrow is thus a reason of the wrong kind for me to believe that it will rain tomorrow, since it is a consideration that bears on the question of whether it is desirable for me to hold that belief, rather than on the question of whether the belief is appropriate or correct.

Some argue that there are no reasons of the wrong kind (Gibbard, 1990, p. 37; Parfit, 2011, p. Appendix A; Skorupski, 2007; Way, 2012). According to this view, reasons of the wrong kind are not reasons to hold the relevant attitude, but rather, are reasons to *desire* or *bring it about* that one holds the relevant attitude. This view is

controversial. Thus, if possible, it would be desirable to find an analysis of the concept REASON (count) that remains neutral on whether there are any reasons of the right or wrong kind (Howard, 2019).

Some existing accounts of reasons fail to meet this desideratum. For example, there are accounts in the literature according to which a reason for an attitude *just is* a consideration that explains why (Chappell, 2012) or is evidence that (Thomson, 2008) it is *fitting* to hold that attitude. While views of this kind may have some plausibility as accounts of reasons of the right kind, it is difficult to see how they could accommodate the existence of reasons of the wrong kind. There are other views in the literature according to which reasons for an attitude are considerations that explain why holding that attitude would be *good* in some way (Finlay, 2006; 2014; Raz, 1999; Searle, 2001; Maguire, 2016). These accounts threaten to rule out the existence of reasons of the right kind (though see Finlay, 2014 for a response to this worry).

6.3.5 Desideratum 5: Does not Give Rise to Counterexamples

Our fifth desideratum is simply that the analysis not give rise to counterexamples. A number of putative counterexamples have been proposed, for example, against Kearns and Star's (2006; 2008) view that reasons for someone to do something are pieces of evidence that they ought to do it. Kearns and Star have been largely successful in defusing these counterexamples (see, for example, Kearns and Star, 2008; Kearns and Star, 2013). There is one type of counterexample to their view that I myself find persuasive, however: on this view, certain facts *about* reasons, that intuitively are not themselves reasons, are counted as reasons. For example, the fact that there is conclusive reason for someone to do something, which is evidence that they ought to do it, but not itself a reason for them to do it, constitutes a reason for them to do it, on this view (Brunero, 2009; 2018).

6.4 The Closeness Analysis of REASON (Count)

In this section, I provide an analysis of REASON (count) that is a natural companion to the closeness analysis of REASON (mass) set out above. According to this analysis, judgements about reasons and their strengths are all judgements about the *amount*

of reason that is *provided by* some fact (Fogal, 2016), and the amount of reason for something that is provided by a fact is determined by the relative closeness of worlds at which that fact *explains why* it is *permitted*. The resulting view is one according to which how strong a reason is for something to be so is a matter of how close it comes to explaining why it is permitted, and how far it is from not explaining why it is permitted. On this view, if a fact is sufficient to explain why something is permitted, then it is a reason for it to be so. How strong a reason it is depends on how *robust* an explanation it is—that is, it depends on how different things would need to be in order for it *not* to explain why it is permitted. If, on the other hand, a fact is not sufficient to explain why something is permitted, then it is a reason for it to be so if it *could have*, under the right circumstances, explained why it is permitted. In that case, how strong a reason it is depends on how *close* it comes to explaining why it is permitted—that is, it depends on how different things would need to be in order for it to explain why it is permitted. In what follows, I will offer one way of making this idea precise.

I begin with the intuitive idea that that all judgements about reasons can be understood as judgements about *amounts of reason* that are *provided* by facts (Fogal, 2016). Thus, for all *p*, *q*, *r*, *t, w*, and *v*:

- To judge that the fact that *p* is a *stronger reason* at *w* for it to be the case that *q* than the fact that *r* is at *v* for it to be the case that *t just is* to judge that the amount of reason for it to be the case that *q* that is provided by the fact that *p* at *w* is *greater than* the amount of reason for it to be the case that *t* that is provided by the fact that *r* at *v*;
- To judge that the fact that *p* is an *equally strong reason* at *w* for it to be the case that *q* as the fact that *r* is at *v* for it to be the case that *t just is* to judge that the amount of reason for it to be the case that *q* that is provided by the fact that *p* at *w* is *equal to* the amount of reason for it to be the case that *t* that is provided by the fact that *r* at *v*;
- To judge that the fact that *p* is *at least as strong a reason* at *w* for it to be the case that *q* as the fact that *r* is at *v* for it to be the case that *t just is* to judge that the amount of reason for it to be the case that *q* that is provided by the fact that *p* at *w* is *at least as great as* the amount of reason for it to be the case that *t* that is provided by the fact that *r* at *v*;

- To judge that the strength of the fact that *p* as a reason at *w* for it to be the case that *q* is *on a par with* the strength of the fact that *r* as a reason at *v* for it to be the case that *t just is* to judge that the amount of reason for it to be the case that *q* that is provided by the fact that *p* at *w* is *on a par with* the amount of reason for it to be the case that *t* that is provided by the fact that *r* at *v*;
- To judge that the fact that *p* is a *sufficient reason* at *w* for it to be the case that *q just is* to judge that the amount of reason for it to be the case that *q* that is provided by the fact that *p* is at least as great as the relevant *threshold* amount of reason;
- To judge that the fact that *p* is a *conclusive reason* at *w* for it to be the case that *q just is* to judge that the amount of reason for it to be the case that *q* that is provided by the fact that *p* at *w* is at least as great as the relevant *threshold* amount of reason and there is insufficient reason at *w* for it to be the case that ¬*p*; and
- To judge that the fact that *p* is a *pro tanto reason* at *w* for it to be the case that *q just is* to judge that the amount of reason for it to be the case that *q* that is provided by the fact that *p* is greater than the *minimum* amount of reason.

