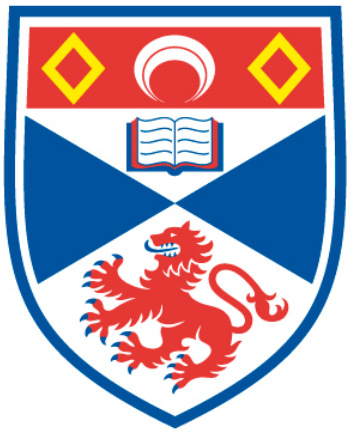
Attitudes first:
rationality attributions and the normativity of rationality

Lisa Bastian

A thesis submitted for the degree of PhD
at the
University of St Andrews



2020

Abstract

This thesis has two distinct aims. The first is to shed light on our practice of attributing rationality to others. To begin, Chapter 2 demonstrates that we cannot rely on questions of what rationality requires to make sense of this practice. Chapter 3 explores a different strategy and directly engages with rationality attributions. It lays out some desiderata for an adequate account of such attributions. Chapter 4 develops a novel account of rationality attributions. This account focusses on explicitly mentioning sets of an agent's attitudes, and also includes a measure for the attribution's robustness. Thanks to these features, the account meets the desiderata, and also allows for progress on persisting disagreements in the debate. Chapter 5 further illustrates the account by contrasting it with an alternative contextualist understanding of rationality attributions.

The second aim of this thesis is to defend the Normativity of Rationality. In Chapter 6, I consider problems for a reasons-based understanding of the Normativity of Rationality which arise from so-called transmission principles, and point out strategies to defend rationality's normativity. Chapter 7 provides further support for one of these strategies. Finally, Chapter 8 presents my positive argument. I propose to understand rationality's normativity in terms of *commitment* – if you are rationally required to x, you are committed to x. Commitment can avoid the counter-examples of alternative understandings in terms of *reasons* or *ought* by combining features of both notions. This makes commitment a promising normative notion in its own right.

Acknowledgements

So much more went went into this thesis than its seventy-thousand-something words. Throughout this journey, I was aided and supported by many people.

I count myself lucky to have had the support of my supervisory team, Dr Justin Snedegar and Prof Jessica Brown. Thank you for accompanying and supporting this project throughout its many stages, for tirelessly reading and commenting on countless versions, and for sharing your experience and wisdom about navigating the academic landscape with me.

Over the past four years, I have received many kind and critical words from the academic community. This feedback has been invaluable for the progress of this thesis.

Thanks to Prof Mark Schroeder, who challenged and encouraged me to approach my work from an entirely different perspective during my research visit at the University of Southern California.

I would also like to thank my colleagues in Arché and SASP for taking the time to read my work, for coming to my talks, for thinking through ideas with me and for sharing coffee breaks. I am proud to have been part of this community.

Thanks to all my safe havens, dotted across the globe:

Joseph Bowen, Ethan Landes, Anh Quan Nguyen, Saranga Sudarshan – we have embarked on this journey together and I could not have wished for a better cohort.

Marvin Backes, Miguel Egler, Claire Field, Lorenzo Lazzarini, Stefano Lo Re, Quentin Pharr, Lewis Ross, Clotilde Torregrossa – thank you for never tiring of reading and proof-reading my work, and for getting me out of my own head when I really needed to.

Arola Moreras Marti – you have been so much more than a flatmate. Thank you for being my friend, my ally, and for making this house a home.

Mook Attanath, Tessa Simpson – no mountain is too high, no scramble too steep for you. Walking up hills with you has kept me going. Thank you for always being up for an adventure.

Leonie Bleiker, Lena Cohrs, Theresa Nover, Bettina Schwab – you have been there for all of it. Thank you for believing in me, and for lending me an ear when I couldn't say I needed one.

And finally, I want to thank my parents. I am grateful for your unconditional support, and for always trusting me to make the right choices. You are the wisest people I know. This thesis is for you.

Journal Acknowledgements

I would like to thank an anonymous referee for *Philosophical Studies*, whose comments greatly improved the paper that is based on Chapter 7 of this thesis.

Funder Acknowledgements

This work was supported by the Arts and Humanities Research Council [grant number 1796583]; the Stirling and St Andrews Graduate Programme in Philosophy; the Royal Institute of Philosophy; and the McGlashan Trust.

Candidate's declaration

I, Lisa Bastian, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 77,200 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2016.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 14 July 2020 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 14 July 2020 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Lisa Bastian, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

Embargo on all of print copy for a period of 3 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for printed embargo request

The material of this thesis provides a novel contribution to the field, which I am currently preparing for submission to journals. I would like to request an embargo on this material so that this material is not made available to the public before the reviewing process of this material is finished.

Electronic copy

Embargo on all of electronic copy for a period of 3 years on the following ground(s):

- Publication would preclude future publication

Supporting statement for electronic embargo request

The material of this thesis provides a novel contribution to the field, which I am currently preparing for submission to journals. I would like to request an embargo on this material so that this material is not made available to the public before the reviewing process of this material is finished.

Title and Abstract

- I agree to the title and abstract being published.

Date 14 July 2020 Signature of candidate

Date 14 July 2020 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Lisa Bastian, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

Date 14 July 2020 Signature of candidate

Contents

Chapter One

Introduction: Rationality, Normativity, and Criticism

Rationality seems obviously normative – after all, it features heavily in our practices of evaluating agents, their thought processes and actions. We criticise our interlocutors for being irrational and praise students for displaying rationality.

This practice motivates one of the two main themes of this thesis: getting rationality attributions right. Because we use rationality in this important way, we better have a precise understanding of rationality attributions, so that we can decide when a rationality attribution is warranted, and when not. Ultimately, this has the further advantage of shedding some light on the property of rationality, and what it takes for agents to be rational.

The other main theme of this thesis is the Normativity of Rationality. The two main themes obviously interact. The Normativity of Rationality can provide an explanation for our practice of relying on rationality attributions in criticism or praise – the reason why someone is criticisable if they are irrational is because rationality is normative. So being irrational would amount to not living up to the demands of a normative notion. As we will see, it is notoriously difficult to provide an argument for the Normativity of Rationality that is not subject to a number of immediate counter-examples. Developing such an argument is the aim connected to this second main theme of the thesis.

Next to these two main themes, there is an underlying methodological current, or lesson, to this thesis. I will highlight at various points how greater complexity and specificity in our debates is helpful and might even be required. Specifically, my treatment of the matter will show that often, taking a closer and more fine-grained look at rationality can help to solve persisting problems. This is particularly obvious in my discussion of rationality attributions (Chapter 3), of transmission principles (Chapter 6) and of reasons of the right kind for belief (Chapter 7).

For now, let's turn to the first theme – rationality attributions.

1.1 Rationality Attributions

Agents do things all the time. They form opinions, make decisions and sometimes act on them. In philosophy, our investigation of these behaviours is often located at the level of the propositional attitudes that are involved. Forming an opinion is analysed in terms of the beliefs that play a role. Decisions and the resulting actions are examined via the relevant intentions. For example, an agent's opinion that democracy is the best political system is understood as her *belief in the proposition* that democracy is the best political system. Her decision to live a healthier life can be broken down into a number of intentions, like her *intention* to eat a balanced diet and her *intention* to exercise regularly.

In what follows, I introduce three examples which will accompany us throughout this entire discussion. They are supposed to show that i) rationality attributions present themselves quite naturally to us if we are presented with an agent's situation that is described at the propositional level, ii) that those attributions seem to naturally provide the grounds for criticism of or praise for the agent and iii) that despite their ready availability, which rationality attribution is correct is a non-trivial and potentially more complicated matter.

> *Example 1 – Sara without means*
>
> Sara intends to send off her paper by the end of the day. She believes that skipping her lunch break is a necessary means for achieving this. But Sara's favourite Italian restaurant runs a pizza lunch special that day and so she goes for a long lunch break.

Intuitively, we are inclined to criticise Sara for her behaviour. It does not seem rational for her to not intend the means to her ends; her attitudes do not seem to fit together in the right way. Based on this, we might want to attribute irrationality to Sara. But on further reflection, this verdict does not seem so obvious, even in this straightforward example. Suppose that Sara also believes that she should make a habit of prioritising regular lunch breaks over deadlines and that meeting this deadline is not hugely important for her career. If we take into account these additional beliefs of hers, dropping her intention to skip lunch does not seem so irrational anymore. Maybe we are now more inclined to attribute rationality to Sara in virtue of her decision.

Attributing rationality to agents becomes even trickier in less straightforward cases. Consider a slightly altered version of Sara's case:

> *Example 2 – Sara without ends*
>
> Sara intends to send off her paper by the end of the day, believes that skipping lunch is necessary for this and lacks the intention to skip lunch. But this time, Sara responds by dropping her intention to send off her paper today.

Can we attribute rationality to her? By dropping her intention, Sara avoids the criticism we raised against her in the previous case: her attitudes no longer stand in conflict with each other. If she does not intend to send off her paper today, there is nothing wrong with taking a lunch break. However, we might be reluctant to call her rational. Suppose that Sara believes that she has good reason to submit her paper on time in the first place. And giving up an intention that is strongly supported by one's reasons does not seem to warrant a rationality attribution.

Finally, consider a third case, adapted from Schroeder (2005), in which the move from an agent's performance, cashed out in terms of their attitudes, to a rationality attribution is also not straightforward.[1]

> *Example 3 – Muriel, the axe-murderer*
>
> Muriel intends to become the world's most prolific axe-murderer. She believes that in order to become the world's most prolific axe-murderer, it is necessary that she practices her swings and sharpens the blade. But she does not intend to practice her swings and sharpen the blade.

Because this case is similar to the first one, one available response is to consider Muriel irrational, since her attitudes stand in some kind of conflict with each other. But at the same time, not intending the necessary means to becoming an axe-murderer seems like the rational thing to do. After all, many factors speak against becoming an axe-murderer. And so we might be inclined to attribute rationality to Muriel.

These examples show how we rely on rationality attributions when we criticise agents, but also how developing a nuanced understanding of them is a serious challenge. Chapters 3 and 4 of this thesis are concerned with developing this understanding. At this point, the underlying methodological lesson of this thesis becomes apparent for the first time – adopting a more complex view of rationality attributions will be helpful in dissolving disagreements and in guiding our responses to the above examples. But for now, let me turn my attention to the second main theme of this thesis – the Normativity of Rationality.

1.2 The Normativity of Rationality

If rationality is normative, then we have a straightforward *explanation* of why we rely so much on rationality attributions in criticism. Unfortunately, it has proven very difficult to provide an argument for the Normativity of Rationality. Drawing on a parallel to morality only goes so far. For example, whilst non-compliance with the norms of morality, such as not keeping a promise, constitutes a serious failing, it is not obvious that, say, believing a contradiction would be a similarly serious kind of failing on the part of the agent. Hence, the Normativity of Rationality

[1] We will re-encounter this case in my discussion of transmission principles in Chapter 6.

requires a distinct argument in its own right. This is what has become known as the 'Normativity Problem' – the challenge of providing a theoretical argument that supports our intuitions and practices regarding rationality.

The Normativity Problem can be captured more precisely as the search for a true bridge principle – a principle that links the requirements of rationality to a normative notion.[2] This term appears in the related debate about the normativity of logic and has its predecessor in the debate about the normativity of morality. The challenge to be met here was raised by Harman (1984), who observed that the laws of logic and the norms that govern belief revision are not the same – deductive logic does not amount to a theory of reasoning. And whilst the claim that logic is normative for reasoning sounds intuitively plausible, there is a gap between the laws of logic on the one side and the norms of belief revision on the other. This gap makes it unclear how we should understand the normativity of logic. MacFarlane's (ms) response to this challenge is to establish bridge principles which are supposed to bridge said gap by "linking claims about logical validity with norms for belief" (ibid.: 5). Bridge principles follow a general template:

> (BP) If $A, B \models C$, then (normative claim about believing A, B, and C).[3]

They take the form of material conditionals, where logical validity features in the antecedent and the norm for belief in the consequent. Given the antecedent, we can then detach the consequent via modus ponens, making it such that we can link validity to norms of belief.

The template can be filled in a number of ways. In our case, the requirements of rationality take the place of claims about logical validity in the antecedent, and whatever satisfies the rational requirement is combined with a normative claim in the consequent. The two main candidates for taking up this spot in the consequent are *reasons* and *ought*. This gives us two potential bridge principles:

> *(BPO)* If rationality requires you to x, you ought to x.
>
> *(BPR)* If rationality requires you to x, you have a reason to x.

The question about the Normativity of Rationality then becomes a question about whether it is true that rational requirements give us reasons to do a certain thing, or whether it is true that we ought to do that thing. In this introduction, I focus on the reasons-version of bridge principles – *(BPR)* – for two reasons. One, *(BPR)* is the weaker of the two and most of the problems it faces apply to *(BPO)* as well and are often even more taxing there. And two, I will discuss *(BPO)* at length in Chapter 8.

[2] In the literature, discussions of the Normativity Problem can also be found in terms of the question whether a version of 'Normativity Claim' is true (e.g. Broome, 2013).

[3] (MacFarlane, ms: 6). Steinberger (2019) revises the template slightly, but maintains its structure.

1.2.1 A Problem for *(BPR)*: Detachment

In this introduction, I briefly introduce detachment as one of the main problems for *(BPR)* and hence for the Normativity of Rationality in general. This will be discussed further in Chapters 2, 6 and 8. For now, it should suffice to get clear on the structure of the problem, as it will reappear later on.

A famous argument against *(BPR)* is based on a particular understanding of the logical structure of rational requirements. It has been observed on many occasions that requirements like 'You are rationally required to not believe p and not-p' or 'You are rationally required to intend p, if you intend q and you believe that p is a necessary means for q-ing' are ambiguous: it is not clear which scope the *requires*-operator takes.[4] On a natural understanding that follows everyday language use, the operator takes *narrow scope* – it only governs the consequent of the conditional:

> *(No Contradictions-NS):* If you believe that p, rationality requires of you that [you do not believe that not-p].
>
> *(Means-End Coherence-NS):* If you intend to e, and you believe that intending to m is necessary for e-ing, rationality requires of you that [you intend to m].

Despite its intuitiveness, this understanding of rational requirements creates problems for *(BPR)*. Remember that if *(BPR)* is to be true, it needs to be the case that whenever a rational requirement like the ones above applies, we have a reason to satisfy it; to carry out the required action. But as we will see, this is not the case with narrow scope requirements because they allow for bootstrapping, which is one instance of a more general problem known as detachment.

Because of their structure, we can apply modus ponens to narrow scope requirements and detach the consequent, given the truth of the antecedent. So if an agent believes p and we accept *(No Contradictions-NS)*, it is true that the agent is rationally *required* to not believe not-p. Combining this with *(BPR)* then makes it the case that the agent has a *reason* to not believe not-p. This all sounds plausible. If I believe that today is Wednesday, it seems just right that I am rationally required to not believe that today is Thursday. It also seems right that I have a reason to not believe that today is Thursday. But there are many problem cases in which these consequences of narrow scope requirements seem less plausible, if not outright false.

In the domain of theoretical rationality, we see that narrow scope requirements tend to make beliefs self-justifying. A recent discussion of this problem is provided by Kolodny (2005).[5] He starts with a plausible putative rational requirement to believe what is entailed by one's beliefs. The narrow scope reading would be 'If you believe that p, rationality requires you to believe what p entails'. Now, amongst other things, p obviously entails p. So one is rationally required

[4] For a more detailed discussion of scope, see Section 2.1.3.
[5] Previous occurrences include Gensler (1985) and Greenspan (1975).

to believe p if one believes p. If one satisfies the antecedent of the requirement, i.e. if one does in fact believe p, the narrow scope requirement allows for detachment of the consequent. One is now required to believe p. Furthermore, if we also accept *(BPR)*, one has a reason to believe p if one believes p. But this seems absurd. Why should I be rationally required to believe that the earth is flat just because I believe that the earth is flat? What is more, it does not have to be the case that I have a reason to believe the earth is flat if I believe the earth is flat. It seems more plausible that I do *not* have a reason to believe this, since all the available evidence speaks against it. And even if I do have a reason for it, say, because my otherwise reliable teacher told me so, this is not the reason that features in the above example. That reason is provided merely by me believing that the earth is flat (in combination with the narrow scope requirement and *(BPR)*).

But the problems of narrow scope requirements are not limited to theoretical rationality. Requirements of practical rationality also give rise to problem cases, if they take narrow scope. For example, take a means-end coherence requirement like the one above. Again, the problem is that the narrow scope formulation not only lets intentions and means-end beliefs go through unchecked, it also 'ennobles' them by turning them into rationally required attitudes and, in combination with *(BPR)*, by providing a reason for them.

Imagine a case where someone intends to get a promotion and believes that a means to getting the promotion is to kill their competitor. Call it *(The Ruthless Careerist)*. Having these attitudes makes the antecedent of *(Means-End Coherence-NS)* true and we can now detach the consequent: our ruthless careerist is now rationally required to intend to kill her competitor. Again, in combination with *(BPR)*, she also has a reason to intend to kill her competitor, merely by having those attitudes. We can construct such cases not only for immoral means-end beliefs but also for atrocious ends, like in Muriel's case. She intends to become a famous axe-murderer and believes that a means to this end is to sharpen her blade. Now we can detach the consequent, and she is rationally required to sharpen her blade and, by *(BPR)*, also has a reason to do so.

These examples illustrate that the Normativity of Rationality faces serious problems. On the narrow scope conception of rational requirements, we end up with undesirable consequences. Because they allow for detachment via modus ponens, things are being rationally required that we do not want to be rationally required. Even worse, because of *(BPR)* we are also given reasons for attitudes in cases where we would normally deny that we have a reason. So detachment is a serious problem. One obvious solution would be to simply abandon *(BPR)*. But given the intuitive appeal of the idea that rationality is normative, and how it is connected to our practice of basing criticism on rationality attributions, this should only be a last resort. What is more, *(BPR)* seems to be otherwise well-supported.[6] And even if we abandoned *(BPR)*, only part of the problem would be solved. Narrow scope formulations still rationally require attitudes which we do not think should be required. So rather than giving up *(BPR)* right away, it seems like a

[6] See e.g. Raz (2005) and Southwood (2008).

better option to focus on the formulation of the requirements.

1.2.2 Saving *(BPR)*: The Wide Scope Solution to Detachment

Indeed, it has been argued that the narrow scope reading is the wrong way of cashing out the ambiguity of rational requirements. This reading might sit well with our everyday language use but, as we have seen, allows for detachment and the resulting problems. The alternative is to interpret the *requires*-operator as taking *wide scope* – as governing the entire conditional:

> *(No Contradictions)*: Rationality requires [that if you believe p, you do not believe not-p].
>
> *(Means-End Coherence)*: Rationality requires [that if you intend q, and you believe that p is a necessary means for q-ing, you intend to p].

Wide scope requirements no longer allow for detachment via modus ponens. They can be satisfied in multiple ways and thereby avoid the kind of detachment problems that we encountered so far. To see this, let us start with Kolodny's entailment requirement. On a wide scope reading, this is would be: 'Rationality requires that if you believe that p, you believe what p entails'. The conditional that is governed by the *requires*-operator is true if its antecedent is false or its consequent is true. So if an agent does believe that p, she can satisfy this requirement by either believing what p entails or by no longer believing that p. It is not the case that she is rationally required to believe what p entails just because she believes p. Consequently, *(BPR)* no longer automatically gives her a reason to believe p.

Detachment is blocked by the same mechanism in the case of *(Means-End Coherence)*. This requirement can be satisfied by either intending to p, no longer believing that p is a necessary means for q-ing, or no longer intending to q. So we do not end up being rationally required to intend to kill our competitors or to sharpen our blades. Dropping the belief that killing one's competitor is a necessary means for getting the promotion or no longer intending to become a mass murderer would equally satisfy the requirement. Because there no longer is a rational requirement to do just one of these options, *(BPR)* does not automatically give us a reason for each option. What is rationally required is a certain pattern of attitudes, rather than a single one: to believe what p entails or to not believe p; to not believe not-p or to not believe p; to not intend q or to not believe that p is a necessary means for q or to intend q. Therefore, the reason that *(BPR)* provides is a reason for instantiating these patterns of attitudes.

So, adopting the wide scope reading avoids detachment. And since detachment created a problem for *(BPR)*, it seems like the wide scope strategy is successful: if we adopt the wide scope reading, we can maintain the requirements that we thought plausible, and also their normativity, in the

form of *(BPR)*.[7] Unfortunately, it has been pointed out that even wide scope requirements are susceptible to detachment, thus facing the same problems as narrow scope requirements. It is still possible to detach reasons for individual attitudes if we accept some other, independent principles about how reasons are transmitted. The discussion of these principles and the mechanism that reintroduces trouble for rationality's normativity is subject of Chapter 6.

As we will see, this leaves us a with a number of different responses. If we want to uphold the Normativity of Rationality, we need to reject the principles that reintroduce detachment for wide scope requirements. At this point, my methodological call for a more fine-grained and explicit approach to debates about rationality becomes relevant again. We will see that a more nuanced understanding of these principles might actually reveal them as not very plausible from the outset.

Alternatively, we can maintain these principles and abandon the Normativity of Rationality. But we have already seen that this would come at a cost: we would lose our explanation for why we care so much about rationality attributions when we criticise agents. This would also take away some motivation from the first main theme of this thesis, i.e. the investigation of rationality attributions. Given this, I argue for another option in Chapter 6: to uphold the normativity of wide scope requirements by making the detachment problem less pressing.

Before eventually turning to my positive argument for the Normativity of Rationality in the final chapter, I spend some more time engaging with my methodological cause. As a case study of how a more precise and explicit approach can help to shed light on issues in debates about rationality, I explore how this pans out in the context of reasons for belief. As we will see, following this methodological approach leads to the result that pragmatic reasons could actually be reasons of the right kind for belief, in the sense that they bear on the (epistemic) rationality of such a belief. This result can supplement some of the strategies explored in the previous chapters and hence make detachment less of an issue for the Normativity of Rationality.

With a less threatening detachment in place, I finally turn to my positive argument – I propose to understand the Normativity of Rationality in terms of *commitment*.

Here, then, is the plan for the thesis: given the importance of rationality attributions for our practices of criticism and praise, the first half is concerned with the question of how to attribute rationality. I first turn to rational requirements for help (Chapter 2). In the course of investigating whether we can rely on requirements as a stepping stone for attributions, I discuss a number of choice points about the correct formulation of such requirements. Ultimately, we will see that the lack of consensus regarding these choice points means that requirements are not much help in settling the question of when we can attribute rationality. This motivates a change in strategy, which I turn to in Chapter 3. In this chapter, I directly engage with rationality

[7] To reiterate, what we have a reason for on the wide scope picture is a combination of attitudes expressed as a disjunction.

attributions and some of the controversies that surround them, to achieve a better understanding of the challenges to be met when making rationality attributions. Chapter 4 then develops my account of rationality attributions. Here, I argue that we should understand these as relative to specific subsets of an agent's attitudes. In addition, I also introduce a robustness measure and a mechanism for selecting relevant subsets of attitudes that reflects the particular question under discussion. Thanks to these features, my account is able to meet the challenges raised in Chapter 3. Chapter 5 further illustrates my account by contrasting it with a contextualist understanding of rationality attributions.

With this in place, the second half of the thesis is explicitly concerned with the Normativity of Rationality. In Chapter 6, I consider problems for a reasons-based understanding of the Normativity of Rationality that arise from so-called transmission principles. I conclude this chapter by pointing out various strategies by which we can soften the blow for rationality's normativity. Building on this, Chapter 7 provides further support for one of these strategies and also highlights the methodological undercurrent of this thesis. The final chapter then presents my defence of the Normativity of Rationality.

Chapter Two

Rational Requirements

Whether we can attribute rationality to an agent is crucial for whether we can subject them to criticism. But the preceding examples showed that deciding whether agents merit such attributions is no easy task. As a first pass, rational requirements could be a promising starting point. They are often used to bridge the gap from an agent's propositional attitudes to a decision about whether they should be attributed the property of rationality. In short, rational requirements can take us from attitudes to attributions. This idea, that rational requirements "determine when you are rational and when you are not rational" (Broome, 2013: 134), enjoys wide support and is most clearly stated in Broome (Broome, 2007c: 363): "Perhaps the most important question a system of rational requirements needs to settle is whether you are rational– have the property of rationality".[1] And his answer is as follows: "The proposition 'You are rational' is true at a world w if and only if all the requirements of rationality that apply to you at w are satisfied at w. [...] To say you are rational is to say you have the property of rationality" (Broome, 2007a: 362).

This suggests the following procedure. First, we check whether an agent complies with the relevant rational requirements. Second, we can then simply read off a rationality attribution, depending on whether the agent complies with the requirements. Endorsing this picture lets us shift the focus of inquiry away from attributions and onto requirements.

In order for requirements to play this crucial role, we need to know what these requirements are. This means that on this approach, we now need to inquire into the nature and form of rationality requirements, instead of inquiring into the nature and form of rationality attributions, and arrive

[1] This picture can also be traced in earlier and later works of Broome, e.g. Broome (1999: 410, 419), (2007b: 27) and (2013: 117). Some recent exceptions to this include Fogal (forthcoming: 3), who engages critically with this "requirements-based account". He rejects that agents are rational or irrational only in virtue of their compliance with rational requirements and argues that rational pressure, instead of rational requirements, is fundamental for rationality. In a similar vein, Fink (2018) contends that it is possible to take the opposite approach to Broome's picture. Instead of starting out with rational requirements in order to shed light on the property of rationality, we can start with the property of rationality itself to gain insights into the nature of rational requirements.

at rationality attributions in a second step. This leads us to a new problem. Unfortunately, there is no consensus on the definitive list of rational requirements – authors disagree about almost every choice point connected to rational requirements. In the next sections, I turn my attention to the various choice points we need to take a stand on in order to arrive at something that comes at least close to a definitive list of the things that rationality requires. The lesson to be drawn from this discussion is the following: the task of settling all questions related to finding the correct requirements is monumental. I ultimately abandon this strategy in favour of a more promising one. Instead of focussing on rational requirements in order to settle rationality attributions, I propose to approach things the other way around by focussing on the nature of rationality attributions directly. This way, we can maintain the promise of progress without getting caught up in the overwhelming complexity of the issues surrounding rational requirements. But for now, let us get these issues on the table. The choice points that I will be considering concern the nature (Section 2.1), the extent (Section 2.1.2), the scope (Section 2.1.3) and the temporal nature (Section 2.1.4) of rational requirements.

2.1 The Nature of Rational Requirements

In the search for the authoritative list of rational requirements, the first question to answer is *what* to look for. *What* do agents have to do in order to be attributed rationality? This seemingly obvious question actually leads to one of the most fundamental and dividing debates within the literature on rationality, which is dominated by two competing views about the *nature* of rational requirements. The intuitive attributions or reactions that we observed in connection with the three cases in the previous section can each be mapped onto one of them.

Criticising Sara for not intending to skip lunch, not taking issue with Sara dropping her intention to submit today and criticising Muriel for not intending to sharpen her blade all fall out of a picture that takes rational requirements to be *coherence* requirements. Being hesitant about criticising Sara for not intending to skip lunch, criticising Sara for dropping her intention to submit today and having reservations about criticising Muriel for her lack of blade-sharpening intentions sits well with a view that takes rational requirements to be *substantive* in nature. For a better understanding of these terms, I now deal with each view separately.

2.1.1 Coherence vs Substantive Requirements

What does rationality require of you? One side holds that rationality requires you to have your propositional attitudes in proper order. On this view, rationality is all about having coherent attitudes. So, the requirements of rationality are coherence requirements. I refer to this as 'the coherence view' throughout the thesis. The other side thinks that rationality requires you to have specific attitudes, instead of just patterns of attitudes, given a particular situation. This is most commonly understood in terms of reasons. What rationality requires, so the thought goes, is to respond correctly to reasons. The rational requirements endorsed by this view are substantive

requirements. I refer to this view as 'the reasons view'.[2] These views are best illustrated by appealing to examples. In a case like Sara's, where she lacks the intention to skip lunch, and given Broome's requirement-based picture, the coherence view does not attribute rationality to Sara, because she violates a plausible rational requirement:

> *(Means-End Coherence)* Rationality requires that [if you intend to e, and you believe that intending to m is necessary for e-ing, then you intend to m].[3]

The coherence view's requirement prohibits a pattern of attitudes. Whenever an agent intends the end, believes that intending the means is necessary for achieving the end and does not intend the means, they are displaying this prohibited pattern and thereby violate the requirement. This leaves us with three ways in which agents can comply with *(Means-End Coherence)*:[4]

i. intend the end, believe that intending the means is necessary for achieving the end, and intend the means

ii. not intend the end

iii. not believe that intending the means is necessary for achieving the end

Sara displays the prohibited pattern. She intends to submit her paper today, she believes that skipping lunch is necessary for this and yet she does not intend to skip lunch, as illustrated by her going for a lunch break instead.[5] And so she violates *(Means-End Coherence)*. All else being equal, this is enough for the coherence view to call Sara irrational.

The verdict of the reasons view is less straight-forward. Take a plausible rational requirement of the view:

> *(Reason-Responsiveness)* Rationality requires that you intend what you have decisive reason to intend.[6]

[2] Throughout, I sometimes use 'substantive requirements', 'reasons-responsiveness' or the 'reasons view' to refer to requirements of this kind. For my purposes, they can be used interchangeably.

[3] This formulation can be found in Brunero (2012: 127). Many different versions exist, e.g. in Lord (2014), Broome (2013), Way (2010a) and Schroeder (2004). The differences between them need not concern us here, since I take them to all express the same idea.

[4] One might have noticed that I have presented *(Means-End Coherence)* as a wide scope requirement without arguing for it, because it mainly serves illustrative purposes at this point. The choice between narrow and wide scope formulations and their connections to coherence and substantive requirements will be discussed in Sections 2.1.3 and 2.2.

[5] I am assuming here that going for lunch is sufficient evidence for the absence of an intention to skip lunch. If this is a point of disagreement, we can rephrase the original example in terms of intention: Sara's favourite Italian restaurant runs a pizza lunch special that day and so she does not have the intention to skip lunch.

[6] This requirement should be understood as expressing the main idea behind the reasons view, rather than as a fully developed and defensible requirement. Similar versions can be found in e.g. Kiesewetter (2017), Lord (2018) and Broome (2007a). They all come with details specific to the underlying accounts, which we can safely disregard at this point.

If we only take into account Sara's initial situation, then the reasons view might concur with the coherence view. Plausibly, she has decisive reason to intend the means necessary for submitting today and violates this requirement by lacking the intention to skip lunch. All else being equal, this makes her irrational. But if we also consider possible beliefs of Sara about lunch breaks taking priority over deadlines, the two views can come apart. The reasons view could now hold that Sara satisfies *(Reasons-Responsiveness)* by taking a lunch break, if she has decisive reason to intend to go for lunch, given her beliefs about the priority of lunch breaks.

We can observe similar disagreement between the views in the second case. We have seen that the coherence view can attribute rationality to Sara when she drops her intention to send her paper off today, because she would satisfy *(Means-End Coherence).* But the reasons view can plausibly hold that she is irrational because she violates *(Reasons-Responsiveness).* It is plausible that Sara has decisive reason to intend to send her paper off today: the deadline is important, meeting the deadline would be good for her career, she needs to move on to other projects etc. And if rationality attributions reflect an agent's performance regarding rational requirements, then Sara can be seen as irrational because of her violation of *(Reasons-Responsiveness).*

Muriel's case brings out yet another similar disagreement between the views. A defender of the coherence view would point to Muriel's violation of *(Means-End Coherence)* and then move to attributing irrationality to her. After all, she displays the prohibited pattern of attitudes by not intending the means believed necessary to her end. On a reasons view, we at least have the option to attribute rationality to Muriel, if we can argue that she satisfies, or at least does not violate, *(Reasons-Responsiveness).* And given how many reasons – moral, prudential, or other – speak against becoming an axe-murderer, Muriel plausibly does not have decisive reason to intend to take the means necessary for becoming one.[7]

This quick discussion of the three cases shows that the coherence view and the reasons view can and do come apart. This makes the question about the right view salient. If we are to rely on rational requirements to obtain rationality attributions, which of the views (with their respective set of requirements) should we adopt? In the next sections, we will see that both views have some appeal.

2.1.1.1 Motivating the Coherence View

Why should we adopt the coherence view? To many, linking rationality with coherence seems almost like a platitude. This is illustrated by the common practice of merely stating this link without providing a supporting argument. Broome, as one of the most famous coherence-theorists,

[7] The above remarks should not be taken as authoritative statements on what these agents have decisive reason to intend. Instead, they should be seen as pointing to possible positions, admittedly without much argument, to illustrate the possible divergence between the coherence and the reasons view. Whether the views actually diverge will depend on what the agent in question has decisive reason to intend. But the possibility of convergence between the views is not problematic for my purposes. All we need to establish is that they *can* come apart – not that they always do.

is a good example: "What rationality requires of you is proper order in your mind. It requires your mental states to be properly related to each other. That is to say, it requires your mind to be coherent in particular aspects" (Broome, 2013: 152). He is in the good company of Kolodny, one of his main interlocutors in the debate, who states that "[i]n each instance in which one is under a rational requirement, what it ought to require of one is to avoid or resolve some specific conflict among one's attitudes" (Kolodny, 2005: 516). Another example is Brunero, whose views will be of particular interest later: "The requirements of rationality specify conflicts concerning our attitudes which we should resolve (or avoid if we do not yet have them) if we are to be rational" (Brunero, 2012: 28).

We can find further examples of the assumed link between coherence and rationality, which focus on practical rationality and the means-end coherence requirement in particular. Broome (2005: 2) strongly connects *irrationality* to a failure to comply with this requirement: "A person is necessarily irrational if she does not intend whatever she believes is a necessary means to an end she intends". The same idea is expressed by Bratman (2009: 412), who links the requirement to a *rational demand*: "There is a rational demand that one's intentions be means-end coherent in the sense, roughly, that it not be true that one intends E, believes that E requires that one now intends necessary means M, and yet not now intends M". And similarly, Raz (2005: 18) establishes a link between *malfunctioning* and failure to satisfy (*Means-End Coherence)*: "People who fail to pursue the means to their ends display or manifest a form of malfunctioning criticisable as a form of irrationality".

Anyone not already convinced by the coherence view will not find these statements or platitudes helpful. Luckily, there are more substantial arguments in favour of the view.[8] In what follows, I focus on one coherence requirement in the practical domain – the by now familiar (*Means-End Coherence)* – and one coherence requirement in the theoretical domain:

> *(No Contradictions)*: Rationality requires of you that [if you believe that p, you do not believe that not-p].[9]

The arguments can be classified according to their *directness*. On one end of the spectrum, we have arguments that favour the coherence view indirectly, because of its desirable consequences. On the other end, we find arguments that support the coherence view directly, mostly by appeal to a close and intimate connection to agency. And some arguments fall in between.

Indirect arguments for the coherence view can be found in Reisner (2011) and Rippon (2011). Reisner's argument actually focusses on the theoretical case but he also considers how it could be extended to the practical case. In the theoretical case, Reisner thinks that we have very

[8] For a helpful overview, see Kiesewetter (2017).

[9] This is Way's (2010b: 1059) version. Many other versions have been proposed, see e.g. Kolodny (2008), Fogal (forthcoming) or Broome (2013). Again, the differences between the versions need not worry us here.

good epistemic reasons to be coherent. That is because being coherent increases our chances of believing truly. Take the *(No Contradictions)* requirement. We have strong epistemic reasons to satisfy this coherence requirement, since having contradictory beliefs makes it impossible for all our beliefs to be true. And since we want to have true beliefs, we have epistemic reasons to do whatever gets us closer to having true beliefs. Coherence requirements are one way of doing so, and so we have epistemic reasons to have coherent beliefs. In the practical domain, we can imagine a similar argument. By requiring coherence among our mental states, it could be argued that we are more likely to achieve our ends, which is something that we have practical reasons to do. Note that Reisner does not endorse this application to the domain of practical rationality. But the general pattern of this indirect argument for the coherence view should be clear: the coherence view is to be preferred because it is conducive to a goal for which we have strong epistemic or practical reasons.[10]

Rippon (2011) defends the coherence view against a number of objections, thereby concluding that it is plausible, since it can withstand such scrutiny. He outright accepts that *(Means-End Coherence)* comes with a reason: "people have reason to take the known, necessary means to the ends they intend" (ibid.: 1). But importantly, Rippon should not be understood as defending a reasons view here. The reason in question is "a wide-scope conditional reason [which] says that we have reason to make it the case that (if the antecedent is true, then the consequent is true)" (ibid.: 4), i.e. a reason to intend the means, if we intend the end and believe in the necessity of the means. This does not give us a reason for one specific attitude but rather a reason that "constrain[s] our attitudes and actions by governing combinations of them" (ibid.: 4), which is in keeping with the coherence view.

Moving further along the spectrum, Raz (2005) discusses an argument for the coherence view that is still indirect but already makes use of the notion of agency. The argument focuses on *(Means-End Coherence)* specifically. It assumes a strong connection between rationality and agency: "[r]ationality consists in part in proper functioning", where "proper functioning" is to be understood as "the standard of effective agency" (ibid.: 18). Being means-end incoherent means that an agent is not living up to this standard of effective agency, i.e. is not functioning properly. That is because "when people's mental processes fail to function properly with the result that they fail to take appropriate actions to facilitate realization of their ends, they are left with inconsistent – or at any rate incoherent – sets of beliefs, or beliefs and intentions", which in turn "reveals an executive defect in the agent, a limitation on his ability to be an effective agent" (ibid.: 19, 17). I take it that this line of reasoning can be extended to the theoretical realm as well. For example, if we think of beliefs as the necessary precursors to action, an agent

[10] This might sound like a defence of the normativity of rationality already. At this point, I am not concerned with this question. The first step to answering the normativity of rationality-question is to decide on a conception of rationality that can be evaluated for normativity in a second step. Now, it is true that the indirect arguments for the coherence view also represent one way of arguing for the normativity of rationality. They correspond to the reasons-version of the Normativity Claim (see Sections 1.2 and 8.2.2). But we cannot always lump the two steps together in this way. For the sake of clarity, I prefer treating them separately at first.

who has contradictory beliefs about which actions would facilitate their ends is unable to take the appropriate actions. On my reading of Raz's presentation of this argument, effective agency is an ideal of rational agency that we can sometimes achieve and should aim to achieve whenever we can. With this in place, we can see why the coherence view should be preferred: it lets us avoid the kind of malfunctioning that stands in the way of us being effective agents and us living up the the ideal of rational agency. But this is still an indirect argument: coherence is favoured because of one of its consequences, i.e. avoiding malfunctioning.