6.4.1 Amount of Reason Provided by a Fact

To make the idea of the amount of reason provided by a fact precise, I will draw on the account of amounts of reason presented above. Again, I provide two different analyses: a simple analysis, which is a natural counterpart to the simple analyses of REASON (mass) and OUGHT presented so far, and a strengthened analysis, which is a natural counterpart to the strengthened analyses of REASON (mass) and OUGHT presented so far. My task here is to define a function that will play the role of assigning to each fact and proposition at each world an amount of reason that is provided by that fact in favour of that proposition at that world. We thus define a function s: $\wp(W) \times \wp(W) \times W \rightarrow \{0, 1\} \times \wp(D)$ such that, for all p, q, and w:

- If, at w, it is true that p, and the fact that p is sufficient to explain why it may be that q, then $s(p, q, w) = (1, x)$, where:
  - $x = \{y \in D \mid$ there is some $v \in W$ such that:
    - At v, it is not the case that the fact that p is sufficient to explain why it may be that q; and

        - $y = d(v, w)$}; and
- If, at w, it is true that p, and the fact that p is not sufficient to explain why it may be that q, then $s(p, q, w) = (0, x)$, where:
    - $x = \{y \in D$ | there is some $v \in W$ such that:
        - At v, the fact that p is sufficient to explain why it may be that q; and
        - $y = d(v, w)$}; and
- If, at w, it is not true that p, then $s(p, q, w) = (0, \varnothing)$.

s is thus a function that assigns to each triple consisting of two propositions and a world an amount of reason that can be thought of as the amount of reason for the second proposition to be true that is provided by the fact that the first proposition is true at that world. For the strengthened analysis, as above, we modify the definition so as only to include 'fully determinate' worlds. Thus, we define t such that, for all p, q, and w:

- If, at w, it is true that p, and the fact that p is sufficient to explain why it may be that q, then $s(p, q, w) = (1, x)$, where:
    - $x = \{y \in D$ | there is some $v \in W$ such that:
        - At v:
            - It is not the case that the fact that p is sufficient to explain why it may be that q; and
            - For all r, either it must be that r or it is not the case that it may be that $\neg r$; and
        - $y = d(v, w)$}; and
- If, at w, it is true that p, and the fact that p is not sufficient to make it the case that it may be that q, then $s(p, q, w) = (0, x)$, where:
    - $x = \{y \in D$ | there is some $v \in W$ such that:
        - At v:
            - It is not the case that the fact that p is sufficient to explain why it may be that q; and
            - For all r, either it must be that r or it is not the case that it may be that $\neg r$; and
        - $y = d(v, w)$}; and
- If, at w, it is not true that p, then $s(p, q, w) = (0, \varnothing)$.

6.4.2 Stronger Reason, Equally Strong Reason, and at Least as Strong a Reason

I provided the tools for comparing amounts of reason in my analysis of REASON (mass) above. In order to make sense of judgements comparing strengths of reasons, then, we can simply apply these tools to the amounts of reason that are provided by these reasons. Thus, we can say that, for all *p*, *q*, *r*, *t*, *w*, and *v*, to judge that the amount of reason for *q* to be the case that is provided by the fact that *p* at *w* is *greater than* the amount of reason for *t* to be the case that is provided by the fact that *r* at *v just is* to judge that $s(p, q, w) >_3 s(r, t, v)$. This is so just in case either:

1. The fact that *p* explains why it may be that *q* at *w*, and the fact that *r* does not explain why it may be that *t* at *v*; or
2. The fact that *p* explains why it may be that *q* at *w*, and the fact that *r* explains why it may be that *t* at *v*, but the fact that *p* is a more *robust* explanation for the fact that it may be that *q* at *w* than the fact that *r* is for the fact that it may be that *t* at *v*—that is, roughly, the closest worlds to *v* at which the fact that *r* does not explain why it may be that *t* are closer to *v* than the closest worlds to *w* at which the fact that *p* does not explain why it may be that *q* are to *w*; or
3. The fact that *p* does not explain why it may be that *q* at *w*, and the fact that *r* does not explain why it may be that *t* at *v*, but the fact that *p* comes *closer* to explaining why it may be that *q* at *w* than the fact that *r* does to explaining why it may be that *t* at *v*—that is, roughly, the closest worlds to *w* at which the fact that *p* explains why it may be that *q* are closer to *w* than the closest worlds to *v* at which the fact that *r* explains why it may be that *t* are to *v*.

Similarly, we may say that, for all *p*, *q*, *r*, *t, w*, and *v*, to judge that the amount of reason for *q* to be the case that is provided by the fact that *p* at *w* is *equal to*, *at least as great as*, or *on a par with* the amount of reason for *t* to be the case that is provided by the fact that *r* at *v just is* to judge that $s(p, q, w) =_3 s(r, t, v)$, $s(p, q, w) \geq_3 s(r, t, v)$, or $s(p, q, w) \approx_3 s(r, t, v)$, respectively.

The fact that $>_3$ is, by definition, a strict partial order makes sense of the fact that it is *a priori* that STRONGER REASON picks out a property that has the properties of a strict partial order. The fact that $=_3$ is, by definition, an equivalence relation makes sense of

the fact that it is *a priori* that EQUALLY STRONG REASON picks out a relation that has the properties of an equivalence relation. The fact that $\geq_3$ is, by definition, a quasi order makes sense of the fact that AT LEAST AS STRONG A REASON picks out a relation that has the properties of a quasi order. And the fact that $\approx_3$ is, by definition a reflexive and symmetric relation makes sense of the fact that it is *a priori* that ON A PAR WITH RESPECT TO STRENGTH OF REASON picks out a reflexive and symmetric relation.

6.4.3 Threshold Amount

(1, *D*) is an amount of reason that marks the threshold between a reason's explaining why something is permitted and its not explaining why it is permitted. That is, for all *p*, *q*, and *w*, $s(p, q, w) \geq_3 (1, D)$, if, and only if, at *w*, the fact that *p* explains why it may be that *q*. It is thus natural to analyse SUFFICIENT REASON and CONCLUSIVE REASON in terms of this threshold amount. So we may say that, for all *p*, *q*, and *w*, to judge that, at *w*, the fact that *p* is a sufficient reason for it to be the case that *q just is* to judge that $s(p, q, w) \geq_3 (1, D)$; and to judge that, at *w*, the fact that *p* is a conclusive reason for it to be the case that *q just is* to judge that $s(p, q, w) \geq_3 (1, D)$ and $(1, D) >_3 r(\neg p, w)$.

6.4.4 Minimum Amount

As discussed above, it is natural to think of (0, Ø) as the minimum amount of reason. It follows that, for all *p*, *q*, and *w*, to judge that, at *w*, the fact that *p* is a *pro tanto* reason for it to be the case that *q just is* to judge that $s(p, q, w) >_3 (0, \varnothing)$. One possible objection to this analysis is that it implies that there are very many more *pro tanto* reasons than we should be willing to countenance. This is because, for all *p*, *q*, and *w*, $s(p, q, w) >_3 (0, \varnothing)$ just in case there is some world in the set of worlds *W*—no matter how distant—at which the fact that *p* explains why it may be that *q*. One might think that, for almost any *p* and *q*, there is some world at which the fact that *p* explains why it may be that *q*. For example, take any two apparently unrelated propositions—say, the proposition that it is Thursday and the proposition that I blink my eyes 20 times in the next five seconds. One might think that there is some world at which the fact that it is Thursday explains why I may blink my eyes 20 times in the next five seconds. For example, there is presumably a world at which I have been offered an enormous financial reward for blinking my eyes 20 times in five seconds

on a Thursday. At this world, it seems natural to say that the fact that it is Thursday explains why it is permissible for me to blink my eyes 20 times in the next five seconds—after all, if were not Thursday, then it would presumably not be worth the effort. On the closeness account, it follows that the fact that it is Thursday is *in fact* a reason for me to blink my eyes 20 times in the next five seconds. But this is absurd. Therefore the closeness analysis of REASON (count) is false.

There are three ways the proponent of the closeness analysis could respond to this objection. The first is to deny that, at the world at which I have been offered the reward for blinking my eyes on a Thursday, the fact that it is Thursday explains why I may blink my eyes 20 times in the next five seconds. One might argue instead that the fact that it is Thursday is merely a part of some larger fact—for example, the fact that it is Thursday and I have been offered an enormous reward for blinking my eyes 20 times in five seconds on a Thursday—that itself explains why it is permissible for me to blink my eyes 20 times in the next five seconds. This would avoid the unwanted consequence that the fact that it is Thursday is in fact a reason for me to blink my eyes 20 times in the next five minutes.

While this may solve the problem, however, adopting this kind of response threatens to result in there being too *few* reasons, according to the closeness analysis. For example, suppose that I have *in fact* been offered an enormous reward for blinking my eyes 20 times in five seconds on a Thursday. In this case, it seems plausible that the fact that it is Thursday is indeed a reason for me to blink my eyes 20 times in the next five seconds. But, on the view we are considering, this is false, since this fact is only a part of the explanation of why it is permissible for me to blink my eyes 20 times in the next five seconds. It may be that avoiding each of these undesirable outcomes could be avoided by making a distinction between ‘derivative’ and ‘non-derivative’ reasons (Maguire and Snedegar, 2021; Nair, 2016). We might say that a non-derivative reason for something to be so, on this account is a fact that, at some world, is the full explanation why it may be so, and a derivative reason for something to be so is a fact that is part of a non-derivative reason for it to be so. This may well be an appropriate way of dealing with this problem, although it is not clear to me that the distinction between derivative and non-derivative reasons is one that is to be found in our common-sense conception of reasons.

A second response to the too-many reasons objection is to point out that the set of worlds W need not contain every possible world. It is not implausible that people ignore very distant worlds when making judgements about reasons. A third response, however, is simply to bite the bullet, and accept that the fact that it is Thursday is a reason for me to blink my eyes 20 times in the next five seconds—albeit a vanishingly weak one. One might make this conclusion more palatable by engaging in the following kind of reasoning. There is an extremely miniscule chance that, if I were to blink my eyes 20 times in five seconds on a Thursday, then that would bring about the end of world poverty. This is clearly absurd, but not impossible. With this in mind, one might think that the fact that it is Thursday provides me with an extremely miniscule amount of reason to blink my eyes 20 times in the next five seconds, and is thus an extremely weak reason for me to do so—though one that is so weak that it does not come anywhere close to making it worth doing. I take it, then that the advocate of the closeness account of REASON (count) can deal with this objection one way or another.

6.4.5 Judgements about Reason and Judgements about Reasons

These analyses of REASON (mass) and REASON (count) vindicate the rational principles that govern how judgements involving these concepts relate to each other. The reason for this is that it follows from these analyses that the total amount of reason that there is for some response is always at least as great as the amount of reason that there is for that response that is provided by some fact. That is, for all p, q, and w, $r(q, w) \geq_3 s(p, q, w)$.

Assume for *reductio* that there is some p, q, and w such that it is not the case that $r(q, w) \geq s(p, q, w)$. Assume that there is some x, y, A, and B such that $r(q, w) = (x, A)$ and $s(p, q, w) = (y, B)$. It follows that either (i) $x = 0$ and $y = 1$; (ii) $x = y = 0$ and it is not the case that $A \geq_2 B$; or (iii) $x = y = 1$ and it is not the case that $B \geq_2 A$. Assume that (i) is true. It follows that, at w, the fact that p explains why it may be that q, and, at w, it is not the case that it may be that q. Since this is impossible, it follows that (i) is false. Assume that (ii) is true. It follows that there is some $u \in W$ at which the fact that p explains why it may be that q, and there is no v such that $d(v, w) \geq_1 d(u, w)$ and, at u, it may be that q. But there is some such u, namely $u = v$. It follows by *reductio* that (ii) is false. Finally, assume that (iii) is true. It follows that there is some

$u \in W$ at which it is not the case that it may be that q and there is no $v \in W$ such that $d(v, w) \geq_1 d(u, w)$ and, at v, it is not the case that the fact that p explains why it may be that q. But there is some such v, namely $v = u$, since u meets this condition. It follows by *reductio* that (iii) is false. Since (i), (ii), and (iii) are all false, it follows by *reductio* that there is no p, q, and w such that $\neg(r(q, w) \geq s(p, q, w))$.