Bratman (2009) takes this line of thought further. He first provides another indirect argument for the coherence view. On his planning theory, the fact that our computational capacities are limited creates the need for us to make long-term partial plans that we fill in more precisely as we go along. This picture requires a stable and coherent set of mental attitudes, including both intentions and beliefs, in which "threats of means-end incoherence pose deliberative problems" (ibid.: 413). For example, suddenly dropping the intention to an end will jeopardise the filling in of an agent's partial plan that was based on that very intention. Similarly, holding contradictory beliefs about the long-term plan will make filling it in significantly more difficult. Coherence requirements let us avoid such deliberative problems by providing guidance to our practical deliberation and are "central to the proper functioning of planning in our agency" (ibid.: 413). This is Bratman's indirect argument – the coherence view is attractive because it sits well with and might even be required by his planning theory. But he goes further. He thinks that "there is a distinctive, noninstrumental practical reason" to be means-end coherent (and presumably also to not believe contradictions), namely "a reason of self-governance" (ibid.: 436). This is based on two assumptions. One, that "we have an intrinsic reason to govern our lives" and two, that satisfying *(Means-End Coherence)* "is a necessary constitutive element in the corresponding self-governance of planning agents like us" (ibid.: 432). And so whenever the possibility of self-governance is not excluded from the outset, the coherence view is to be favoured since it features so predominantly in the notion of self-governance that applies to us.

At the very end of the spectrum, we find Southwood's (2008) direct argument. The argument builds on proper functioning arguments along the lines of Bratman (2009) but supplements them with a first-personal perspective that Southwood thinks is often unjustly neglected. The argument starts with the idea of a "first-personal standpoint", which corresponds to the mental make-up of an agent, i.e. their beliefs, intentions, desires, values and so on.[11] Southwood then argues that being subject to the demands of a first-personal standpoint is (in part) constitutive of having such a standpoint. And the demands of a first-personal standpoint are coherence requirements, like intending the necessary means to your intended ends, or not believing contradictions. So in virtue of having a first-personal standpoint, agents are subject to such demands. He then assumes that "the particular kind of agency we possess" involves having a first-personal standpoint (Southwood, 2008: 29). And so we can conclude that the coherence view is correct – it falls out of the nature

[11] This already touches on my solution to the Normativity Problem that I will develop in Chapter 8.

of our agency, and in particular out of the demands of the first-personal standpoint that this involves.

A final source of support for the coherence view is its intuitiveness. As Kolodny (2008: 437) puts it, "[t]he intuitive idea is that formally incoherent attitudes give rise to a certain normative tension, or exert a kind of rational pressure on one another, and this tension, or pressure, is relieved, just when one of the attitudes is revised." There just seems to be something very odd about an agent who wants to have dinner but has no intention of eating food; or someone who believes that trees have leaves but also believes that they do not have leaves. These incoherencies are intuitively incompatible with an agent's rationality and should be ruled out by its requirements, as far as the intuitive story behind coherence requirements is concerned.

2.1.1.2 Motivating the Reasons View

But the coherence view is not the only well-supported view. Taking the the reasons view to be correct is often considered a platitude about rationality as well. One example is Nozick (1993: 107), who holds that "[t]he rationality of a belief or action is a matter of its responsiveness to the reasons for and against, and of the process by which those reasons are generated". A more general statement of this platitude can be found in Raz (1999: 68): "An account of rationality is an account of the capacity to perceive reasons and to conform to them". And as a last example, take Gibbard (1990: 161), who thinks that "[t]o say it is rational to do something is to say that doing it is supported by the preponderance of reasons".

But just like with the coherence view, we are not restricted to mere platitudes. Various arguments have been advanced in favour of the reasons view, too. It should be noted that these arguments often aim to support one specific version of the reasons view, e.g. that rationality consists in responding correctly to *possessed* reasons, or *objective* reasons, or *subjective* reasons.[12] I will not discuss the merits of these different versions here. For my purposes, it suffices to keep in mind the general idea: that rationality consists in responding correctly to reasons, where *(Reasons-Responsiveness)* is an application of this general idea to practical rationality. I will discuss three arguments in favour of the reasons view: an argument that relies on theoretical virtues, one that appeals to both theoretical virtues and normativity, and one that argues directly from normativity.

In his recent book, Lord (2018) defends a reasons-based account of rationality. To him, "[i]t's clear: Sometimes rationality requires me to have particular attitudes and perform particular actions" (ibid.: 18). Remember that this is not the case on the coherence view.[13] If rational requirements are coherence requirements, then they only ever require or prohibit patterns of

[12] For discussion of these variations, see e.g. Lord (2018).

[13] To be exact, the coherence view can sometimes require particular responses of agents, e.g. when its requirements are formulated as having narrow scope (see Section 2.1.3) or in specific cases where some of the available options are ruled out due to external circumstances (see Setiya, 2007; Greenspan, 1975). But it is primarily concerned with patterns, rather than specific attitudes.

attitudes, and not specific attitudes. But cases like Sara's and Muriel's seem to motivate the need for rational requirements to require specific attitudes or their absence, e.g. Sara's intention to skip lunch and the absence of Muriel's intention to sharpen her axe. Cases like this provide strong intuitive support for a reasons view. Building on this, Lord's argument can be understood as claiming that in addition to doing justice to such cases, the reasons view has the additional advantage of being more *fundamental*. That is because the coherence view can ultimately be reduced to the reasons view. Every coherence requirement can be translated into a corresponding reasons requirement. And every instance of incoherence can ultimately be traced back to a lack of reasons-responsiveness on the side of the agent. So the reasons view can account for all cases that the coherence-account can account for, and some more: it can also account for cases where we are in fact required to hold certain attitudes.[14] Since the reasons view is more fundamental and also more explanatory forceful, it should be preferred.

Kiesewetter's (2017) preference for the reasons view is similarly rooted in weighing up the theoretical advantages of the two views. He discusses the coherence view at length, only to conclude that it is incompatible with a central claim about rationality: that rational requirements are normative.[15] Giving up this claim comes with a lot of theoretical costs for Kiesewetter, since he thinks that "our practice of rational criticism essentially involves the assumption that rationality is normative" and that giving up on this would commit us "to a radical error theory about ordinary attributions of irrationality" (ibid.: 160, 43). So if there is a view that can easily account for the normativity of rational requirements, it ought to be preferred, since we would not have to unnecessarily overhaul our practices of rational criticism. One example of a feature of the coherence view that is widely taken to be incompatible with the normativity of rational requirements is the so-called symmetry objection (e.g. Schroeder, 2004).[16] This objection will play an important role in later sections, too (see Section 3.2.1).

For now, it suffices to understand the gist of it. Recall Sara's case. A coherence view requirement like *(Means-End Coherence)* allows for three different ways of compliance. That is because the relevant operator in such requirements (here: 'requires') governs the entire conditional, which means that there are three different ways of making the conditional true: Sara can intend to skip lunch (which would make both antecedent and consequent true); she can drop the belief that skipping lunch is necessary for submitting today (which would make the antecedent false); and she can drop her intention to submit today (which would make the antecedent false, too). These options are *symmetrical* in that they all lead to satisfaction of the requirement. Precisely this feature is seen as problematic for the claim that rational requirements are normative. It has been noted that the three options differ in rationality. Sometimes, it might seem more rational to drop

[14] This result could also be achieved on a coherence view, but only at the cost of endorsing the controversial narrow-scope formulation of rational requirements. This formulation is widely regarded as untenable since it leads to problematic bootstrapping; see Sections 2.1.3, 1.2.1.

[15] This point will also be addressed in my discussion of scope in Section 2.1.3.

[16] Kiesewetter discusses a number of other features, such as bootstrapping, detachment, lack of guidance and Kolodny's (2005) Why be rational? challenge, which will also be dealt with in Chapter 8.

the end rather than the belief about the means-end relation (e.g. Bedke, 2009), if for example this belief is well-supported by evidence. At other times, it might seem most rational to hold on to our initial intention and to just see it through, i.e. to intend the means. Dropping the intention to submit today could be seen as the result of laziness and of rationalising one's way out of an uncomfortable task. In any case, it is widely accepted that the options sometimes differ with regard to their rationality – in other words, they are asymmetrical. For example, we sometimes want to and do criticise people for being irrational if they flippantly drop their intentions. But this criticism is not available to us on the coherence view. And since the normativity of rational requirements is intimately bound up with our practice of rational criticism, the coherence view is not entirely compatible with Kiesewetter's normativity claim. The reasons view, in contrast, can allow for rational criticism of this kind: it is easy to argue that Sara does not intend what she has decisive reason to intend if she drops her intention to submit her paper by the end of the day. What is more, it also gives us an easy explanation of the normativity of rational requirements: it is the normativity of reasons, with which we are familiar.[17] And so it is to be preferred to the coherence view.

The last argument I want to mention invokes normativity explicitly. To be clear, Kiesewetter does so as well, but in the larger context of showing that the reasons view has theoretical advantages over the coherence view, where upholding his normativity claim is one among other theoretical advantages. Raz (1999) prefers the reasons view to the coherence view because only the former guarantees that agents apprehend the normative features in the world – "[t]he core idea is that rationality is the ability to realize the normative significance of the normative features of the world, and the ability to respond accordingly" (ibid.: 68). And for Raz, "the normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons" (ibid.: 67). If rationality is the ability to recognise normativity and if normativity amounts to reasons, an account of rationality that does not closely connect it to reasons is simply inadequate.

Moreover, it might even amount to a reductio of the coherence view. The mere fact that we have to engage with the question whether there are reasons to be coherence-rational, and the apparent difficulty of arguing that there (always) are might be taken to indicate that the coherence view is simply false (see Broome, 2013; Kolodny, 2005). But note that there are a number of different interpretations available here. For one, one might instead reject the Normativity of Rationality. But a large part of this thesis is dedicated to avoiding precisely this. Alternatively, one might want to broaden the understanding of 'normativity' at work here. And indeed, this is already foreshadowing my defence of the Normativity of Rationality as developed in Chapter 8. For the time being, let us grant Raz his understanding of 'normativity' as reasons, and 'rationality' as the

[17] Again, this corresponds to the reasons-version of the Normativity Claim, which is not uncontroversial and will be discussed in more detail in Chapter 8. What is more, if one accepts that "[w]hat it is to be rational is to correctly respond to reasons", as Lord (2014: 155) puts it in his "Determination Principle", there is not even a need for a principle like the Normativity Claim that links rationality to reasons, because rationality is already understood in terms of reasons.

ability to recognise these. It then makes perfect sense to consider "the accounts of normativity and of reason and rationality, though not identical, [to be] inter-related. An account of rationality is an account of the capacity to perceive reasons and to conform to them, and of different forms of conforming to reasons, and their appropriateness in different contexts" (Raz, 1999: 68). Raz's understanding seems especially attractive in cases that involve not only rational but also moral considerations. The perceived mismatch between the requirements of rationality and morality, when an aspiring axe-murderer like Muriel is required to sharpen her axe, disappears when we adopt this broad understanding of reasons and their normativity. On Raz's account, moral considerations might well contribute to or constitute an agent's reasons. This allows for the possibility of Muriel not being required to sharpen her axe, since she does not have decisive reason to do so.

This discussion has shown that both the coherence and the reasons view have good and important arguments that speak in their favour. After having considered only one of the four choice points, we can already see that there are serious and deep disagreements about rational requirements. It seems like we cannot even agree on *what* they require. With an eye to our initial question of how to decide on the rationality of agents, this means that we have not made a lot of progress. Rational requirements seemed like a good pathway to rationality attributions. But if it is unclear which set of requirements, endorsed by either the coherence view or the reasons view, we should rely on, then maybe they are not much help in providing rationality attributions after all. However, let us have a look at the remaining three choice points.

2.1.2 Local vs Global Requirements

Unfortunately, the issues surrounding rational requirements do not stop here. The second choice point is also surrounded by controversy. Fogal (2018: 32) calls this choice point the "extent" of rational requirements. Even if we know *what* rational requirements ask of us, e.g. to respond correctly to reasons or to be coherent, we also need to know *where* they make these demands. Do they govern particular subsets of the agent's attitudes or their entirety? These are the two options on the table: are rational requirements *local* or *global*?

Unlike some of the other choice points to be discussed in this section, this one has received fairly little attention in the literature – with the works cited here as notable exceptions. But since it will be crucial in my later discussion of rationality attributions, and for the account of such attributions that I ultimately propose, I want to spend some time on spelling out the distinction at the level of requirements.

Starting with the idea of *local* requirements, Kolodny (2005) characterises local rational requirements as governing specific conflicts amongst attitudes. An example of such a specific conflict would be an agent who believes both p and not-p. The proponent of local rational requirements would argue that conflicts of this kind, which only involve the subset of the agent's attitudes that contains the conflicting beliefs, are precisely what rational requirements should

rule out (as done by *(No Contradictions)*). As Kolodny puts it, "we typically identify rational requirements by first identifying conflict-states to be avoided or resolved. This is, in essence, what it means to say that rational requirements are 'local' " (Kolodny, 2005: 518). Instead of making demands on all of the agent's attitudes, local requirements only deal with smaller 'pockets', so to speak, of an agent's attitude set. Kolodny illustrates this idea further by comparing local requirements to "a referee with authority over a different part of a playing field, or to an inspector with authority over a different stage in a production process" (ibid.: 516).

Brunero (2010; 2012) endorses this characterisation of local requirements for the most part, but seems to be a bit less concerned with conflict states. Whilst he adopts Kolodny's characterisation in terms of "conflict of attitude-states" at first, he later refers to the extent of rational requirements as a "specific attitude-state" more generally (Brunero, 2010: 33). I take this to mean that local requirements need not be silent about attitude sets that are not plagued by outright conflict just yet. For example, this characterisation would allow local requirements to require that an agent believe the consequences of the beliefs in the subset it governs. But this is just a minor point, especially when we adopt a broader reading of Kolodny's use of "conflict states". It is in keeping with his points and examples to understand this as not only involving outright contradictory attitudes, but also broader inconsistencies and maybe also attitude states that are conflicting because of a lack of attitudes, like an agent not believing q, despite believing that p and that if p, then q. More importantly, both characterisations are structurally the same. The hallmark of a local requirement is that it is only concerned with a specific subset of an agent's attitudes.[18]

With this characterisation of local requirements in place, let us move on to their *global* counterpart. If local requirements govern specific subsets of attitudes, then global requirements govern the entire set of an agent's attitudes. But this seemingly simple idea can be understood in two different ways. On the one hand, 'governing the entire set of attitudes' might be taken quite literally to refer to a requirement that makes a demand on all of an agent's attitude at once, like a referee that has authority over the entire playing field. Call this the 'literal-global' understanding. On the other hand, we might understand this idea of global governance as the result of somehow combining individual local requirements into one overall requirement. Call this the 'aggregated-global' understanding. Moving forward, I will mostly rely on this second, aggregated-global understanding for two reasons: I take it that it is most commonly adopted (albeit rarely explicitly), and it underwrites Brunero's work, which I draw on extensively in this and the next chapter. But before I spell this out further, let me quickly address the competing literal-global understanding which I cast aside. Getting clearer on this alternative will improve our grasp of the aggregated-global understanding we are taking forward.

Examples of literal-global requirements are difficult to come by. Presumably, they could take the

[18] Broome (2007a), (2013) also uses this terminology, albeit in a slightly different way. For him, claiming that requirements are local is to be understood more as saying that what they require depends on the actual world or the actual situation of the agent, rather than some ideal. This understanding is in line with his semantics for requirements, which I do not engage with here.

form of

> *(Global Coherence)* Rationality requires that you have coherent attitudes.

This requirement makes a demand on the entirety of an agent's attitude set (namely, it requires that this set be coherent), which makes it a global requirement according to the first understanding of 'global'. But one might note right away that this does not seem a very appealing requirement. For one, the situations, or conflicts, that it prohibits would still be more local in nature. Whenever agents have incoherent attitudes, this is due to some specific attitudes that are conflicting. This seems at odds with the global nature of the requirement. And even if we were not overly troubled by this mismatch, it would also mean that we could encounter conflicts amongst different ways of satisfying the requirement. For example, dropping a belief that q might avoid being incoherent with regard to also believing not-q, but might introduce a new incoherence, e.g. if the agent also holds the beliefs that p, and that if p, then q. As it stands, *(Global Coherence)* could not say anything about this, other than that the agent flouts a rational requirement either way.

This is problematic, for two related reasons. One, as will be discussed in Section 2.1.3.1, we might want our rational requirements to be action-guiding, that is, to tell us how we can improve our rationality. The 'damned if you do, damned if you don't'-verdict that *(Global Coherence)* yields is of no help here. Two, we might also think that this is the incorrect verdict. It might be that all things considered, it would make more sense for the agent to maintain the belief that q. But in order to reflect this, we would need to make use of some sort of all-things-considered-judgment. However, on this literal-global understanding as simply ranging over all attitudes, there is no available weighing procedure that would yield such an all-things-considered judgment. The importance of being able to appeal to such judgments further motivates the aggregated-global understanding over the literal-global understanding.

But at this point, one might think that I have not given careful enough consideration to the literal-global understanding I cast aside. After all, I have only considered one potential requirement. And indeed, it seems like once we adopt a reasons-responsiveness understanding of rationality, such literal-global requirements seem less odd. *(Reasons-Responsiveness)* might be understood along these lines, even though it only governs intentions. A truly literal-global reasons-responsiveness requirement could look like:

> *(Global Reason)* Rationality requires you to have the attitudes that you have decisive reason to have.

Again, this requirement would make a demand on all of an agent's attitudes. And it has the advantage that we avoid situations that seemed problematic for *(Global Coherence)* because all-things-considered judgments are already built into the requirement, via the appeal to 'decisive

reason'. Nevertheless, I set this literal-global understanding to the side, the reason being that I want to be able to operate within a coherence framework for rationality, or at least be able to be neutral about the correct rationality framework. If global requirements were only compatible with the reasons-responsiveness framework, this would substantially limit my options. The other, aggregated-global understanding of global requirements does not entail such a commitment to one rationality framework over the other.[19]

The case for understanding global requirements in this aggregated way is further bolstered by the fact that we actually have examples in the literature that we can appeal to here. On this understanding of global requirements, we need to somehow aggregate the existing local requirements in order to arrive at global judgments. In other words, we have to combine the individual pockets; we have to reconcile the verdicts of the individual referees. This picture is endorsed by Brunero (2012), who understands global or ""overall" (or "all-attitudes-considered") [requirements] [to] specify what's rational in light of the totality of an agent's attitudes" (ibid.: 129).[20]

He suggests to understand local and global requirements and their connection in parallel to Rossian prima facie duties and duties proper: Ross (1930) introduces the notion of prima facie duties to deal with competing moral demands.[21] For example, I have promised my friend to meet her at 5pm but close to five, I receive a call from another friend, asking me to drive her to the train station so that she can make her 5pm train. I am now under two competing moral demands: on the one hand, I should keep my promise. On the other hand, I should help a friend in need. Ross thinks that these competing demands are not "duties proper" but rather "prima facie duties". Prima facie duties are would-be duties proper, if it were not for another morally salient feature of the situation – if my other friend was not in need of a lift to the train station, keeping my promise would be a duty proper.