PRO TANTO REASON implies SOME REASON.

$$\frac{s(p, q, w) \geq_3 (0, \varnothing)}{r(q, w) \geq_3 (0, \varnothing)}$$

Suppose that there is some p, q, and w such that $s(p, q, w) \geq_3 (0, \varnothing)$. Since $r(q, w) \geq_3 s(p, q, w)$, it follows by the transitivity of $\geq_3$ that $r(q, w) \geq_3 (0, \varnothing)$.

SUFFICIENT REASON (count) implies SUFFICIENT REASON (mass).

$$\frac{s(p, q, w) \geq_3 (1, D)}{r(q, w) \geq_3 (1, D)}$$

Suppose that there is some p, q, and w such that $s(p, q, w) \geq_3 (1, D)$. Since $r(q, w) \geq_3 s(p, q, w)$, it follows by the transitivity of $\geq_3$ that $r(p, q, w) \geq_3 (1, D)$.

CONCLUSIVE REASON (count) implies CONCLUSIVE REASON (mass).

$$\frac{s(p, q, w) \geq_3 (1, D);\ (1, D) >_3 r(\neg q, w)}{r(q, w) \geq_3 (1, D);\ (1, D) >_3 r(\neg q, w)}$$

This follows straightforwardly since $s(p, q, w) \geq_3 (1, D)$ implies that $r(q, w) \geq_3 (1, D)$.

6.4.6 Summary

In this section, I have presented an analysis of the concept REASON (count) according to which judgements about reasons are judgements about the amounts of reason that are provided by certain facts, where this is understood in terms of the relative closeness of worlds at which these facts explain why certain things may be so. I showed that this analysis makes sense of the idea of reasons of different strengths, satisfying our first desideratum. I also showed that the analysis validates the intuitive connections between the concepts REASON (count) and REASON (mass), satisfying our

second desideratum. The analysis also satisfies our third desideratum, in so far as it entails that reasons are facts that *explain* certain deontic facts at various more or less close-by worlds. And I also went some way toward showing that this analysis satisfies our fifth desideratum, by defending it against some putative counterexamples.

The analysis also appears to satisfy our fourth desideratum, since it seems to be neutral with respect to the existence of the right and wrong kinds of reasons. In fact, if the discussion of different kinds of modality in chapter one is on the right track, then this analysis seems well placed to account for the distinction between the right and wrong kinds of reason. In chapter one, I discussed the possibility that there are different kinds of modality, such that a distinction can be drawn between the *practical* MAY and MUST, the *evaluative* MAY and MUST, and the *rational* MAY and MUST. If these distinctions are tenable, and REASON (count) is analysable in terms of MAY, then we may be able to distinguish reasons of the right and wrong kinds in the following way: to judge that something is a reason of the right kind is to make a judgement about worlds at which something may be the case in the *rational* sense, whereas to judge that something is a reason of the wrong kind is to make a judgement about worlds at which something may be the case in either the *practical* or *evaluative* sense. That this analysis offers a way of distinguishing the right and wrong kinds of reason is thus another added advantage of this view.

6.5 Conclusion

In this appendix, I have provided new analyses of the concepts REASON (mass) and REASON (count), and shown that they satisfy a number of desiderata for concepts of this kind. According to these analyses, REASON (mass) and REASON (count) are analysable in terms of the concepts MAY and MUST. In chapter one, I argued that these concepts are themselves analysable in terms of fitting-attitude concepts. In chapter two, I provided an analysis of fitting-attitude concepts in terms of the concept RATIONAL. Combined, these views constitute an attractive rationality-first account of normativity.

Bibliography

Anscombe, G. E. M. (1957) *Intention*. Cambridge: Harvard University Press.

Aristotle (2009) *Nicomachean Ethics*. Oxford: Oxford University Press.

Baker, D. C. (2017) 'The Verdictive Organization of Desire'. *Canadian Journal of Philosophy* 47 (5), 589-612.

Barnett, D. J. (2021) 'Self-Knowledge Requirements and Moore's Paradox'. *Philosophical Review* 130 (2), 227-262.

Bennett, J. (1974) 'Counterfactuals and Possible Worlds'. *Canadian Journal of Philosophy* 4 (2), 381-402.

——— (2003) *A Philosophical Guide to Conditionals*. Clarendon Press.

Bilgrami, A. (2012) *Self-Knowledge and Resentment*. Cambridge: Harvard University Press.

Boghossian, P. (2005) 'Is Meaning Normative?'. in *Philosophy – Science – Scientific Philosophy. Main Lectures and Colloquia of Gap.5, Fifth International Congress of the Society for Analytical Philosophy, Bielefeld, 22–26 September 2003.* ed. by Christian, N., and Ansgar, B. Paderborn: Mentis, 205-218.

Braddon-Mitchell, D., and Jackson, F. (1996) *Philosophy of Mind and Cognition*. Cambridge: Blackwell.

Bratman, M. (1987) *Intention, Plans, and Practical Reason*. Cambridge: Harvard University Press.

——— (1996) 'Identification, Decision, and Treating as a Reason'. *Philosophical Topics* 24 (2), 1-18.

——— (2007) *Structures of Agency: Essays*. Oxford: Oxford University Press.

Bratman, M. E. (2009) 'Intention, Belief, Practical, Theoretical'. in *Spheres of Reason: New Essays in the Philosophy of Normativity.* ed. by Robertson, S. Oxford University Press.

Broome, J. (1999) 'Normative Requirements'. *Ratio* 12 (4), 398-419.

——— (2005) 'Does Rationality Give Us Reasons?'. *Philosophical Issues* 15 (1), 321-337.