When it comes to deciding what we are morally required to do in this situation, we have to weigh the competing prima facie duties against each other and decide which one takes priority – keeping a promise or helping a friend in need. Here is how Ross describes this competition of duties: "when I am in a situation, as perhaps I always am, in which more than one of these prima facie duties is incumbent on me, what I have to do is to study the situation as fully as I can until I form the considered opinion (it is never more) that in the circumstances one of them is more incumbent than any other; then I am bound to think that to do this prima facie duty

[19] I investigate the connections between these choice points in Section 2.2.

[20] Brunero casts this point in terms of rationality *judgments* rather than *requirements* but he moves seamlessly between the two. Given the close connection between rationality attributions (or judgments) and rational requirements that we established in Chapter 1, I think this practice is unproblematic.

[21] There are a number of problems with this analogy. For one, Brunero (2012) does not develop it consistently. On p. 129 he likens local rationality *judgments* to prima facie duties, whereas in a footnote on p. 133, he states that local rationality *requirements* are analogous to such duties. The latter would move Brunero closer to Kolodny, who claims that rational requirements, as well as attributions, are local. What matters for our purposes here is just the difference between the local and global perspectives and that they could lead to different results.

is my duty sans phrase in the situation" (ibid.: 19). The picture that Kolodny describes and that Brunero endorses runs in parallel. Local rational requirements issue demands for subsets of attitudes. To get to an aggregated-global requirement that takes into account all attitudes, the (potentially conflicting) demands of the local requirements have to be weighed against each other. As Kolodny puts it, whilst being sceptical of it, "[w]hat rationality requires of one 'all things considered,' as it were, is some function of all of these local requirements: presumably, what will maximize one's overall satisfaction of these various local requirements, perhaps with weightings for more and less important local requirements" (Kolodny, 2005: 516).

Reisner (2013: 440) suggests that the 'function' or aggregation procedure which takes us from local to aggregated-global requirements could be understood as a supervenience relation:

> "It is tempting to think that a global judgement of an agent's rationality supervenes in a straightforward way on the satisfaction of all of her local rational requirements. The idea is roughly that A is fully rational if she has satisfied all of her local rational requirements. Alternatively, R is a local rational requirement for A, only if A is not fully rational unless she satisfies R".[22]

This should give us a good enough understanding of the distinction between local and aggregated-global requirements. Going forward, I will use 'global' and 'aggregated-global' synonymously, since I have rejected the competing literal-global understanding. But we still have to take a stance on this choice point. What is the extent of rational requirements? Are they local or global?

2.1.2.1 Motivating Local Requirements

Kolodny, as perhaps the most vocal defender of local requirements, puts forward some brief arguments in their favour. The first argument consists of two main premises. One, Kolodny thinks that rational requirements should reflect our ordinary rationality judgments. Two, he claims that our ordinary rationality judgments are local. And from this he concludes that rational requirements should be local too:

> "Our ordinary judgements about rationality have [...] important features, which rational requirements should reflect. First, these judgements are local. They are focused on specific conflicts among one's attitudes. We might judge, for example, that a person is being weak-willed in believing that he has conclusive reason to X, but not intending to X. And we might judge, at the same time, that he is giving in

[22] Since this is not the main concern of his paper, Reisner merely mentions this as a possibility without endorsing it or arguing for it. Fogal (forthcoming) also engages with this idea but ultimately rejects supervenience as the correct relation between the local and the global level in his framework.

> to wishful thinking in believing that he has conclusive reason to X, having himself decided that the evidence for that belief is flimsy" (Kolodny, 2005: 515-516).

The success of this argument clearly depends on whether its premises are plausible, i.e. whether our ordinary rationality judgments really are local in this way, and whether they are good guides when settling the questions of correct nature and form of rational requirements. Especially the first premise will be discussed at length in later sections (see e.g. Section 3.2.1). But for now, let us grant Kolodny this assumption.[23] He then moves on to consider the alternative – global rational requirements, conceived as the result of some aggregation of local requirements – and concludes that these are inferior because (i) they do not reflect our ordinary judgments, (ii) they might be superfluous or (iii) simply not feasible. (i) will be addressed later in my discussion (see Section 3.2.1). (ii) and (iii) have to be understood with Kolodny's first pass at aggregating local requirements in mind. On a picture where local requirements co-exist and are simply lumped together to make a global requirement, this global requirement might indeed be superfluous. In cases where the individual local requirements do not conflict, the global requirement presumably would only require that an agent satisfy as many local requirements as possible.[24] It would not say anything about the substantial demands that rationality makes on agents, over and above what is already contained in the local requirements. And in cases where such a global requirement could actually provide new insights into the demands of rationality, i.e. in cases where the local requirements conflict, Kolodny is sceptical that we could even construct such a global requirement, since this would require a "principle of triage, determining which from among conflicting requirements I ought to satisfy. It isn't clear to me that there is, or needs to be, a determinate answer to the question of which from among conflicting requirements one ought to satisfy" (ibid.: 516).

But not everyone is as convinced of the uselessness of global requirements. Brunero (2012) allows for both global and local requirements. Being able to appeal to the two different levels plays an important role for him. It allows him to reconcile conflicting verdicts in response to symmetry-objection-like scenarios. For example, Sara can satisfy the local *(Means-End Coherence)* by dropping her belief that skipping lunch is necessary for submitting her paper tonight, and by dropping her intention to submit the paper tonight. Those that find the symmetry objection persuasive would argue that it is implausible that both these ways of satisfying *(Means-End Coherence)* are equally rational. They are inclined to think that surely it is more rational to drop the intention rather than the belief.[25] This is where the appeal to global requirements can be helpful, according to Brunero: "It's perfectly consistent to claim both that (1) as far as

[23] For what it is worth, I think that this assumption is also shared by Brunero (2010; 2012).

[24] There is a related thought here that takes the number of satisfied (local) requirements to be indicators of degrees of rationality – the more requirements satisfied, the more rational. I take it that both Fink (2012) and Cherniak (1986) endorse this picture. I do not think that we can understand degrees of rationality this way, for reasons specific to my account of rationaltiy attributions. For more discussion of this, see Section 4.2.

[25] Bedke (2009) is an example.

[*(Means-End Coherence)*] goes, giving up an end is just as rational as giving up the belief, and (2) most often, all-attitudes-considered, it will be more rational to give up the end than give up the belief" (ibid.: 130). We can check an agent's compliance with an individual local requirement like *(Means-End Coherence)* that is only concerned with a small subset of attitudes to evaluate them for rationality, but we can also check their compliance with a global requirement, which groups together a number of local requirements, to evaluate them. The global perspective could include *(Means-End Coherence)* but also a requirement asking agents to conform their beliefs to the evidence, which would open up the possibility of holding that giving up the belief is less rational than giving up the means. Keeping both local and global requirements in the picture enables us to make these (no longer inconsistent) statements, which is reason enough for Brunero to also allow for global requirements in an account of rationality.

Some even go further and seem to prefer global to local requirements. Brunero mentions Schroeder (2004) as an example. Schroeder contends that it should never be considered rational if an agent simply drops their means-end belief, absent any reason to do so other than the fact that this would lead to complying with a rational requirement. This is an unacceptable outcome, even if we keep in mind that it is only located at the local level and might well change once we consider the global level. Here's how Schroeder puts the point:

> "So we don't get a problem [...] merely by noticing that so far as it says, it may be rational for [Sara] to change [her] belief. The problem [...] is that if [Sara] does respond to [her] situation in this way, she has to allow that though [Sara] is being epistemically irrational, [she] is in fact behaving impeccably, when it comes to subjective instrumental rationality. And that is a bizarre thing to say. Surely a good account of subjective instrumental rationality should not tell us that so far as instrumental rationality goes, this kind of behavior is okay" (ibid.: 346).

Allowing only for global requirements rules out these 'bizarre' things, since it would take into account a number of local requirements, e.g. the ones concerned with epistemic and instrumental rationality. And it would issue one unified demand that would deem the behaviour in question, e.g. dropping a means-end belief, as either rational or irrational, depending on whether it would violate the global requirement.

It does not strike me a 'bizarre' thing to say that an action that would be deemed epistemically irrational could be instrumentally rational and so I am hesitant to accept Schroeder's criticism. But for now, let us simply note Schroeder's criticism, as it will reappear in my discussion of symmetry (see Section 3.2.1). This section has shown that the disagreements surrounding rational requirements are even more numerous. Not only is it not obvious *what* they require, as illustrated by the discussion of the choice point regarding the nature of rational requirements, it is also unclear *where* they do so.

2.1.3 Wide vs Narrow Scope

We started this chapter by turning to rational requirements in the hope that these would be good guides to the rationality attributions we need for our practices of praising and blaming agents. But the choice points of nature and extent have turned out to be open questions. I now turn to the third choice point which unfortunately will also not be settled definitively. This choice point is concerned with the *structure*, or *scope* of rational requirements.

In particular, the question at issue here is whether rational requirements take narrow or wide scope. This point is easily conflated with the question of their extent, as dealt with in the previous section. But this should be resisted. Whilst the question of extent was concerned with where rational requirements apply, the question of scope is potentially more far-reaching. It can be interpreted as a question merely about form, or structure, of rational requirements, but can also be seen as a crossroads for more substantial commitments, such as the choice between a coherence and a reasons-responsiveness conception of rationality, as discussed in Section 2.2. But before we engage with these questions, let me explore the three different ways in which one can understand the scope-question.

One, asking about the scope of rational requirements can be understood as asking about the conditionality of these requirements (Broome, 2007b). On a narrow scope reading, the *existence* of rational requirements turns out to be conditional. On a wide scope reading, it is the *content* of rational requirements that is conditional. Specifically, the choice between narrow and wide scope applies to the 'rationality requires' operator. If 'rationality requires' ranges over the entire conditional, we are facing a wide scope requirement. If 'rationality requires' only ranges over the consequent, we are dealing with a narrow scope requirement. We can now see that the earlier *(No Contradictions)* requirement is in fact a wide scope requirement:

> *(No Contradictions)*: Rationality requires of you that [if you believe that p, you do not believe that not-p].

The conditionality of this requirement lies in its content – what rationality requires of you is to satisfy a conditional. Compare this to a possible narrow scope version of the same requirement:

> *(No Contradictions-NS)*: If you believe that p, rationality requires of you that [you do not believe that not-p].

Now, what rationality requires of you is merely that you not believe that not-p. The content of the requirement is no longer conditional. Instead, the conditionality has moved to the existence of the requirement – it is only in place if you satisfy the antecedent, i.e. you actually believe that p. If you have no opinion about the matter, you are not subject to a rational requirement governing your stance towards p or not-p. Broome (ibid.) compares this to a law requiring you to

drive on the left in Britain. On a narrow scope reading of this, you are only required to drive on the left if you are actually in Britain. If you are not in Britain, the law simply does not apply to you. In contrast, the wide scope reading of the law would place you under a conditional demand – if you are in Britain, drive on the left – where this demand obtains even if you are not in Britain right now. This brings out an important feature of narrow scope requirements, insofar as they are understood as being conditional in their existence. Since they are basically material conditionals, such requirements allow for *detachment* via modus ponens. Whenever the antecedent is given, we can detach the consequent, which contains the rational requirement. As we will see, this feature plays a crucial role in the debates about the scope of rational requirements, as it is the source of many objections raised against narrow scope formulations.

Two, the choice between narrow and wide scope can also be interpreted as the choice about what rational requirements ask of us, or how we can comply with them (Brunero, 2010; Brunero, 2012). Narrow scope requirements require a specific attitude and hence can only be satisfied in one specific way. For example, the only way to comply with *(No Contradictions-NS)* is to not believe not-p. In contrast, wide scope requirements do not require specific attitudes, but rather *patterns* of attitudes. This also means that wide scope requirements give agents options. There is usually more than one way of complying with the requirement. In our example, the requirement would be satisfied by either not believing not-p or by no longer believing p. This feature of the scope debate also suggests a natural alignment between a narrow scope view about rational requirements and a reasons-responsiveness conception of rationality on the one side, and between a wide scope view and a coherence conception. If rationality is largely a matter of having coherent attitudes, understanding rational requirements as prescribing or prohibiting patterns of attitudes seems fitting. But if rationality is about responding correctly to reasons, then it seems to make sense to think of requirements as demanding specific attitudes (e.g. the one that is supported by reasons).

This gives us an additional reason to be careful to keep apart the local/global and wide/narrow scope debates (or, to use the choice point terminology, the questions of extent and scope). The wide/narrow scope debate seems to entail a commitment to a view about the nature of rationality, whereas the local/global debate does not immediately do so. But let me note that this seemingly natural alignment should not be taken too seriously, as will be discussed in Section 2.2.3.

Fogal (2018: 37) refers to the question of whether rational requirements demand patterns or specific attitudes as "the philosophical issue" that is raised by the scope debate. He contrasts this with the linguistic issue, which is the third way of understanding the scope debate. The linguistic issue is about finding the correct interpretation of natural language uses of modal expressions like 'ought' or 'require' in conditionals. If approached from the linguistic point of view, the scope debate is no longer concerned with rationality or with what it requires. Instead, it captures different ways of providing a descriptively correct representation of these modal expressions in general. Rational requirements are only one field of application. As Fogal points out, the

non-linguistic, philosophical scope debate is the more important one, at least if we are interested in the notion of rationality. In that spirit, I will leave the linguistic issues aside from now on.

Now that we have a sufficient understanding of what is at stake when we decide on the scope of rational requirements, let me turn to discussing respective advantages and disadvantages of the two options. This discussion will also provide additional examples and will illustrate the distinction further.

2.1.3.1 Discussing Scope – Guiding Questions

I structure my discussion of scope around seven guiding questions, following a helpful summary provided by Fink (2018). The questions are not intended to be settled definitively. Instead, they serve to bring out prior theoretical commitments that bear on the scope debate and that could be reasons to give preference to one option rather than the other. Whatever one believes to be the correct answer to each question will suggest narrow or wide scope respectively as more plausible. For now, let us start with what are commonly regarded as problems for narrow scope views.

1 – Can rational requirements conflict? Narrow scope requirements can conflict with each other or be inconsistent. As a result, it is sometimes impossible to satisfy all narrow scope rational requirements that apply to you. This has often been seen as a problem for narrow scope views since "it seems plausible that rationality should not issue inconsistent requirements" (Broome, 2007b: 28).[26] As a notion that is concerned with good thinking in the broadest sense, it seems that "we should expect of rationality [that it] should not conflict with itself" (Broome, 2007c: 364), regardless of the overall conception of rationality.[27]

To illustrate this, Broome appeals to deontic dilemmas. Imagine an agent who is faced with such a dilemma – they believe both that they ought to ϕ and that they ought not to ϕ. Now, add the narrow scope version of a popular rational requirement, namely the enkratic requirement, to the mix:

> *(Enkrasia-NS)* Necessarily, if you believe you ought to [ϕ], then rationality requires you to intend to [ϕ].[28]

The fact that narrow scope requirements allow for *detachment* now comes to play a crucial role. Given the narrow scope structure of *(Enkrasia-NS)*, and given that the agent satisfies the antecedent of the conditional, we can now detach the consequent in both applications of the requirement: rationality requires the agent to intend to ϕ and rationality requires the agent to

[26] Broome is not alone in holding this view. A more recent example is Fogal (2018).

[27] One might think that Broome moves too quickly here and that rationality might actually conflict with itself. But this need not worry us. As I said earlier, the questions might not have definitive answers. Disagreement with Broome here only amounts to contesting one objection against narrow scope views.

[28] (Broome, 2007c: 359).

intend not to ϕ. This already seems like an odd result. But this is not a case of conflicting requirements yet. For this, we need to add another requirement that rules out contradictory intentions. Once we do, the requirements of rationality conflict. The agent cannot satisfy *(Enkrasia-NS)* and a (*No Contradictory Intentions)* requirement at the same time. By satisfying *(Enkrasia-NS)* they necessarily violate (*No Contradictory Intentions)* and vice versa.

Compare this result to the wide scope versions of the above requirements. Again, the agent is facing a deontic dilemma; they believe that they ought to ϕ and also that they ought not to ϕ. But now, the enkratic requirement takes wide scope:

> *(Enkrasia-WS)* Necessarily, rationality requires that [if you believe you ought to [ϕ], then you to intend to [ϕ]].

We have not further specified the *(No Contradictory Intentions)* requirement, so we can leave this one unchanged. *(Enkrasia-WS)* does not allow for detachment via modus ponens. The agent can satisfy it by either intending to ϕ (or intending not to ϕ respectively), or by no longer believing that they ought to ϕ (or no longer believing that they ought not to ϕ respectively). This way, they do not necessarily violate a rational requirement, because they could opt for intending to ϕ while no longer believing that they ought not to ϕ, and vice versa, which would satisfy both *(Enkrasia-WS)* and *(No Contradictory Intentions)*. Wide-scoping avoids the result of rational requirements potentially conflicting with each other.

This brief discussion shows that we should prefer a wide scope view, if we think that the requirements of rationality should not conflict.[29]

2 – Do rational requirements pick out the necessary conditions for being fully rational? Broome (2007b) takes the answer to this question to be yes. This commitment is rooted in his logic of requirements, which contains the following Necessity Principle:

> If rationality requires you to ϕ, then necessarily, if you are rational, you ϕ.

But on narrow scope formulations, this principle can come out false. Consider *(No Contradictions-NS)*, which holds that rationality requires you not to believe that not-p, if you believe that p. Combining this with the Necessity Principle yields that necessarily, if you are rational you do

[29] Of course, this is not the only possible result. One might want to agree with this but also maintain a narrow scope view by attacking some of the other elements of the example, such as the possibility of (genuine) deontic dilemmas, the plausibility of *(No Contradictory Intentions)* or the understanding of *(Enkrasia-NS)* as a material conditional. But each of these options would require an independent argument and would have substantial theoretical consequences. This might or might not be warranted and need not concern us at this point in the dialectic – I am merely pointing out the various commitments that might make the narrow or wide scope view more attractive.

not believe that not-p. But that is false – you could be rational and also believe that not-p, e.g. when the evidence supports that not-p. Not believing not-p was only required, given that the agent already believed that p. The problem is that the Necessity Principle only focusses on the part of the requirement that is governed by the rationality-requires operator, which in the case of narrow scope formulations is only the consequent, independently of its antecedent. But narrow scope formulations rely on the presence of the antecedent for their plausibility.

The case is different for wide scope requirements. Combining the Necessity Principle with the wide scope version of *(No Contradictions)* yields that if you are rational, you do not believe that not-p if you believe that p. Since the entire conditional is governed by the rationality-requires operator, both antecedent and consequent find their way into the Necessity Principle.

As a result, we can maintain that on wide scope views rational requirements pick out the necessary conditions for being fully rational, as set out by the Necessity Principle, but not on narrow scope views.[30]

3 – Are rational requirements normative? On this question too, the narrow scope view faces more problems than its wide scope counterpart. But before we can see why, we need to address what the question about the normativity of rational requirements amounts to. A popular way of cashing this out is in terms of so-called Normativity Claims, which link rational requirements to normative notions like *ought* or *reason* (Broome, 2013).[31] In what follows, I will only engage with a reasons-version of a Normativity Claim. As we will see, the narrow scope view runs into problems here. And since the reasons-version is weaker than the ought-version, we can infer that the same, or worse, problems would apply there as well. The claim I will be working with is a simplified version of Broome (ibid.: 193):

> *(NC)* If rationality requires you to x, you have a reason to x.

According to *(NC)*, the normativity of rationality is the normativity of reasons.[32] Whenever a rational requirement applies, we have a reason to satisfy it, to carry out the required action. Rational requirements come with reasons, so to say. But I have already touched on how this creates tension for narrow scope formulations because they allow for detachment. Again, take *(No Contradictions-NS)*, requiring an agent not to believe not-p if they believe p. Combining this with *(NC)* then makes it the case that the agent has a reason to not believe not-p. But, as I mentioned earlier, it might be the case that not-p is actually well supported by the evidence, whereas p is not. In those cases, it might seem that you actually do not have a reason to not believe not-p.

[30] Again, another option could be to deny that the Necessity Principle is appropriately capturing this thought. But for reasons set out in the previous footnote, I do not engage with this possibility here.

[31] I deal with different versions of these in Chapter 8, where I refer to them as Bridge Principles.

[32] This corresponds to *(BPR)*.

The tension between narrow scope views and *(NC)* increases when we consider the problem of *bootstrapping*, which is an instance of detachment. We have seen in the previous chapter how narrow scope requirements can lead to problematic bootstrapping and to self-justifying beliefs (see Kolodny, 2005). In the domain of practical rationality, narrow scope requirements like

> *(Means-End Coherence-NS)*: If you intend q, and you believe that p is a necessary means for q-ing, then rationality requires you to intend p.

lead to the same problematic results of over-generating reasons and potentially requiring the wrong things. Our examples of *(The Ruthless Careerist)* and Muriel illustrated as much. Narrow scope requirements like the one above, in combination with *(NC)*, lead to the result that our agents are now required to intend to kill their competitor, or sharpen their blade, and also have a reason to do so.

To sum up, the narrow scope conception of rational requirements leads to undesirable consequences. Because they allow for detachment via modus ponens, they give us reasons for attitudes in cases where we would normally deny that we have a reason, via *(NC)*. So detachment, and bootstrapping in particular, are serious problems.

As noted earlier, wide scope requirements do not allow for detachment via modus ponens. Such requirements can be satisfied in more than one way and can hence avoid detachment and the ensuing problems. In the case of theoretical rationality, and in particular of bootstrapping, the agent has the option of either believing what p entails or no longer believing that p. Both would satisfy the requirement, and both would be supported by reasons via *(NC)*. In the practical rationality case, our agents are now required to instantiate patterns of attitudes, rather than dropping or adopting specific ones. This is what *(NC)* provides reasons for. For example, the options of dropping the belief that killing one's competitor is a necessary means for getting the promotion, or no longer intending to become a famous axe-murderer would be reasons-supported ways of satisfying the wide scope requirement.

Since wide scope views avoid detachment (and bootstrapping in particular) which created the problem for *(NC)*, we can maintain that rational requirements are normative, as spelled out by *(NC)*, if we adopt a wide scope view rather than a narrow scope one.

4 – Can rational requirements be guiding? The previous discussion might seem to indicate that wide scope views are superior to narrow scope views in all regards. But this is far from obvious, as the following guiding questions will highlight. To start, one might think that it is an important feature of rational requirements that they are guiding. They should tell us how to be rational and, more importantly, how to improve ourselves or move out of irrational states. As Reisner (2009b: 245) puts it, "a theory of rationality that does not tell an agent how to change her mental states will be of limited use". Narrow scope requirements are well suited to issue this

sort of guidance.[33] Because they allow for detachment, their guidance is clear and determinate. Reisner (Reisner, 2009b: 249) even goes further and claims that "rational requirements must be fully determinate to be guiding at all [...] [, and] by 'determinate', I mean that the changes required are not disjunctive options amongst multiple mental states".

If, for example, you intend to see a concert in the neighbouring town and you believe that this necessitates taking the bus at 7pm, then *(Means-End Coherence-NS)* tells you exactly what to do – intend to take the bus at 7pm. Compare this to *(Means-End Coherence-WS)*. This requirement only tells you about the options you have – you could either intend to take the bus at 7pm, or drop your intention to see the concert, or drop your belief about taking the bus being necessary. It provides no further guidance pertaining to how you should choose, or what you should do precisely, other than that you should make this disjunction true.

So wide scope views are inferior to narrow scope views with regard to the ability to provide guidance.

5 – Are there genuinely symmetric options for satisfying rational requirements? This feature of wide scope views, namely that it gives agents a number of options for compliance, is an issue for yet another guiding question. I have already touched on the symmetry objection at earlier points, when we were discussing motivations for the reasons view (see Section 2.1.1.2). The objection holds that wide scope views posit a false symmetry between the different ways of satisfying the requirements. As far as *(Means-End Coherence-WS)* is concerned, all three options are equally good ways of satisfying the requirement. For example, it would be equally rational for you to drop your intention to go see the concert as it would be for you to form the intention to take the 7pm bus. But as Brunero (2012: 126) puts it, "there appears to be [...] an asymmetry in the rationality of the different ways one can escape from a state of means-ends incoherence". This thought is shared by Schroeder (2004), who holds that it should not be the case that giving up an instrumental belief just for the sake of complying with *(Means-End Coherence-WS)* is acceptable, even from the point of view of instrumental rationality. Moreover, Bedke (2009) criticises that wide scope requirements like *(Means-End Coherence-WS)* do not account for the directionality of rationality. By this he means that the various ways of satisfying requirements might be on a par if all else is equal, but not if we consider other relevant features of the situation, such as necessary means-beliefs that are well-supported, or intentions to ends that we care about. If one wants to make room for it sometimes being less rational to give up the intention to an end rather than forming the intention to the means, then wide scope views become unattractive. They do not allow for such asymmetry between the options.

For comparison, narrow scope views do not face this problem, since they do not give the agents options in the first place. So the question about symmetry between options does not even arise.

[33] The guidance point will become salient once more in Section 2.1.4, where it can be used to motivate process requirements.

6 – Do rational requirements provide standards for correct reasoning? One putative function of rational requirements is to provide standards for correct reasoning. They should tell us how to reason from some of our attitudes to others, in a rational way. One specific instance of this is concerned with reasoning that would resolve conflicts. The idea is that if we are in a state of conflicting attitudes, rational requirements (which, at least on the coherence conception, govern such conflict states) should tell us how we can reason ourselves out of this conflict, so to speak. In this way, this is connected to the earlier question about guidance, but applies to reasoning specifically. Take an akratic agent as an example (Kolodny, 2005; Fink, 2018). Such an agent believes that they ought to ϕ, but does not intend to ϕ. If we want to maintain that the akratic agent can avail themselves of rational requirements in order to resolve this conflict, the narrow scope interpretation of *(Enkrasia)* seems favourable.

Here is why. It seems like there is only one way in which the akratic agent can reason themselves out of this conflict state – by reasoning from their belief that they ought to ϕ to forming the intention to ϕ. That is because on closer inspection, the reverse option turns out not to be an option after all. Kolodny (2005) maintains that one cannot reason from the *lack* of one's intention to ϕ to dropping one's belief that one ought to ϕ, because there is nothing that could serve as the starting point for such reasoning. The lack of an intention should not be confused with an intention to, say, not-ϕ. Genuinely lacking an intention just amounts to the absence of an intention and nothing more. Since one cannot reason from nothing to something, there seems to be only one way of reasoning oneself out of conflicts like the one described, which corresponds to what *(Enkrasia-NS)* would require. In general, there being only one option of satisfying a rational requirement is a feature of narrow scope interpretations, as we have previously seen. So if we want to maintain that rational requirements can provide standards for correct reasoning to resolve conflict states, narrow scope views are more fitting.

This observation is closely linked to Kolodny's (ibid.) general argument that some rational requirements *just are* narrow scope requirements. On that line of thought, we do not have to worry about which interpretation is the correct one, since sometimes, there only is the narrow scope option. In order to identify these cases, Kolodny proposes the "Reasoning Test", according to which a requirement has narrow scope, whenever there is only one way to resolve an existing conflict through reasoning from the content of one of the conflicting attitudes to the conflict's resolution (ibid.: 520-521). The example of the akratic agent is one such case. These deliberations also matter for yet another guiding question, to be discussed next.

7 – Do rational requirements depend on attitudes? One way of understanding this question is to ask whether the application of a requirement to an agent depends on attitudes of that agent. On narrow scope views, whether a requirement applies to an agent is sensitive to their attitudes. Since the requirement only comes in with the consequent, agents are under no rational requirement if they do not fulfil the corresponding antecedent. In other words, narrow

scope rational requirements are only triggered once agents hold the antecedent attitude(s).[34] On wide scope interpretations however, rational requirements are not neatly structured into parts that are governed by the requirement and parts that are not. And so they always apply to agents, regardless of whether they have the respective attitudes.[35]

Broome (2007b) thinks that this feature speaks against wide scope interpretations because they consequently collapse the distinction between compliance and avoidance. In particular, he argues that wide scope requirements make it the case that agents can satisfy them by avoiding them. For example, *(Enkrasia-WS)* is satisfied merely by not believing that one ought to ϕ. Similarly, *(No Contradictions-WS)* would be satisfied merely by not believing that p. The same story can be told for the other examples of wide scope requirements. To highlight why this might be seen as an odd result, Broome again relies on his analogy to the requirements of the law. The law requires that, if you are in Britain, you drive on the left. On the wide scope interpretation, this requirement is fulfilled either by driving on the left when actually in Britain (which would amount to compliance) but also by simply not being in Britain (which would amount to mere avoidance). Broome thinks that compliance and avoidance are importantly different phenomena and should not be collapsed.

With an eye to our practice of relying on rational requirements to praise or blame agents for their (ir)rationality, this point seems particularly important. If we praise agents for satisfying rational requirements, we might want to make sure that this satisfaction is the result of actual compliance, rather than mere avoidance. To put the point differently, we might think that an agent who satisfies a rational requirement by complying with it is more deserving of praise than an agent who satisfies a rational requirement by merely avoiding it. If we want to capture these differences, a narrow scope view might be more appealing.

To wrap up this section, we have seen again that no consensus is to be found on yet another choice point pertaining to rational requirements. It should be noted at this point that if we first accept a coherence view of rationality and then inquire about scope within that framework, wide-scoping is by now seen as the default option, mostly because detachment problems are deemed too serious. But this does not mean that wide-scope views are universally endorsed. For example, asymmetry intuitions very much still present a live challenge. Moreover, this has actually motivated a shift in the debate, away from the correct formulation of rational requirements and towards an inquiry about the nature of rationality. In that sense, asymmetry worries can be seen as motivating a reasons-responsiveness approach to rationality, which makes specific demands on agents rather

[34] I take it that this is what Broome has in mind when he talks about the conditionality of the *existence* of rational requirements (Broome, 2007a: 25).

[35] For what it is worth, I think there is a third option on which I rely in my account (see Chapter 4). This option would hold that while technically speaking, wide scope requirements always apply to agents, regardless of their attitudes, they are not always relevant for evaluating the rationality of an agent. For example, a requirement like *(Modus Ponens Closure)* only becomes relevant for whether an agent is rational, if they actually hold a belief that p and a belief that if p, then q; i.e. if they have the attitudes that the (wide-scope) requirement governs.

than merely requiring patterns. But as I have presented things here, with the scope debate being one choice point amongst others, we can still note that there is enough persisting disagreement such that this choice point is no help in our search for the definitive list of rational requirements, which we could then base our rationality attributions on.

2.1.4 State vs Process Requirements

Unfortunately, this story continues when we turn our attention to the last choice point I want to engage with here. This choice point is concerned with whether we should think of rational requirements as requiring agents "to be a certain way at a given time" or to "do something over time" (Kolodny, 2007: 371). The former corresponds to understanding rational requirements as state requirements, whereas the latter interprets rational requirements as process requirements.

This is related to what Fogal (2018: 32) calls the "temporal nature" of rational requirements, which comes down to the question whether they are synchronic or diachronic. Not only seem those two questions related, we might even think that we can match state requirements with synchronic requirements and process requirements with diachronic requirements – that is, we can use the terms interchangeably. But not everyone agrees. For example, Fink (2012: 119) argues that not all diachronic requirements are process requirements. That is because in order for something to be a diachronic requirement, it merely needs to be the case that "its content consists of a cross-temporal relation among a subject's attitudes or actions [...] [, as in:] at t_1, rationality requires of S that [if, at t_1, S believes that p, then, at t_2, S believes that q]". On his understanding, an additional condition needs to be satisfied for something to be a process requirement, namely that it "require[s] a subject to change in a certain way" (ibid.: 118). For example, a requirement to retain one's attitudes would be diachronic, because it makes demands on agents over an extended period of time, but would not be a process requirement because it does not involve change. But this need not worry us here in our discussion of process requirements, since I take it that Fink would agree that all process requirements are also diachronic requirements (but not vice versa). For the sake of completeness, let me briefly consider the opposite pair: synchronic state requirements. If we accept Broome's (2013: 152) definition of state requirements as "concerned only with attitudes that exist at a single time", such that "[i]n the formal statements of them, the same time index 't' is attached to each attitude", then it seems like state requirements and synchronic requirements do in fact run in parallel.

So far, all the requirements I have been considering have been state requirements, either implicitly or explicitly. They have merely specified specific states that agents should make sure to be in, in order to be rational. And in fact, this is not just a matter of convenience but has also been explicitly defended. A recent example is Hedden's (2015) case for "Time-Slice Rationality". This features a commitment to *Synchronicity*: "What attitudes you ought to have at a time does not directly depend on what attitudes you have at other times" (ibid.: 452), and given the congruence of synchronic requirements and state requirements, it also amounts to a commitment

to state requirements. Hedden provides two arguments in support of *Synchronicity*. The first follows from a form of internalism, which holds that the attitudes you are rationally required to have are determined by your mental states, or your point of view, and not by other external factors, such as facts about the world. Hedden takes it that the most natural interpretation of this only involves an agent's *present* mental states, which would put past or future mental states in the same category as external facts about the world. As a result, process requirements cannot be correct since they necessarily involve past or future attitudes of an agent. So if we want to uphold Hedden's internalism, we have to favour state requirements over process requirements.

Hedden's second argument builds on the first. Given internalism, external considerations should not impact what is rationally required of an agent. Arguably, questions about personhood belong to these external considerations. Parfit's (1971) famous personhood puzzles have shown that it is a substantial question whether personal identity persists over time.

Hedden points out that process requirements (and diachronic requirements more generally) assume the truth of a specific position in the debate, namely that persons can be extended over time. Only then can we make sense of requirements that govern the attitudes of the same agent at different points in time. The fact that state requirements avoid this theoretical commitment should make us favour them over process requirements. As he puts it, "[d]etermining what an agent ought to believe does not require first figuring out the correct theory of personal identity over time. This means that requirements of rationality should not make reference to the relation of personal identity over time; what you ought to believe does not depend on who you are" (Hedden, 2015: 452).

However, this preference for state requirements is not universally shared. Many have thought that what really matters for the rationality of agents is not so much what they are like at a given point in time, but rather how they transition from one state to another over time. One subscriber to this view is Podgorski, who holds that "the assessment of an agent's rationality is primarily concerned with processes rather than states" (Podgorski, 2017: 1). Part of his defence of this claim is constituted by his earlier (2016) reply to Hedden. Here, he argues that Hedden's arguments actually do not preclude diachronic process requirements. That is because Hedden restricts his focus to the question of which *attitudes* agents are rationally required to have. However, according to Podgorski, there are other phenomena of interest to rationality, that cannot be cashed out in terms of attitudes because they are essentially diachronic – for example, processes like belief formation, or reasoning more generally. If we broaden the question of what is rationally required to include these necessarily temporally extended phenomena, diachronic process requirements are a live option again. In other words, by restricting his focus to attitudes in the first place, Hedden could be accused of stacking the deck in favour of state requirements. At the very least, diachronic requirements could exist alongside state requirements, where they "govern the most natural diachronic cognitive units – processes such as reasoning". This would still be in keeping with Hedden's internalism constraint, since "those requirements will supervene

on internal features of the agent during those processes" (Podgorski, 2017: 863).

The appreciation of processes that Podgorski shows in his reply to Hedden can also be found in the earlier exchange between Broome and Kolodny. In his (2005) reply, Kolodny asks us to imagine an agent that satisfies *(No Contradictions)* as the result of being struck by lightning, as opposed to having realised that their beliefs that p and that not-p are inconsistent, and having deliberated about which one is more plausible, with the result of giving up the less plausible one. The latter seems to be a rational way of resolving the conflict, the former does not. This difference, so the thought goes, should be captured by rational requirements. But state requirements, which are only concerned with outcomes, would not be able to differentiate between the two cases. In both cases, the agent would satisfy *(No Contradictions)* and be equally rational, all else equal. This goes against the thought that "rationality is, in no small part, a matter of how one forms, retains, and revises one's attitudes" (ibid.: 518). Kolodny relies on the intuitive appeal of these claims to establish that at least some (if not all) rational requirements are process requirements. But he gives us additional reasons to back this up.

One, process requirements subsume state requirements. Whenever you satisfy the process-version of a given requirement, you thereby also satisfy the corresponding state requirement. If process requirements demand us to avoid or escape conflict states (where 'avoiding' and 'escaping' is understood as a process), then we do not end up in such a conflict state. This would suggest that even if we think state requirements are plausible and capture something important about rationality, we can subscribe to a process requirement view, because we get state requirements for free.