——— (2007) ‘Is Rationality Normative?’. *Disputatio* 2 (23), 161-178.

——— (2013) *Rationality through Reasoning*. Malden: John Wiley & Sons.

——— (2016) ‘Responses’. *Philosophical Studies* 173 (12), 3431-3448.

——— (2018) ‘Reason Fundamentalism and What Is Wrong with It’. *The Oxford handbook of reasons and normativity*, 297.

Brown, C. (2014) ‘Minding the Is-Ought Gap’. *Journal of Philosophical Logic* 43 (1), 53-69.

Brunero, J. (2008) ‘Against Cognitivism About Practical Rationality’. *Philosophical Studies* 146 (3), 311-325.

——— (2009) ‘Reasons and Evidence One Ought’. *Ethics* 119 (3), 538-545.

——— (2010) ‘The Scope of Rational Requirements’. *Philosophical Quarterly* 60 (238), 28-49.

——— (2012) ‘Instrumental Rationality, Symmetry and Scope’. *Philosophical Studies* 157 (1), 125-140.

——— (2018) ‘Reasons, Evidence, and Explanations’. in *The Oxford Handbook of Reasons and Normativity.* ed. by Star, D. Oxford: Oxford University Press, 321-341.

Burge, T. (1996) ‘Our Entitlement to Self-Knowledge’. *Proceedings of the Aristotelian Society* 96 (1), 91-116.

Cariani, F. (2009) *The Semantics of ‘Ought’ and the Unity of Modal Discourse.* Doctoral thesis, University of California, Berkeley.

Chang, R. (2016) ‘Parity, Imprecise Comparability, and the Repugnant Conclusion’. *Theoria* 82 (2), 183-215.

Chappell, R. Y. (2012) ‘Fittingness: The Sole Normative Primitive’. *The Philosophical Quarterly* 62 (249), 684-704.

Chisholm, R. M. (1963) ‘Contrary-to-Duty Imperatives and Deontic Logic’. *Analysis* 24 (2), 33-36.

Coates, A. (2013) ‘The Enkratic Requirement’. *European Journal of Philosophy* 21 (2), 320-333.

Coliva, A. (2012) ‘One Variety of Self-Knowledge: Constitutivism as Constructivism’. in *The Self and Self-Knowledge.* ed. by Coliva, A. Oxford: Oxford University Press, 212-242.

Cresswell, M. J., and Hughes, G. E. (1996) *A New Introduction to Modal Logic*. London: Routledge.

Dancy, J. (2000a) ‘Editor’s Introduction’. in *Normativity.* ed. by Dancy, J. Oxford: Blackwell.

——— (2000b) *Practical Reality*. Oxford: Oxford University Press.

Daniels, N. (1980) ‘On Some Methods of Ethics and Linguistics’. *Philosophical Studies* 37 (1), 21-36.

Darwall, S. (1997) ‘Reasons, Motives, and the Demands of Morality: An Introduction’. in *Moral Discourse and Practice: Some Philosophical Approaches.* ed. by Darwall, S., Gibbard, A., and Railton, P. Oxford: Oxford University Press, 305-312.

——— (2001) ‘Normativity’. in *Routledge Encyclopedia of Philosophy.* ed. by Craig, E. London: Routledge.

Davidson, D. (1984) *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Dennett, D. C. (1989) *The Intentional Stance*. Cambridge: The MIT press.

Dworkin, G. (1988) *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.

Ellis, B., Jackson, F., and Pargetter, R. (1977) ‘An Objection to Possible-World Semantics for Counterfactual Logics’. *Journal of Philosophical Logic* 6 (1), 355 - 357.

Fernandez, J. (2013) *Transparent Minds: A Study of Self-Knowledge*. Oxford: Oxford University Press.

Fine, K. (1975) ‘Critical Notice of Lewis, Counterfactuals’. *Mind* 84 (335), 451 - 458.

Finlay, S. (2006) ‘The Reasons That Matter’. *Australasian Journal of Philosophy* 84 (1), 1 – 20.

——— (2014) *Confusion of Tongues: A Theory of Normative Language*. New York: Oxford University Press.

——— (2019) ‘Defining Normativity’. in *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence.* ed. by Toh, K., Plunkett, D., and Shapiro, S. Oxford: Oxford University Press, 187-219.

Fogal, D. (2016) ‘Reasons, Reason, and Context’. in *Weighing Reasons.* ed. by Lord, E., and Maguire, B. Oxford: Oxford University Press, 74-103.

Fogal, D., and Risberg, O. (2021) ‘Explaining Normative Reasons’. *Noûs* early view, 1-30.

Frankfurt, H. (1988) *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press.

——— (1999) *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.

——— (2004) ‘Disengaging Reason’. in *Reason and Value: Themes from the Moral Philosophy of Joseph Raz.* ed. by Wallace, R. J. Oxford: Oxford University Press, 117-128.

Frankfurt, H. G. (1969) ‘Alternate Possibilities and Moral Responsibility’. *The Journal of Philosophy* 66 (23), 829-839.

Gert, J. (2004) ‘Value and Parity’. *Ethics* 114 (3), 492-510.

Gibbard, A. (1990) *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge: Harvard University Press.

——— (2003) *Thinking How to Live*. Cambridge, Massachusetts: Harvard University Press.

Gilabert, P. (2009) ‘The Feasibility of Basic Socioeconomic Human Rights: A Conceptual Exploration’. *Philosophical Quarterly* 59 (237), 659-681.

Glüer, K., and Wikforss, Å. (2009) ‘Against Content Normativity’. *Mind* 118 (469), 31-70.

Goldman, A. I. (2000) ‘Folk Psychology and Mental Concepts’. *ProtoSociology* 14, 4-25.

Gregory, A. (2016) ‘Normative Reasons as Good Bases’. *Philosophical Studies* 173 (9), 2291-2310.