Two, we have already encountered the question about the action-guiding potential of rational requirements in our discussion of the earlier choice point of scope. If we think that rational requirements should be prescriptive, or should be able to provide agents with straight-forward advice, then process requirements seem more suited. That is because they issue such advice (e.g. 'avoid this conflict state'), whereas state requirements are merely descriptive, or at best evaluative, in that they provide the "necessary conditions for qualifying for a certain kind of appraisal [i.e. being rational]" (Kolodny, 2007: 372).[36]

The third argument is directly connected to the scope debate. For one, Kolodny's argument for at least some requirements taking narrow scope (via the reasoning test) assumed a process version of these requirements. So if we accept that (at least) some requirements are process requirements, then we also have to accept that (at least) some requirements take narrow scope, assuming that Kolodny's argument is valid. What is more, Kolodny goes so far as to suggest that the scope debate only makes sense when we allow for process requirements. In his (ibid.) reply to Broome (2007c), he argues that if we only allow for state requirements, there actually is

[36] Given the importance of this kind of appraisal for our practices of praising or criticising agents, this point of Kolodny's might not actually worry us this much.

no difference between the narrow and wide scope versions in terms of violations. Both scope versions are violated in exactly the same scenarios. If we combine this with Broome's claim that the rationality of an agent depends on the absence of violations of rational requirements, we get the result that the choice between narrow and wide scope makes no difference to the rationality of agents. It is only once we accept process requirements that the corresponding scope versions come apart. Only then can we have cases where one is violated, whereas the other is not. The ubiquity of the scope debate, combined with this insight, leads Kolodny to postulate an error theory: "it seemed to me that there was an important difference between narrow and wide scope, because first, there is an important difference when the relevant requirements are process requirements, and second, I was implicitly taking the relevant requirements to be process requirements" (Kolodny, 2007: 375). In other words, we already implicitly subscribe to process interpretations of rational requirements, which is why the scope debate seemed interesting in the first place.

This concludes my discussion of the four choice points regarding rational requirements. As it stands, these choice points are surrounded by disagreements and controversies and so the prospects for finding the definitive list of rational requirements are dim.

2.2 Putting It All Together: Connections Between the Distinctions

But maybe there is hope. With this cornucopia of issues on the table, one might start to wonder whether we can impose some more structure on the choice points. So far, I have presented them as largely independent of each other. However, could it be that this pick-and-mix approach glosses over some of the relations between the choice points? In particular, are there connections between them such that our choice with regard to one of them would commit us to a particular stance with regard to another?

2.2.1 Local Wide Scope State Coherence Requirements

And indeed, if we look at the different choice points, it seems that someone who thinks that rational requirements are coherence requirements would also be drawn to thinking that they are local, that they take wide scope and that they are state requirements. And the complementary combination of reasons-responsiveness with global, narrow-scope process requirements seems to suggest itself equally naturally. In the following sections, I will draw out these connections more explicitly. As an illustration, it will be helpful to think of the choice points as nodes in a decision tree.

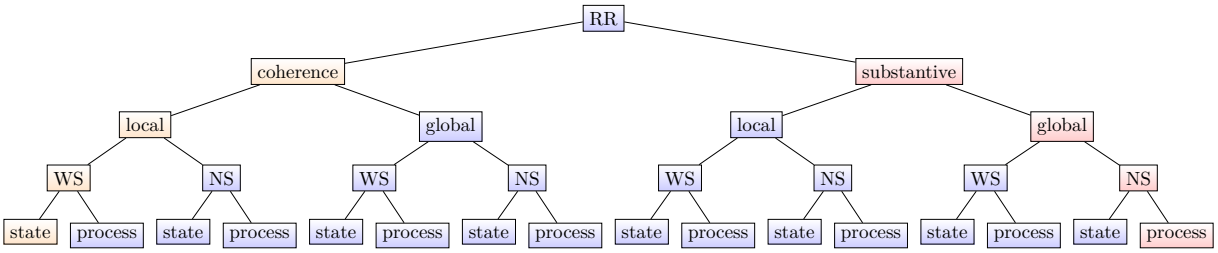


Figure 2.1

Now, it is easy to see that the above-mentioned connections, which are highlighted in the tree, are not the only possible connections. We could, at least theoretically, conceive of a process narrow scope global coherence requirement. Neither is the order in which I present the choice points necessary. But I do not think much hangs on this. Presumably, the same combinations would seem attractive, even if we started with e.g. the state vs process requirements choice point. These remarks aside, we will see that the two paths that are highlighted suggest themselves so naturally that I will focus on them.

Let us first turn our attention to the connection between *coherence* and *local* requirements. This will then serve as the starting point for developing the other connections. So we start our journey with the choice point about the nature of rational requirements. If we think that these are coherence requirements, we think that they require or prohibit specific patterns of attitudes. We have already encountered *(Means-End Coherence)* as an example in the domain of practical rationality; and closure requirements are examples in the domain of theoretical rationality:

> *(Means-End Coherence)* Rationality requires that [if you intend to e, and you believe that intending to m is necessary for e-ing, then you intend to m.]
>
> *(Modus Ponens Closure)* Rationality requires that [if you believe p, and you believe that p implies q, then you believe q.]

With this commitment to coherence requirements in mind, which stance should we take on the question about the *extent* of rational requirements? Once we have engaged on the coherence path, it seems very natural to think of rational requirements as being local. To recall, local requirements are taken to govern small specific sets of attitudes. This seems to sit particularly well with coherence requirements and their focus on specific conflicts. Instead of making demands on the entirety of an agent's attitudes, they explicitly refer to a small number of very specific attitudes: an intention to an end, a means-end belief and an intention to a means in the case of *(Means-End Coherence)*, and three specific beliefs in the case of *(Modus Ponens Closure)*. To repeat Kolodny's words, they require agents to "avoid or resolve some specific conflict among one's attitudes" (Kolodny, 2005: 516). If we understand the idea of 'specific conflict' as necessarily

limited to a small set of attitudes, then this is one way of spelling out what it means for a requirement to be 'local'. And so coherence and local requirements seem closely aligned.

Moving on to the next choice point, it seems natural to opt for rational requirements having wide scope, given a coherence framework. Wide scope requirements specify patterns of attitudes to be instantiated or avoided. This sits well with a coherence view of rational requirements, since such a view also holds that rational requirements are concerned with patterns, rather than isolated individual attitudes.[37] So it seems like endorsing a coherence view generates a commitment, or at least an affinity, for requirements being local and taking wide scope.

With this in place, we can contemplate the final choice point regarding the temporal nature of requirement. The earlier endorsement of the local view now makes state requirements seem more attractive – the idea that rational requirements govern small, local subsets of attitudes is much easier to square with state requirements, which focus on how agents are at a given time, rather than over an extended period of time. Presumably, the number of attitudes we need to look at to decide whether an agent satisfies a state requirement is a lot smaller than the amount we would need to take into account when evaluating how an agent is doing over time. And this more limited focus seems to reflect the idea of local requirements rather well.

To conclude this section, a commitment to coherence seems to motivate the wide scope and local views, which in turn motivates state requirements. As a result, the combination of local wide scope state coherence requirements seems natural. Note that next to being of theoretical interest, there is another potential upshot to this insight: we may no longer need to debate every single choice point individually. Merely deciding one a coherence vs reasons-responsiveness conception could settle the other choice points, too.

2.2.2 Global Narrow Scope Process Substantive Requirements

Similar observations can be made regarding the combination of substantive requirements and the global view. Take our example of a substantive requirement:

> *(Reason-Responsiveness)* Rationality requires that you intend what you have decisive reason to intend.

Unlike the coherence requirements above, this substantive requirement does not explicitly refer to specific attitudes. Instead, it might well be understood as imposing "global constraints on all of one's attitudes" (Brunero, 2012: 129). That is because in order to see whether an agent complies with *(Reasons-Responsiveness)*, we need to take into consideration a large set of her attitudes, if not all of them. How else could we decide what the agent has decisive reason to

[37] The reader might wonder whether this natural alignment only holds between coherence and wide scope requirements, and not necessarily local requirements. This will be addressed in the next section.

intend? This will require considering her other intentions, her beliefs about value and feasibility of her intended ends, general beliefs about her circumstances etc. And so one might think that a commitment to substantive requirements creates a commitment to a *global* view about rationality requirements, i.e. to a view that holds that rational requirements govern the entire set or at least a large subset of one's attitudes.

In a similar manner, substantive requirements seem to sit particularly well with narrow scope requirements. Both require specific attitudes of agents, rather than mere general adherence to a pattern. They do not give the agent options. To drive home this point, consider the alternative: combining a substantive requirement like the above, which demands the adoption of a specific intention, with a wide scope formulation, which would allow for an alteration of what there is decisive reason to intend, seems very odd.

And again, once we have decided on substantive requirements and consequently global and narrow scope requirements, opting for process requirements seems to follow quite naturally. In particular, global requirements seem to sit well with the idea of process requirements. We have seen previously that process requirements evaluate agents at different points in time, which already increases the amount of attitudes that need to be considered when evaluating whether a process requirement is violated. This might not be possible on a local view, but does not pose a problem for the global view. What is more, we might think that next to attitudes at earlier points in time, we also need to evaluate a number of background conditions, or related attitudes, to decide whether the process an agent followed to get to their current state was a good one, or a rational one. For example, it does not seem sufficient to only consider an agent's earlier belief that p and their current belief that q, but also whether they hold the belief that if p, then q and whether they used it in their reasoning process. A global view seems perfectly suited to deal with this potential plenitude of attitudes that play a role for processes.

This discussion suggests that the choice point about the right view about the nature of rational requirements might have further-reaching consequences. It seems plausible that it interacts with the other choice points about extent, scope and temporal nature of requirements, producing two sets of combinations as illustrated in the above figure.

2.2.3 A Merely Conventional Fit

But this seems almost too good to be true. Can we really dodge all the hard questions about the various choice points surrounding rational requirements, once we settle on their nature and realise what this commits us to down the line? The sceptic might think that the combinations I have drawn out in the previous section might merely be the products of convention and convenience, instead of being indicative of an actual or necessary connection. To see this, note that the decision tree I used already indicates that there is conceptual space for other combinations too – we could choose different paths, and we could order the choice points differently. In particular, the previous discussion drew out that the connection between, say, coherence and state requirements

is an indirect one that needs a commitment to local requirements as an intermediary step. The same was true for the connection between substantive requirements and process requirements. Substantive requirements sit well with global requirements, which in turn sit well with process requirements. Changing one's mind about e.g. the extent of requirements could then open up the possibility for new combinations, such as coherence wide scope global process requirements. I take it that the existence of such possibilities can motivate scepticism about the neatness of the picture I presented.

Let me develop this sceptical thought a bit further by taking a closer look at the seemingly most fundamental choice point – the nature of rational requirements. Could there be global coherence requirements and local substantive requirements? In theory, this seems possible. But when we search for actual examples, the literature quickly runs out. Both Kolodny and Brunero, as some of the few authors who explicitly deal with the local/global distinction, only mention examples for local coherence requirements. So it seems like we need to construct examples for other combinations ourselves, as illustrated in the following figure.

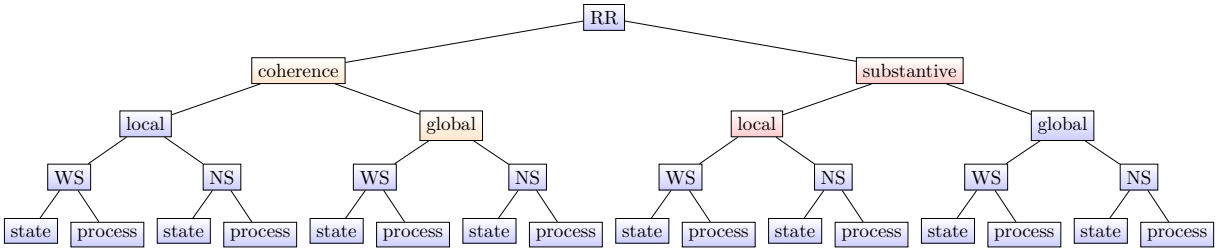


Figure 2.2

Starting with the left hand side, what could global coherence requirements look like? The following might be a good starting point:

> *(Global Coherence)* Rationality requires that you have coherent attitudes.

This requirement is obviously a coherence requirement. It is in keeping with the seeming platitude that rationality is ultimately about coherence amongst one's attitudes, which provides the underlying motivation for the coherence view. And it is also a global requirement, because its demands span over the entire set of an agent's attitudes – it requires coherence everywhere.

Now, there are various things one might want to say about the plausibility of such a requirement. For one, it does not seem to fulfil a commonly held desideratum for rational requirements, which we already encountered in the scope discussion: action-guidingness. According to this desideratum, it is important that we can bring ourselves to be more rational, e.g. through deliberation, where rational requirements are an important guide to the rational response. But if I find myself in a situation like Sara's and I have to decide whether to take a lunch break,

(Global Coherence) is of little help. In order to decide on the rational thing to do, I would need sit down and consider all of my attitudes and how each of my options (forming or abandoning my means-end belief and the intentions to skip lunch and to submit today) would affect the overall coherence of my attitude set. This seems cumbersome at best, if not infeasible at worst. And there might not even be a clear solution. As we have seen, there is often more than one way of maintaining or re-establishing coherence. *(Global Coherence)* would provide no guidance with regard to which option to choose.

Moreover, one might think that *(Global Coherence)* does not even do justice to all aspects of the motivation for the coherence view. Whilst it is undoubtedly concerned with coherence, it is not concerned with specific conflicts. To use Broome's words, this requirement is not one that "requires your mind to be coherent in particular aspects" (Broome, 2013: 152), but only to be coherent in general. Given these points, the prospects for *(Global Coherence)* seem dim.

A very similar story can be told for the possibility of the corresponding alternative pair – local substantive requirements. For this combination to hold for a requirement, it would need to be *local* in its focus and substantive in its demands. Again, the literature runs out here. But maybe we can construct our own potential candidate:

> *(Reasons-Responsiveness$_{local}$)* Given attitude set a_n, intend what you have decisive reason to intend.

Because this requirement involves a restriction to an attitude set, it can be local, namely whenever the set a_n is a small one. For example, it might only include Sara's intention to send her paper off today and her belief that this necessitates skipping lunch. The demands of the requirement are still substantive. Instead of prohibiting patterns, it places the agent under a specific demand such that the requirement can only be satisfied if the agent complies with this particular demand. For comparison, giving up the means-end belief would not make her satisfy the requirement.

Again, one might think that this is not a particularly useful requirement. After all, the appeal to an all-things-considered notion like 'decisive reason' might be seen as in tension with a restriction to a small subset of attitudes. What use is it to do what we have decisive reason to do on a small part of the playing field, when this might not coincide with what we have decisive reason to do on the entire playing field? So again, it seems like whilst it is at least possible to construct local substantive requirements, the results might not be very appealing.

What does this mean for the remaining choice points? I think we should adopt a similarly cautious stance here. Different combinations might be possible. For example, as previously noted, combining local wide scope coherence requirements with state requirements was mostly motivated by a good fit between local requirements and state requirements. If we can conceive of global coherence requirements, the affiliation with state requirements seems less straight-forward.

Instead, we might want to combine global coherence requirements with process requirements. The same can be said for combining local substantive requirements with state requirements (see Figure 2.3).

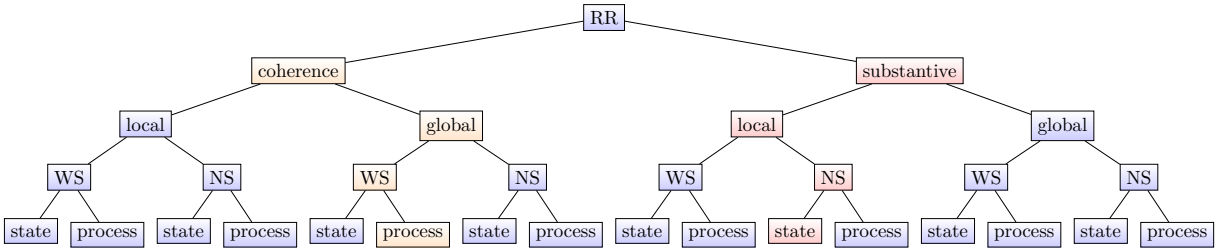


Figure 2.3

What are the implications of all of this for the overall dialectic? I have started out by exploring all the different choice points we would need to settle in our search for the definitive list of rational requirements. This brought to light an abundance of issues, arguments and counter-arguments, such that the prospects on settling on the correct requirements seemed dim. Taking note of some seemingly natural connections between the choice points sparked some hope in this respect. If committing on one option with regards to one choice point committed us to another option with regard to another choice point, maybe we could settle all these hard questions definitively and efficiently. But as this section has pointed out, these connections should be taken with a grain of salt. They might not amount to necessary relations or commitments. However, we have also seen that some of the alternative combinations did not seem very desirable after all.[38] The take away lesson then is this: whilst we cannot hope to settle all the choice points by merely settling one, there still is some pressure to lean towards a certain direction generated by previous choices about choice points. Whilst this does not solve all our problems, it can make our lives a bit easier by pre-structuring the search for the correct requirements. But that is all we can hope for.

2.3 Leaving the Arena of Rational Requirements

Let us take a step back and remind ourselves of what is at issue in this chapter. Given the important role that rationality plays in our practices of praise and criticism, we faced the problem of how to decide when to attribute rationality to an agent. Combining this problem with the popular requirements-based understanding of rationality motivated the turn to requirements. If we could establish the authoritative list of rational requirements, we could easily decide on an agent's rationality, since this would just correspond to them conforming to the requirements. This search for the authoritative list raised a number of issues and choice points, which I have discussed at length. Now that we have considered the manifold options, their strengths and

[38] Of course, my remarks cannot claim to amount to a complete investigation of all possible alternative combinations and the entirety of the conceptual space occupied by rational requirements. But given the modesty of the conclusion I drew from this discussion, this need not worry us much.

weaknesses and their connections, we are no closer to establishing this list. One might take this to motivate agnosticism about requirements – maybe this authoritative list just cannot be established. But even if this seems too extreme of a reaction, the previous discussion provides good reason for changing strategy. Instead of trying to settle the many issues with rational requirements, can we approach the problem the other way around and focus our attention on the nature of rationality *attributions* first?

Chapter Three

Rationality Attributions

Putting this point differently, maybe it is helpful to direct our attention away from the question about *what* rationality requires of us, and towards the question of *where* it does so. The second is a question about the area of evaluation: when deciding on the rationality of an agent, *where* do we need to check for compliance with rational requirements?

So far, we have approached the question of how to make rationality attributions largely in isolation of how we actually make these attributions. Once we pay more attention to these actual practices, we see that the requirements-based approach of the first chapter might be too limited because it does not give due consideration to the *where*-question. Recall the example of Sara. In the first case, Sara seemed rational in intending to skip lunch if we took into consideration her attitudes of intending to submit her paper today and her belief that this necessitates skipping lunch. But once we also took into account her beliefs that regular lunch breaks are important and that today's deadline is not, the opposite could be the case: it might be irrational for her to intend to skip lunch. So it seems like even though we were concerned with whether she is means-end coherent, and hence whether she satisfies a means-end coherence requirement, this did not quite capture all that we care about when we decide whether to attribute rationality to Sara. In other words, even if we have an answer to the *what*-question (we care about means-end coherence), it is not clear that this also amounts to an answer to the *where*-question. Which particular set of the agent's attitudes we focus on is not a neutral choice. Instead, this choice has an impact on which rationality attributions seem plausible.

We can make similar observations about the second example. When taking into account Sara's belief about skipping lunch being necessary for submitting, and the absence of her intention to skip lunch, it seems like we can attribute rationality to her if she drops her intention to submit today. But again, once we factor in other attitudes of hers, like her belief that she has good reason to submit her paper today, the reverse seems more plausible. Maybe it is not rational for Sara to drop her intention to submit.

This brings out two things: one, the *where*-question matters. Which set of attitudes we focus on influences whether we would want to attribute an agent rationality. Two, this is one reason why merely focusing on the *what*-question (i.e. only on finding the correct requirements) is an inferior strategy, since it does not let us sufficiently appreciate the role that attitude selection plays.

What is more, focussing on rationality attributions and the role of attitudes might eventually lead to some progress in the domain of rational requirements, too. It might give us some indications as to which choice points seem more or less desirable. For example, if we acknowledge that we might care about additional attitudes rather than just intentions to means, intentions to ends and beliefs about means-end relations, even if our concern is means-end coherence, this might speak in favour of an understanding of rational requirements that is not strictly local but instead lets us incorporate additional attitudes that seem relevant. In a similar vein, this might motivate process-requirements over state-requirements, since the former seem better suited to accommodate additional attitudes.

Given the preponderance of choice points related to the *what*, it is no surprise that the *where* has not been sufficiently appreciated in the literature. One aim of this chapter is to give this question due consideration. The other aim of this and the next chapter is to provide a more reliable procedure that allows us to move from an agent's attitudes to a rationality attribution, now that we have seen the need for this, given that our intuitions about whether rationality attributions are warranted seem to shift with focus, or, more precisely, with the set of attitudes we are concerned with.

3.1 Local vs Global Attributions

In order to settle the *where*-question, we can make some use of the previous discussion of rational requirements, and in particular of our discussion of the question about their *extent*. And indeed, the question whether *requirements* are local or global is often discussed in close connection with the question whether rationality *attributions* are local or global. On a local conception of rationality attributions, these would be very narrow in focus. The area that we have to check for compliance with rational requirements is a small one: only a small subset of the agent's attitudes is evaluated.

The alternative conception of rationality attributions as global can be seen as the polar opposite: when we want to decide on the rationality of an agent, we have to take into account all of their attitudes. The entire set of attitudes has to be checked for compliance with rational requirements. I have already discussed how this terminology appears in relation to rational requirements, in particular in the works of Kolodny (2005) and Brunero (2012). Now that we have changed strategies, let us revisit these remarks with a focus on attributions.

> "Whereas "local" judgments of rationality specify what's rational in light of a certain

> subset of an agent's attitudes—perhaps what's rational in light of the fact that one intends some end, has a certain instrumental belief, but doesn't intend the means—"overall" (or "all-attitudes-considered") judgments of rationality specify what's rational in light of the totality of an agent's attitudes" (ibid.: 129).[1]

A local rationality attribution (or judgment), according to Brunero, would only focus on a small subset of attitudes, e.g. an intention to an end, a means-end belief, and the presence or absence of an intention to a means.[2] As a second step, we would then refer to our list of rational requirements and check whether the agent satisfies or violates the requirements that are applicable, given this particular subset.[3] As Brunero points out, in a case of means-end incoherence like Sara not intending to skip lunch, the local view might then yield one of the intuitive verdicts discussed earlier: Sara is irrational because she violates a rational requirement that is applicable here. Other attitudes that she might have, like a belief that lunch breaks take priority over deadlines, are not being considered. Compare this to a global rationality attribution. On this perspective, we have to check the agent's entire set of attitudes for compliance with rational requirements. The global view might then have it that Sara is in fact rational. Here is how: whilst it is true that she violates *(Means-End Coherence)* with regard to the small subset of her attitudes, it might still be that she does what is rational overall, e.g. complying with *(Reasons-Responsiveness)* by intending what she has most reason to intend, e.g. having lunch.

While Brunero seems to allow for both local and global rationality attributions, or at least remains neutral on the correct view, Kolodny clearly favours the local view, as the following, by now familiar, quote shows. Note that Kolodny transitions quickly from a claim about the form of rationality attributions (or judgments) to a claim about rational requirements. Whether he is right in claiming that the local view is correct for *requirements* has been discussed previously and is not of concern just now.

> "Our ordinary judgments about rationality [...] are local. They are focused on specific conflicts among one's attitudes. [...] Rational requirements, accordingly, ought to be local. In each instance in which one is under a rational requirement, what

[1] Both Kolodny and Brunero phrase their discussion in terms of rationality 'judgments' instead of 'attributions', which I favour. I take it that not much hangs on this choice of terminology.

[2] Including absences of attitudes in the evaluation is common practice. See for example Kiesewetter (2017: 16), who uses "attitude-state" as "the generic term for both attitudes and lack of attitudes". He follows Brunero (2012: 28) in that practice: "An attitude-state could consist in either the presence or absence of an attitude". For an example that is specific to beliefs, see Worsnip (2018: 6): "D(p) is a possible doxastic attitude (D) of a subject (S) towards a proposition (p): believing p, disbelieving p, or suspending judgment about p".

[3] I take it that this is comparable to Broome's (2007b) remarks about the conditionality of requirements. He contends that narrow scope requirements are conditional in their existence – if the antecedent does not hold, no requirement applies, just like I am under no obligation to drive on left if I am not in Britain. I understand the local view about rationality attributions to entail this conditionality for requirements generally, even for wide scope ones. If the subset we are considering does not contain the attitudes that are governed by the requirement, regardless of its scope, it simply does not apply to the agent.

> it ought to require of one is to avoid or resolve some specific conflict among one's attitudes–as opposed to, say, to satisfy some global constraint on all of one's attitudes. One might liken each application of a rational requirement to a referee with authority over a different part of the playing field, or to an inspector with authority over a different stage in a production process. Various applications of rational requirements may call for one to form some attitudes, to retain others, and to revise still others all at the same time" (Kolodny, 2005: 515-516).

To use Kolodny's analogy, what we as referees evaluate on the local view is only a portion of the playing field – a subset of an agent's attitudes, whereas the global view would hold that we need to keep an eye on the entire playing field – all of an agent's attitudes, when making rationality attributions. Whether this means that the referee really only has authority over a small portion of the playing field (as would be entailed by local *requirements*) is a related, but different question.

These, then, are the key insights gained so far: we have already established that rational requirements are an important guide to making rationality attributions.[4] They tell us *what* to look out for. This led us to the question of *where* to look for it. The local and global view about rationality attributions provide answers to this question.

Let me note that the choice between the local and global perspective at the level of requirements is of course related to the corresponding choice at the level of attributions. For example, one might think that a global view of requirements commits us to a global view about attributions, if we understand 'global' as 'literal-global', i.e. as simply ranging across all attitudes. How else would one check whether a specific global requirement has been violated or satisfied? In this case, it seems like the answer to the question about *where* to look falls out of the answer to the question of *what* to look for when attributing rationality. In contrast, a commitment to a local view about requirements leaves open more options. It might motivate an equally local view of attributions. It just seems fitting that if the requirements we rely on are local, then the set of attitudes that we have to evaluate has to be local too, or so the thought goes. Another option might be to understand a global view about attributions more along the lines of the 'aggregated-global' understanding, i.e. as multiple applications of local requirements, which would leave open the option of mixed perspectives between requirements and attributions: the *what* would not necessarily provide an answer to the *where*.

To further illustrate the difference between local and global rationality attributions, we can again make use of Brunero's parallel to Rossian prima facie duties and duties proper. As previously discussed, Brunero develops this parallel somewhat inconsistently and jumps back and forth between requirements and attributions. But this need not concern us here; the parallel is nevertheless illuminating. Local attributions, on which we evaluate only small subsets of

[4] On a requirements-based conception of rationality, they are the *only* guide to determining rationality.

attitudes at a time, can be likened to Rossian prima facie duties, which also only focus on a small area of moral obligations. For example, we might only focus on Sara's intention and her means-end belief; and only on my promise to meet my friend at 5pm respectively. The resulting rationality attributions (or moral judgments, in Ross's case) would be limited to this small area as well, presumably resulting in negative evaluations (because Sara is means-end coherent by not intending to skip lunch; and because I am breaking a promise by giving a lift to my other friend instead).

Global attributions, in turn, would factor in the entire (or large) set of attitudes, and other morally salient features respectively. Once we take into account a wider range of factors, it might be that even though I have promised to meet my friend, I also have the duty to help a friend in need. It could then be the case that we decide that it is at least permissible for me to give my friend a lift, all things considered. And so there is an option for a positive (moral) evaluation at the global level. A similar story can be told for the rationality case: once we take into account a much larger set of attitudes, if not the entire set, we would include e.g. beliefs about lunch taking priority over deadlines, beliefs about feasibility and importance of submitting today, and other potentially competing intentions. This broader set of attitudes would plausibly lead to a different, larger set of requirements that are now applicable which might then lead to different, and potentially positive, rationality attributions as well.

What is important is that we realise that the choice between the local and the global perspective matters. Depending on which perspective we choose, the results (regardless whether the result is a moral judgment or a rationality attribution) can differ.

3.2 Discussing Local Attributions

We have so far considered two potential answers to the question of *where* to look for compliance with rational requirements when making rationality attributions. According to the local perspective, rationality attributions are concerned with a small portion of an agent's attitudes. On the global perspective, all or almost all of their attitudes matter. Recall Kolodny, who endorses the local perspective without further ado: "[o]ur ordinary judgments about rationality [...] are local. They are focused on specific conflicts among one's attitudes" (ibid.: 515-516). A more moderate commitment to the local view can be found in Brunero (2012), who thinks that both the local and the global perspective are important when making rationality attributions: "Whereas "local" judgments of rationality specify what's rational in light of a certain subset of an agent's attitudes [...] overall" (or "all attitudes-considered") judgments of rationality specify what's rational in light of the totality of an agent's attitudes" (ibid.: 129). The two perspectives simply play different roles.

This way of characterising the two perspectives raises an obvious delimitation worry. Where should we draw the line? How are we to delimit the local subset that we are evaluating? Neither

Kolodny nor Brunero explicitly deal with this problem. But Brunero at least makes a suggestion. He discusses an example of means-end coherence similar to Sara's and suggests that the local subset should include attitudes that are "relevant to instrumental rationality" (Brunero, 2012: 130). When we apply his treatment of the example to Sara's case, he would hold that the "relevant" attitudes only include her intention to send off her paper today, her means-end belief, and her lack of intention to skip lunch.

3.2.1 The Strictly Local View

This suggests the following principle for delimiting local subsets: the attitudes to be included are the ones mentioned in the respective requirement, together with their potential absences. Let us call this the 'strictly local view'. Let me register two issues with this principle right away.

One, it relies heavily on the particular formulation of rational requirements. But we have seen in Section 2.1 that almost every aspect of these formulations is subject to debate and that seemingly natural combinations might actually be rooted in convention or convenience rather than facts about what rationality requires. Hence, it might not be prudent to base important decisions on them, such as the decision about the extent of the local subset.

Two, even if we could rely on the particular formulation of a requirement like *(Means-End Coherence)* to be the correct one, this seems to entail an impoverished understanding of what is relevant to instrumental rationality. The worry is that we exclude relevant attitudes this way. For example, this picture excludes other beliefs of the agent from the subset of attitudes that are deemed 'relevant'. But we already considered an enriched version of Sara's situation, in which she holds the belief that taking a lunch break is in the interest of self-care and ought to take priority, or that she has been able to send papers off in similar situations in the past when she did take a break. These beliefs of her about inconclusive evidence for her means-end belief or about what she ought to do surely seem relevant for her instrumental rationality. But the principle that we derived from Brunero's suggestions does not allow them to enter the strictly local subset.

Trouble from Symmetry In what follows, I provide further support to this initial worry about the strictly local view being too restrictive, and also argue that Kolodny's outright claim that our ordinary rationality judgments (or attributions) are local, is false. This is important, insofar as he relies on this claim to motivate the local view about attributions and requirements.

How plausible is the claim that our *ordinary* practice of making rationality attributions is strictly local? There are various reasons to be doubtful of this. One, we tend to make further reaching rationality attributions. This is illustrated by the more elaborate version of Sara's case. It is not uncommon for agents to have beliefs about the quality of their other beliefs, or general beliefs about what they ought to do. If Sara has successfully submitted papers with similarly tight time frames in the past, then this provides her with a belief that potentially undermines the plausibility of her belief that skipping lunch is necessary. We can say something similar

for a plausible belief of hers that she ought not sacrifice self-care for deadlines: this stands in conflict with her intention to send the paper off today. If we are asked to evaluate the rationality of an agent like Sara, who does not intend the means to her ends, it is a natural reaction to withhold judgment at this point and inquire further about e.g. the evidence available to the agent, their beliefs about feasibility, about what they ought to do, and so on, instead of attributing irrationality to Sara right away. This story implies that our ordinary rationality attributions are not strictly local. The set of attitudes that we want to evaluate for compliance with rational requirements is rarely this clear-cut and rarely this small.

The just described tendency to withhold judgment sits well with another tendency of ours to prefer more information to less as the basis of our attributions. We ordinarily think that more information leads to more accurate results. In making rationality attributions, we then take into account many attitudes. This tendency to include more attitudes in our rationality attributions sometimes even goes so far that we fill in an agent's attitude profile: we attribute attitudes to agents that we think they are likely to have. For example, we are likely to attribute a belief about her means-end belief being sufficiently supported by the evidence to our means-end incoherent agent. Why else would she hold that belief? I take the *symmetry objection* to provide evidence for these claims.

To see this, take Sara's case again. As discussed earlier on, the symmetry objection holds that the various ways in which Sara can satisfy *(Means-End Coherence)* are not equally rational. There seems to be an asymmetry between i) intending the means, ii) dropping the means-end belief and iii) not intending the end. In most cases, intending the means is seen as the best way of satisfying the requirement, whereas the other two seem more like rationalisations.

This intuition cannot be explained by the strictly local view about rationality attributions that we are exploring here. Indeed, if this view is correct, we should not have the intuition in the first place. However, it would be a plausible explanation for asymmetry that we include additional attitudes in our attributions, for example through evaluating multiple conflicts at once. Take dropping the intention to the end. One explanation of why this seems like a less rational option could be because it can lead to the violation of another rational requirement, namely an enkratic requirement that requires agents to intend what they believe they ought to intend. Plausibly, Sara believes that she ought to send off her paper today. If she then no longer intends to send off her paper today, she satisfies *(Means-End Coherence),* but violates the enkratic requirement.

At this point, one might think that everything I have said so far is consistent with Kolodny's conception of local rationality attributions and requirements as referees in charge of different parts of the playing field. After all, we could just understand the strictly local view in terms of repeated evaluations of small attitude sets, or repeated calls over small parts of the playing field. But I contend that unlike on Kolodny's strictly local view, our rationality attributions are not so fragmented. Here is a potential explanation for why: what we are ultimately evaluating in cases like Sara's is instrumental rationality. Requirements like *(Means-End Coherence)* are only

proxies for this; our notion of instrumental rationality is likely to be richer. And this in turn can explain our reluctance to attribute rationality to agents like Sara, who liberally give up their intentions to ends, or means-end beliefs in order to satisfy the requirement. Broadly speaking, when we care about instrumental rationality, we care about how agents perform with regard to achieving their ends. And agents like Sara are not very good at this. To highlight this, take an example from Brunero (2012: 236) which we will reencounter later:

> "Candice decides to go to the post office this afternoon to send out some mailings, but on the way there, she gives up on this end and decides to go buy groceries instead. But on the way to the market, she yet again trades in this end for another: going to hang out with her friend David. But on the way to David's house, she once more changes her mind and intends to spend a relaxing afternoon at home, but by the time she gets home the afternoon is gone and she's accomplished nothing. We're inclined to accuse Candice of some kind of irrationality. And, since Candice failed to follow through on any of her ends, we're inclined to accuse her of being instrumentally irrational."

Our practice of wanting "to accuse Candice of some kind of irrationality" because "she's accomplished nothing", and despite her satisfying *(Means-End Coherence),* is evidence for a broader understanding of instrumental rationality which underwrites a potential rationality attribution. These remarks are incompatible with a strictly local view of rationality attributions. So if we want to be able to "accuse Candice of some kind of irrationality" and are drawn to a richer understanding of instrumental rationality as the capacity to pursue one's ends, we should reject the strictly local view, where the only relevant attitudes are the ones mentioned in the respective requirements. All of this suggests that the fragmented picture suggested by the strictly local view is not very appealing, as it cannot capture all aspects that we deem relevant in making rationality attributions.

Comparative Attributions and Aggregation Worries One might be doubtful of my previous claims about rationality attributions. Instead, one might think that we should take seriously Kolodny's analogy and simply accept that each strictly local rationality attribution simply covers a different part of the playing field. This would still allow us to evaluate different aspects of the agent's rational performance, depending on how we cut up the playing field. However, I think that this picture makes us lose out on an important function of rationality attributions, namely being able to make comparative rationality attributions.

To see this, imagine two agents. One is very close to being ideally rational. In fact, she satisfies all rational requirements that apply to her with the exception of *(Means-End Coherence)*. The other one is entirely irrational and violates all rational requirements, including *(Means-End Coherence)*. If our ordinary rationality attributions are strictly local, they will only focus on

one small subset of attitudes at a time. This means that when we compare the two agents, our attributions regarding the specific subset that includes the attitudes picked out by *(Means-End Coherence)* should be the same: both agents are equally (ir)rational. How plausible is this? Whilst I will admit that this result is not incorrect, since both agents fare equally with regard to this one specific conflict, I still hold that it is not an ideal result. It is misleading to make no difference in the rationality attributions for agents whose performance could hardly be more different.

I should say that a proponent of the strictly local view like Kolodny might simply deny that this is a problem. In a footnote, he explicitly rejects potential views that rely on aggregation mechanisms to move from individual applications of local requirements to a more global, overall verdict.[5] The reason he gives is that "[this idea] does not correspond to any recognisable feature of our ordinary attributions of subjective rationality and irrationality" (Kolodny, 2005: 516). But this is precisely what I deny. Aggregation does correspond to a recognisable feature of our ordinary rationality attributions, with this feature being comparison. To illustrate this, take the less extreme case of asymmetry. Two agents are faced with means-end incoherence. They do not intend what they believe to be the necessary means to their end. Now, one of them responds to this conflict by intending the means. The other one responds by dropping their means-end belief. I take it that in our ordinary rationality attributions, we would want to account for the difference that intuitively exists between the two: the first agent seems to be doing better from the point of view of rationality than the second.

Now, one might suggest that one way to account for this on a strictly local view is to simply appeal to multiple strictly local attributions. Whilst it is true that the same verdict is cast with regard to *(Means-End Coherence)* on the agents in both cases, assessing their performances in connection with other rational requirements will lead to different results. But I take it that this way of accommodating comparisons already amounts to no longer endorsing a *strictly* local view. Even if our aggregated attributions simply take the form of points on a scorecard, where we compare the number of satisfied or violated requirements, we are nevertheless engaging in some form of aggregation. And once we acknowledge this, we might just as well ask whether this very simplistic aggregation method really is satisfactory in order to allow for accurate and informative comparisons. I take it that this is not the case – it does not take into account potential hierarchies amongst requirements or how we might weigh them against each other. This in turn raises hard questions about an appropriate aggregation mechanism. And here I share Kolodny's scepticism about whether such a mechanism can be found.