Harman, G. (1976) ‘Practical Reasoning’. *The Review of Metaphysics*, 431-463.

Hieronymi, P. (2006) ‘Controlling Attitudes’. *Pacific Philosophical Quarterly* 87 (1), 45-74.

——— (2011) ‘Reasons for Action’. *Proceedings of the Aristotelian Society* 111 (3), 407-427.

——— (2021) ‘Reasoning First’. in *Routledge Handbook of Practical Reasoning.* ed. by Chang, R., and Sylvan, K. New York: Routledge, 349-365.

Howard, C. (2019) ‘The Fundamentality of Fit’. in *Oxford Studies in Metaethics.* ed. by Shafer-Landau, R. Oxford: Oxford University Press, 216-236.

Humberstone, I. L. (1971) ‘Two Sorts of 'Ought's’. *Analysis* 32 (1), 8 - 11.

Hume, D. (2000) *A Treatise of Human Nature*. Oxford: Oxford University Press.

Hussain, N. (2007) ‘The Requirements of Rationality’. Unpublished manuscript. Stanford University.

Jackson, F. (1974) ‘Defining the Autonomy of Ethics’. *Philosophical Review* 83 (1), 88-96.

——— (1985) ‘On the Semantics and Logic of Obligation’. *Mind* 94 (374), 177-195.

Jackson, F., and Pargetter, R. (1986) ‘Oughts, Options, and Actualism’. *The Philosophical Review* 95 (2), 233-255.

Kamp, H. (1974) ‘Iv—Free Choice Permission’. *Proceedings of the Aristotelian Society* 74 (1), 57-74.

Kant, I. (1959) *Grundlegung Zur Metaphysik Der Sitten*. New York: Macmillan Library of Liberal Arts.

Kauppinen, A. (2015) ‘Favoring’. *Philosophical Studies* 172 (7), 1953-1971.

Kearns, S., and Star, D. (2008) ‘Reasons: Explanations or Evidence?’. *Ethics* 119 (1), 31-56.

——— (2009) ‘Reasons as Evidence’. in *Oxford Studies in Metaethics.* ed. by Shafer-Landau, R. Oxford: Oxford University Press, 215-242.

——— (2013) ‘Weighing Reasons’. *Journal of Moral Philosophy* 10 (1), 70-86.

Kelly, T., and Mcgrath, S. (2010) ‘Is Reflective Equilibrium Enough?’. *Philosophical perspectives* 24, 325-359.

Kiesewetter, B. (2017) *The Normativity of Rationality*. Oxford: Oxford University Press.

Kolodny, N. (2005) ‘Why Be Rational?’. *Mind* 114 (455), 509-563.

——— (2007) ‘How Does Coherence Matter?’. *Proceedings of the Aristotelian Society* 107 (1), 229-263.

——— (2008) ‘Why Be Disposed to Be Coherent?’. *Ethics* 118 (3), 437-463.

Korsgaard, C. M. (1996) *The Sources of Normativity*. Cambridge: Cambridge University Press.

——— (2008) *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.

——— (2009) *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Kratzer, A. (1977) 'What 'Must'and 'Can'must and Can Mean'. *Linguistics and philosophy* 1 (3), 337-355.

——— (1981) 'The Notional Category of Modality'. in *Words, Worlds, and Contexts.* ed. by Eikmeyer, H. J., and Rieser, H., 38-74.

——— (1986) 'Conditionals'. *Proceedings of the Chicago Linguistics Society* 22 (2), 1-15.

——— (1991) 'Modality'. in *Semantics: An International Handbook of Contemporary Research.* ed. by Stechow, A. V. Oxford: Oxford University Press, 639-650.

Lassiter, D. (2011) *Measurement and Modality: The Scalar Basis of Modal Semantics.* Doctoral thesis, New York University.

——— (2017) *Graded Modality: Qualitative and Quantitative Perspectives*. Oxford: Oxford University Press.

Lee, K. Y. (2020) 'Reasons as Evidence against Ought-Nots'. *Philosophical Papers* 49 (3), 431-455.

Lewis, D. (1973) *Counterfactuals*. Oxford: Blackwell.

——— (1974) 'Radical Interpretation'. *Synthese* 27 (July-August), 331-344.

——— (1979) 'Counterfactual Dependence and Time's Arrow'. *Noûs* 13 (4), 455-476.

Lewis, D. K. (1977) 'Possible-World Semantics for Counterfactual Logics: A Rejoinder'. *Journal of Philosophical Logic* 6 (1), 359-363.

Lord, E. (2011) 'Violating Requirements, Exiting from Requirements, and the Scope of Rationality'. *Philosophical Quarterly* 61 (243), 392-399.

——— (2013) 'The Real Symmetry Problem(S) for Wide-Scope Accounts of Rationality'. *Philosophical Studies* (3), 1-22.

——— (2016) ‘On the Rational Power of Aesthetic Testimony’. *British Journal of Aesthetics* 56 (1), 1-13.

——— (2018) *The Importance of Being Rational*. Oxford: Oxford University Press.

Maguire, B. (2016) ‘The Value-Based Theory of Reasons’. *Ergo: An Open Access Journal of Philosophy* 3.

Maguire, B., and Snedegar, J. (2021) ‘Normative Metaphysics for Accountants’. *Philosophical Studies* 178 (2), 363-384.

Markovits, J. (2010) ‘Acting for the Right Reasons’. *Philosophical Review* 119 (2), 201-242.

Marušić, B., and Schwenkler, J. (2018) ‘Intending Is Believing: A Defense of Strong Cognitivism’. *Analytic Philosophy* 59 (3), 309-340.

Mcdermott, M. (2007) ‘True Antecedents’. *Acta Analytica* 22 (4), 333-335.

Mchugh, C., and Way, J. (2016) ‘Fittingness First’. *Ethics* 126 (3), 575-606.

Moore, G. E. (1959) *Philosophical Papers*. London: Routledge.

Moran, R. A. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

Nair, S. (2016) ‘How Do Reasons Accrue?’. in *Weighing Reasons.* ed. by Lord, E., and Maguire, B. Oxford: Oxford University Press, 56–73.

Nozick, R. (1981) *Philosophical Explanations*. Belknap Press.

Nute, D. (1975) ‘Counterfactuals and the Similarity of Worlds’. *Journal of Philosophy* 72 (21), 773-778.

Parfit, D. (1984) *Reasons and Persons*. Oxford University Press, UK.

——— (2001) ‘Rationality and Reasons’. in *Exploring Practical Philosophy: From Action to Values.* ed. by Egonsson, D., Josefsson, J., Petersson, B., and Rønnow-Rasmussen, T. London: Routledge, 17-39.

——— (2011) *On What Matters*. Oxford: Oxford University Press.

Pigden, C. (1989) ‘Logic and the Autonomy of Ethics’. *Australasian Journal of Philosophy* 67 (2), 127-151.

——— (ed.) (2010) *Hume on Is and Ought*. New York: Palgrave-Macmillan.

Portner, P. (2009) *Modality*. Oxford: Oxford University Press.

Prior, A. N. (1960) 'The Autonomy of Ethics'. *Australasian Journal of Philosophy* 38 (3), 199-206.

Quine, W. V. O. (1960) *Word and Object*. Cambridge: The MIT Press.

Rabinowicz, W. (2008) 'Value Relations'. *Theoria* 74 (1), 18-49.

——— (2009) 'Values Compared'. *Polish Journal of Philosophy* 3 (1), 73-96.

——— (2012) 'Value Relations Revisited'. *Economics and Philosophy* 28 (2), 133-164.

——— (2017) 'From Values to Probabilities'. *Synthese* 194 (10), 3901-3929.

Rabinowicz, W., and Rønnow-Rasmussen, T. (2004) 'The Strike of the Demon: On Fitting Pro-Attitudes and Value'. *Ethics* 114 (3), 391-423.

Rawls, J., - (1971) *A Theory of Justice*. Cambridge, MA. :: Bellknap Press.

Raz, J. (1999) *Engaging Reason: On the Theory of Value and Action*. Oxford: Oxford University Press

——— (2000) 'Explaining Normativity: On Rationality and the Justification of Reason'. in *Normativity.* ed. by Dancy, J. Oxford: Oxford: Blackwell publishers, 34-60.

——— (2010) 'Reason, Reasons and Normativity'. in *Oxford Studies in Metaethics.* ed. by Shafer-Landau, R. Oxford: Oxford University Press.

Ross, A. (1944) 'Imperatives and Logic'. *Philosophy of Science* 11 (1), 30-46.

Russell, B. (2020) 'A Priori Justification and Knowledge'. *The Stanford Encyclopedia of Philosophy.*

Saebo, K. (1985) 'Notwendige Bedingungen Im Deutschen. Zur Semantik Modalisierter Sätze.(Les Conditions Nécessaires En Allemand. La Sémantique Des Phrases Modalisées)'. *Sonderforschungsbereich 99 Universität Konstanz Konstanz* (108), 1-234.

Schroeder, M. (2004) 'The Scope of Instrumental Reason'. *Philosophical Perspectives* 18 (1), 337–364.

——— (2007) *Slaves of the Passions*. Oxford: Oxford University Press.

——— (2009) ‘Means-End Coherence, Stringency, and Subjective Reasons’. *Philosophical Studies* 143 (2), 223 - 248.

——— (2011) ‘Ought, Agents, and Actions’. *Philosophical Review* 120 (1), 1-41.

——— (2012) ‘The Ubiquity of State-Given Reasons’. *Ethics* 122 (3), 457-488.

——— (2015) ‘What Makes Reasons Sufficient?’. *American Philosophical Quarterly* 52 (2), 159-170.

Searle, J. R. (2001) *Rationality in Action*. Cambridge: MIT Press.

Setiya, K. (2007a) ‘Cognitivism About Instrumental Reason’. *Ethics* 117 (4), 649-673.

——— (2007b) *Reasons without Rationalism*. Princeton: Princeton University Press.

——— (2014) ‘What Is a Reason to Act?’. *Philosophical Studies* 167 (2), 221-235.

Shoemaker, S. (1988) ‘On Knowing One’s Own Mind’. *Philosophical Perspectives* 2, 183-209.

——— (1994) ‘Self-Knowledge and "Inner Sense" Lecture Ii: The Broad Perceptual Model’. *Philosophy and Phenomenological Research* 54 (2), 271-290.

——— (1995) ‘Moore’s Paradox and Self-Knowledge’. *Philosophical Studies* 77 (2-3), 211-228.

Silk, A. (2015) ‘What Normative Terms Mean and Why It Matters for Ethical Theory’. in *Oxford Studies in Normative Ethics, Vol. 5.* ed. by Timmons, M. C., 296–325.

——— (2021) ‘Weak and Strong Necessity Modals: On Linguistic Means of Expressing "a Primitive Concept Ought"’. in *Meaning, Decision, and Norms: Themes from the Work of Allan Gibbard.* ed. by Dunaway, B., and Plunkett, D. Ann Arbor: Maize Books.

Skorupski, J. (2007) ‘Buck-Passing About Goodness’. in *Hommage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz* [online]. ed. by Rønnow-Rasmussen, T., Petersson, B., Josefsson, J., and Egonsson, D. Lund: Department of Philosophy, Lund University.

Smith, M. (1994) *The Moral Problem*. Oxford: Blackwell Oxford.

Smithies, D. (2016) ‘Belief and Self-Knowledge: Lessons from Moore's Paradox’. *Philosophical Issues* 26 (1), 393-421.

Southwood, N. (2008) ‘Vindicating the Normativity of Rationality’. *Ethics* 119 (1), 9-30.

——— (2016a) ‘Does “Ought” Imply “Feasible”?’. *Philosophy and Public Affairs* 44 (1), 7-45.

——— (2016b) ‘The Motivation Question’. *Philosophical Studies* 173 (12), 3413-3430.

——— (2018) ‘Constructivism About Reasons’. in *The Oxford Handbook of Reasons and Normativity.* ed. by Star, D. Oxford: Oxford University Press.

Stalnaker, R. (1968) ‘A Theory of Conditionals’. in *Studies in Logical Theory (American Philosophical Quarterly Monographs 2).* ed. by Rescher, N. Oxford: Blackwell, 98-112.

Star, D. (2018) ‘Reasoning with Reasons’. in *Normativity: Epistemic and Practical.* ed. by Mchugh, C., Way, J., and Whiting, D. Oxford: Oxford University Press, 241-259.

Strawson, G. (2003) ‘Mental Ballistics or the Involuntariness of Spontaniety’. *Proceedings of the Aristotelian Society* 103 (3), 227-257.

Street, S. (2008) ‘Constructivism About Reasons’. in *Oxford Studies in Metaethics.* ed. by Shafer-Landau, R. Oxford: Oxford University Press, 207-245.

——— (2010) ‘What Is Constructivism in Ethics and Metaethics?’. *Philosophy Compass* 5 (5), 363-384.

Studtmann, P. (2003) ‘Counterfactuals and Inferences a New Form of the Three-Parameter Account of Counterfactuals’. *Australasian Journal of Philosophy* 81 (1), 51 – 61.

Tappolet, C. (2011) ‘The Normativity of Evaluative Concepts’. in *Mind, Values, and Metaphysics.* ed. by Reboul, A. Cham: Springer, 39-54.

Tappolet, C., and Voizard, A. (2011) ‘The Philosophy of Normativity, or How to Try Clearing Things up a Little’. *Dialogue: Canadian Philosophical Review/Revue canadienne de philosophie* 50 (Special Issue 02), 233-238.

Taylor, C. (1976) ‘Responsibility for Self’. in *The Identities of Persons.* ed. by Rorty, A. O. Berkeley: University of California Press, 281--299.

——— (1985) *Human Agency and Language*. Cambridge: Cambridge University Press.

Temkin, L. S. (2012) *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.

Thomson, J. J. (2008) *Normativity*. Chicago: Open Court Press.

Vallentyne, P. (1989) 'Two Types of Moral Dilemmas'. *Erkenntnis* 30 (3), 301-318.

Velleman, D. (1989) *Practical Reflection*. Princeton: Princeton University Press.

——— (2000) *The Possibility of Practical Reason*. Oxford: Clarendon Press.

——— (2005) 'Précis of the Possibility of Practical Reason'. *Philosophical studies* 121 (3), 225-238.

——— (2009) *How We Get Along*. Cambridge: Cambridge University Press.

Von Fintel, K. (2012) 'The Best We Can (Expect to) Get? Challenges to the Classic Semantics for Deontic Modals'. *Central meeting of the american philosophical association.*

Von Fintel, K., and Iatridou, S. (2008) 'How to Say Ought in Foreign: The Composition of Weak Necessity Modals'. in *Time and Modality.* ed. by Guéron, J., and Lecarme, J. Springer, 115-141.

Wallace, R. J. (2001) 'Normativity, Commitment, and Instrumental Reason'. *Philosophers' Imprint* 1 (3), 1-26.

Walters, L., and Williams, R. (2013) 'An Argument for Conjunction Conditionalization'. *Review of Symbolic Logic* 6 (4), 573-588.

Watson, G. (2003) 'The Work of the Will'. in *Weakness of Will and Practical Irrationality.* ed. by Watson, G., Stroud, S., and Tappolet, C. Oxford: Clarendon Press, 172-200.

Way, J. (2010a) 'Defending the Wide-Scope Approach to Instrumental Reason'. *Philosophical studies* 147 (2), 213-233.

——— (2010b) 'The Normativity of Rationality'. *Philosophy Compass* 5 (12), 1057-1068.

——— (2011) 'The Symmetry of Rational Requirements'. *Philosophical Studies* 155 (2), 227-239.

——— (2012) 'Transmission and the Wrong Kind of Reason'. *Ethics* 122 (3), 489-515.

——— (2017) 'Reasons as Premises of Good Reasoning'. *Pacific Philosophical Quarterly* 98 (2), 251-270.

Wedgwood, R. (2007) *The Nature of Normativity*. New York: Oxford University Press.

——— (2009) ‘The "Good" and the "Right" Revisited’. *Philosophical Perspectives* 23 (1), 499-519.

——— (2013) ‘Normativity’. *The International Encyclopedia of Ethics.* John Wiley & Son.

Whiting, D. (2015) ‘The Glass Is Half Empty: A New Argument for Pessimism About Aesthetic Testimony’. *British Journal of Aesthetics* 55 (1), 91-107.

Williams, B. (1981) *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press.

Wittgenstein, L. (1953) *Philosophical Investigations*. Oxford: Blackwell.

Wodak, D. (2021) ‘Approving on the Basis of Moral and Aesthetic Testimony’. in *Oxford Studies in Metaethics.* ed. by Shafer-Landau, R. Oxford: Oxford University Press.

Worsnip, A. (2015) ‘Narrow-Scoping for Wide-Scopers’. *Synthese* 192 (8), 2617-2646.

——— (2018) ‘What Is (in)Coherence?’. in *Oxford Studies in Metaethics.* ed. by Shafer-Landau, R. Oxford: Oxford University Press, 184-206.

Zangwill, N. (1998) ‘Direction of Fit and Normative Functionalism’. *Philosophical Studies* 91 (2), 173-203.

——— (2005) ‘The Normativity of the Mental’. *Philosophical Explorations* 8 (1), 1-19.

——— (2010) ‘Normativity and the Metaphysics of Mind’. *Australasian Journal of Philosophy* 88 (1), 1–19.

Zimmerman, M. J. (2006) ‘Moral Luck: A Partial Map’. *Canadian Journal of Philosophy* 36 (4), 585-608.