This motivates scepticism about the strictly local view. The two cases illustrate that being able to make comparative judgments is an important part of our ordinary practice of attributing

[5] This problem is mirrored by the parallel to prima facie duties and duties proper that Brunero suggested. An aggregation mechanism that specifies how prima facie duties combine or compete is needed here too in order to settle on the duty proper. Furthermore, this is a problem for all views that allow for multiple local (i.e. non-global) rationality attributions, even if they are not construed as strictly as the strictly local view.

rationality. This in itself could constitute a rejection of the strictly local view. And even if we wanted to maintain it by allowing for some form of aggregation, we encounter difficult problems with aggregation. So is there a way of accommodating comparative judgments that lets us avoid these problems?

3.3 Discussing a Liberal View

This motivates the turn away from a strictly local to a more liberal understanding of rationality attributions. Maybe this can render local rationality attributions plausible. Adopting a more liberal view means that we construe the subset of attitudes that we evaluate for compliance with rational requirements more broadly. Could we not say that the *relevant* attitudes also include e.g. beliefs about evidence or correct reasoning? In this way, we might be able to account for the comparative attributions that would allow us to capture the differences between various ways of satisfying requirements and between vastly different agents. I will not say much in defence of this option, mainly because I am pessimistic about its feasibility. I take it that two desiderata for a plausible liberal view are (i) having a principled way of deciding which attitudes to include in the broader set of attitudes and (ii) remaining distinct from the strictly local view and its global counterpart, which would evaluate the entire set of attitudes for compliance with requirements. I am sceptical that these two desiderata can be met.

Turning to (i), how can we decide in a principled way which attitudes are to be included? In our example of the means-end incoherent agent who also believes that she lacks sufficient evidence for her means-end belief, it seems straight-forward that this belief should be included in the set of relevant attitudes. But this would not be the result of a principled decision rule but rather of a specific presupposed result that we have in mind (i.e. that beliefs about insufficient evidence can excuse means-end incoherence). There might be other cases of means-end incoherence where beliefs about feasibility are more important. So these should be included as well. So it seems as if we can only decide on the relevant attitudes on a case-by-case basis. And as far as I can see, there is no way of coming up with a suitably principled and general decision rule that can guide our selection of relevant attitudes. Absent such a decision rule, our choice to include one attitude rather than another in the set of relevant attitudes that is evaluated on a liberal view is arbitrary.

This creates a problem for (ii). It follows that we also have no principled reason to only include one or a few additional attitudes in our set that is to be evaluated. This runs the risk of opening the floodgates to the inclusion of many more attitudes and might eventually lead to a set of attitudes that includes all or almost all of the agent's attitudes. In turn, this would mean that the distinction between the local and the global perspective has lost its bite – the two perspectives would have collapsed into one.

With this result in mind, another option would be to endorse the global view from the start.

This would mean that we would check *all* of the agent's attitudes for compliance with rational requirements whenever we make rationality attributions. This would avoid some of the previously raised problems. A global rationality attribution would not hold that the different ways of complying with *(Means-End Coherence)* are rationally on a par because the ensuing violations of other requirements are already taken into account.[6] We also would not arrive at the same attributions for the two vastly different agents. And since the global view does not partition the agent's attitude set into different parts of the playing field with different referees for each, there is no need for an aggregation mechanism that would take us from local to global attributions.

However, going fully global seems to be an overreaction. It would be extremely burdensome to always have to evaluate entire sets of attitudes whenever we want to attribute rationality to an agent. After all, these sets also include a large number of attitudes that are not central to the question of an agent's rationality – e.g. her beliefs about the merits of different kinds of cheese. Moreover, this task might be even more challenging, depending on how fine-grained we take attitudes to be and whether we want to include, say, auxiliary beliefs in our attitude sets as well.

In addition to these worries, I think that the global view would lose out on an important level of information. Whilst the strictly local view on its own did not seem like a satisfactory account of rationality attributions, it nevertheless carried some important information about an agent's rationality. That we do not *only* care about an agent's performance with regard to a small subset of their attitudes does not mean that we *never* care about their performance on the strictly local level. The preponderance of local coherence requirements in the literature and their intuitive appeal speak to this.

3.4 Discussing a Mixed View

Taking stock, we have seen that strictly local attributions alone cannot capture all aspects of rationality. We need the global perspective to somehow feature in our theory of rationality attributions. Relaxing the understanding of 'local' did not help either because that would only dilute the distinction between local and global attributions.

This motivates a mixed view: one that maintains the strict understanding of the local and global perspective but allows for both. The hope is that they can make up for their respective shortcomings this way.

And indeed, the mixed view fares rather well: since we would not be restricted in the amount of attitudes we can evaluate, the global perspective could account for those judgments that have a more global flavour, e.g. when comparing an agent who is perfectly rational except for one requirement violation with an agent who is perfectly irrational except for one requirement

[6] At this point, one might wonder whether such a global view really just is a version of the local view that is being applied over and over again – especially if one wants to maintain locality of requirements. I discuss this possibility in the next section.

satisfaction. What's more, the mixed view also provides an elegant way of accommodating the symmetry objection to wide scope views, without denying asymmetry intuitions.

One explanation of why we judge dropping the means-end belief as a less rational way of satisfying *(Means-End Coherence)* than intending the means is because we implicitly assume a global perspective. On this global perspective, additional attitudes enter the judgment. The agent might also believe that she has sufficient evidence for her means-end belief. And giving up beliefs that one deems well-supported by evidence is not rational. Similarly, we can explain why sometimes dropping the end seems like the most rational option. On a global perspective, an additional belief about the value or dis-value of the end can be taken into account, which can explain why dropping an intention that we do not deem very valuable is more rational than going through with it. Once we adopt the global perspective, our rationality judgments take into account all of an agent's attitudes and this can recommend different courses of action as most rational. To repeat Brunero's words, there is "no inconsistency in claiming both that (1) two ways of proceeding are on a par as far as instrumental rationality goes, but (2) considering the other attitudes of an agent, one way of proceeding is more rational than another" (Brunero, 2012: 129).

So far, the mixed view has proven attractive. However, the mixed view is fairly inflexible, at least on the characterisation that is suggested by Brunero. It jumps from one end of the spectrum, the strictly local one, to the other end of the spectrum, the global one. I have already argued against our attributions being *always local* and we have seen that they are *sometimes global*. But often, what we are interested in are rationality attributions for sets of attitudes that are located somewhere along the spectrum, between the two extremes. They are concerned with more attitudes than the strictly local subset but not with *all* of an agent's attitudes. The mixed view is not flexible enough to truly capture this. All it can do is jump from one extreme to the other.

This discussion lets us draw a useful lesson. The mixed view gets one thing right – we need both a local and a global perspective in our theory of rationality attributions, but ideally, we also want the resources to talk about the in-between space.

Chapter Four

A New Account of Rationality Attributions

In what follows, I introduce my account of rationality attributions. The previous discussions have brought out a number of issues and problems that we should aim to solve or avoid if we are going to suggest a new and better account of rationality attributions. These can be summarised as three desiderata:

First Desideratum: Flexibility We have seen that the local, the global and the intermediary perspectives carry information that we care about when making rationality attributions. My account should therefore be able to allow for the local and the global perspective but also for intermediary perspectives. This includes the ability to explain asymmetry intuitions (see Section 3.2.1).

Second Desideratum: Informativeness Our account of rationality attributions needs to allow for comparative attributions that capture the difference between various ways of satisfying requirements and between agents who differ significantly in their rationality. Since the mixed view might already be able to do so, whatever account I propose instead needs to allow for this as well (see Sections 3.2.1, 3.4).

Third Desideratum: Delimitation If we are going to allow for intermediary subsets to be evaluated for compliance with rational requirements, we need to avoid the shortcomings of the liberal view. In particular, we need a principled way of explaining why we settled on a particular subset of attitudes whenever we are concerned with the space between the strictly local and the global perspective (see Section 3.2).

In what follows, I will first lay out the account and refer to the desiderata along the way. My account suggests that we should think of rationality attributions as taking the form of 'S is

rational with regard to subset a_n of her attitudes and degree of robustness r'. I suggest that we should understand the commonly used locution 'S is rational' as shorthand for this more precise attribution.[1]

My account is characterised by three features: a) an explicit reference to a set of attitudes, b) a robustness measure and c) a question under discussion-mechanism. I will now develop my account with the help of these three features.

4.1 Reference to a Set of Attitudes

On my account, attitudes are central for rationality attributions. Attitudes can be assessed for rationality or irrationality. They are the locus of rationality, and agents are rational or irrational in virtue of the attitudes they do or do not have. So whenever we talk about the rationality of agents, we should think of this as talking about the rationality of some of their attitudes, in virtue of which the agent themselves can be called rational. This understanding of attitudes as the bearers of rationality explains the explicit reference to them (or sets of them) in rationality attributions.

So what do I mean by the phrase 'with regard to subset a_n of her attitudes'? I propose to think of the set and subsets of an agent's attitudes as nested circles. The outer circles correspond to bigger subsets than the inner circles, with every inner circle being a subset of the next outer circle. There are multiple smallest circles (in fact, as many as the agent has attitudes), each containing only one of the agent's attitudes. We can then extend these smallest circles by adding more attitudes and moving outwards. Some of these enriched, bigger circles might not be interesting for the agent's rationality. They might be random collections of attitudes, like an intention to book plane tickets, a belief that crimson is a dark shade of red and a belief that Pythagoras's theorem is true. The biggest, outermost circle represents the entirety of an agent's attitudes and includes all other subsets. Call these attitude subsets a_1, a_2, ..., a_n with a_1 referring to the innermost subset and a_n referring to the entire set of attitudes.

The property of rationality also reflects this set-subset relation. If an agent is rational with regard to a subset of her attitudes (say a_2), she is also rational with regard to the subsets that are included in this one (in this case, a_1). Being rational in a_{n+1} implies being rational in a_n but not vice versa.

So much for sets and subsets. The next idea that I have to clarify is what it means to be rational *with regard to a subset of attitudes*. I take this to simply mean that one satisfies all rational requirements that apply to that subset, where a rational requirement applies to a subset

[1] I do not wish to take a stand on the issue of whether we always have the more complex and precise attribution in mind when we use the shorter locution. For what it is worth, I think there is good reason to think so. But all I am committed to here is that this is is a helpful way to think about rationality attributions, and not necessarily that this is how we actually think about it.

of attitudes if the attitudes (and their absences) that it governs are elements of that subset. For example, *(Means-End Coherence)* applies to a subset containing an intention to an end, a means-end belief and the presence or absence of an intention to the means, but not to a subset containing the belief that p, and the belief that if p, then q, and the presence or absence of the belief that q. *(Modus Ponens Closure)* however would apply to this latter subset.

As we have previously seen (see Chapter 2), there is ample support for the idea that rationality is strongly connected to compliance with rational requirements. To re-iterate this stance, which is sometimes called 'the requirements-based account', take Broome (1999: 410) who claims that "a large part of rationality consists in conforming to normative requirements". If rationality is in large parts a matter of complying with rational requirements, then we only need to check whether an agent complies with the relevant rational requirement(s), given the subset in question, which allows us to then simply read off rationality attributions.

Let me note at this point that this feature of my rationality attributions account meets the first desideratum – flexibility. There are no a priori restrictions on the selection of the subset that the attribution is attached to. It can be as small or big as we want, thereby enabling us to make attributions at the local level (where the subset in question would be very small), at the global level (where the subset would in fact correspond to the entire set of attitudes) and anything in-between, as would be specified by the explicitly mentioned a_{n}. This allows my account to accommodate the symmetry objection in very much the same way as the mixed view: we can maintain that some of the more questionable ways of satisfying *(Means-End Coherence)* would not be rational on a broader perspective, i.e. with regards to an attitude set a_{n} that includes many or all of an agent's attitudes, whilst also holding that all ways of satisfying the requirement would be rational with regard to a much smaller, local subset a_{n} – we are simply evaluating the agent at different subsets.

To see how this plays out when applied, consider again our initial example of Sara, who intends to submit her paper today, believes that this necessitates skipping lunch, but does not intend to skip lunch. We have noted that there are at least two different verdicts that seem plausible here: we might think that Sara is irrational in not intending to skip lunch because this makes her means-end incoherent. But once we took into account additional attitudes of hers, like a belief that regular lunch breaks should be prioritised in the interest of self-care, not intending to skip lunch could be seen as rational. The account of rationality attributions that I have developed so far gives us the resources to better deal with this case. We can now see that both these verdicts are adequate and are not necessarily in tension with each other because the attributions simply refer to different subsets of Sara's attitudes. We can maintain that Sara is not rational with regard to the subset of her attitudes that contains her intention to the end, her means-end belief and the absence of her intention to the means. And we can also maintain that she *is* rational with regard to the subset of her attitudes that includes the absence of her intention to skip lunch and the belief that not skipping lunch should be a priority.

4.2 Robustness

Why not stop here? One might suggest that the notion of sets and subsets already provides a sufficient level of detail to construct a plausible view. I am not opposed to these ideas. However, I think that not only can we make our rationality attributions more flexible but also more informative if we also introduce the notion of *robustness.*

Robustness is a modal notion: if we *were* to assess an agent's rationality with regard to a larger, further out subset of her attitudes and she would still qualify as rational there, her rationality *at the initial point of evaluation* is more robust than if she did not qualify as rational with regard to the larger, further out subset.

Robustness makes the view more complicated. But this is a price I am willing to pay because it allows us to include multiple points of view in the same rationality attribution. Not only do we assess the agent relative to the very narrow focus of a particular subset of her attitudes, we also appeal to how she is doing overall by including robustness in our rationality attribution. This has the additional advantage of accounting for the second desideratum – informativeness. Before I explain how, let me lay out the mechanics of the robustness parameter in more detail.

Robustness is measured by a robustness parameter r which reflects the distance from the outmost subset in which the agent does not violate any requirements to the outmost subset overall (i.e. the entire set of attitudes). Hence, a smaller r-value corresponds to a higher degree of robustness, where r=0 corresponds to maximal robustness and ideal rationality. As an example, take an agent whose attitudes can be represented as four subsets, a_1 – a_4. If we evaluate the agent at the smallest subset, a_1, and she satisfies all rational requirements that apply to her at a_1, we make a rationality attribution with regard to that particular subset. The robustness of that attribution depends on how far out we could go (i.e. how many more attitudes we could take into account), whilst still attributing rationality to that agent. In this case, r will take a value between 3 and 0, counting inwards from the outermost set (a_4) to the innermost set (a_1). If we were to take into account the attitudes that are elements of a_2 and if the agent were to violate a requirement that applies to her at a_2, our initial rationality attribution would receive the value of r=3 which corresponds to the lowest degree of robustness. In the case where she also satisfied all the requirements that apply to her at a_2, we would have r=2, i.e. she would be more robustly rational than in the previous scenario. If we could take into account all of her attitudes without encountering a violation of applicable requirements, our agent would be maximally robust and ideally rational (r=0).

As an example, take a means-end inconsistent agent like Sara, who intends to submit her paper today, believes that this necessitates skipping lunch but does not intend to skip lunch. We have seen that on wide-scope versions of *(Means-End Coherence)*, Sara has three ways of escaping this means-end incoherent state: she can drop her belief, drop her intention, or form the intention to skip lunch. As it stands, she would be rational in all three cases. In the first case, we could say

that Sara is rational with regard to the subset of her attitudes that consists of her intention, the absence of her belief and the absence of her intention to the means. Call this a_1. In the second case, we could attribute rationality to her with regard to the subset that consists of the absence of her intention, her belief, and the absence of her intention to the means. Call this b_1. In the third case, the attribution would be with regard to the subset consisting of her intention, her belief, and her intention to the means. Call this c_1. So far, this is just the picture developed in the previous section.

Let us now add robustness. Presumably, the robustness of the attribution with regard to c_1 is more robust than the ones with regard to a_1 or b_1. That is because in the latter two cases, it is unlikely that we are able to still attribute rationality, once we leave these narrow subsets behind and move out to bigger subsets. Take the first case where Sara drops her means-end belief in order to satisfy the requirement. If we now consider a bigger subset a_2, which contains the same attitudes as a_1 and in addition the belief that submitting comparable paper drafts in similar stages of completion has called for skipping lunch in the past, we can see that we are no longer able to make a rationality attribution. Here is why: next to *(Means-End Coherence)*, an additional rational requirement now applies, namely one that requires agents to conform their beliefs to the available evidence.[2] Sara still satisfies *(Means-End Coherence)*, but violates this other requirement at a_2 and so cannot be attributed rationality with regards to a_2. Note that this does not change the fact that she can be attributed rationality with regards to a_1. This means that the r-value of the initial attribution at a_2 is n-1 – we cannot move even to the next largest subset without encountering a violation of a rational requirement. The initial rationality attribution is minimally robust.

The story looks similar for the second case. Sara is rational with regard to b_1 since she satisfies the rational requirement that applies to her at this subset. But once we move to a slightly bigger subset, b_2, which also contains, say, a belief that it is crucial for her career that she intends to submit the paper today, an additional requirement plausibly applies to her, namely a version of *(Reasons-Responsiveness)*. And Sara now violates this requirement by not intending to submit the paper today. We cannot attribute rationality to Sara at the next biggest subset, b_2, and so the initial rationality attribution is minimally robust, with an r-value of n-1.

But the third case is different. Sara is still rational with regard to the initial subset c_1, since she satisfies the applicable *(Means-End Coherence)* by intending to skip lunch. But if we now move out to a bigger subset, e.g. one that also contains the belief that it is crucial for her career that she submits the paper today, Sara also satisfies the additional rational requirement that applies to her at this bigger subset, namely *(Reasons-Responsiveness)*. And so we can attribute rationality to her at the bigger subset c_2, too. This means that r-value of the initial attribution is at most n-2, or even smaller. Since this r-value is smaller than the other ones, this rationality

[2] We could tell a similar story by appealing to higher-order beliefs, namely if a_2 includes her belief that her means-end belief is sufficiently supported by her evidence.

attribution is more robust than the other two.

This demonstrates how adding robustness to our rationality attributions meets the second desideratum. Robustness makes it possible to acknowledge the fact that there can be different ways of complying with rational requirements that are equally rational in one sense, but also differ in their rationality in another sense. They are equally rational in that they would all justify a rationality attribution (assuming that all applicable rational requirements are satisfied). But the way in which they are different is accounted for by the different r-values that attach to the attributions. As we have seen in the previous example, dropping a means-end belief or dropping an intention to an end warrant a rationality attribution, but this attribution is less robust than the attribution that we can make if an agent forms the intention to the means. Robustness allows us to compare different ways of complying with rational requirements, be it intrapersonal, when we consider the different options available to one and the same agent, or interpersonal, when we compare two agents that respond to a rational conflict in different ways.

Robustness supplies us with an improved understanding of this difference, and with a precise way of cashing it out: the sense in which one way of complying with the requirement is 'better' or 'more rational' is that it is more robustly rational, or, put differently, would make for a more robust rationality attribution. In this way, we can also account for asymmetry intuitions – the way in which multiple ways of satisfying *(Means-End Coherence)* seem asymmetrical, or different in their rationality, is to be understood in terms of different robustness values. Whilst all options that lead to the satisfaction of *(Means-End Coherence)* would warrant a rationality attribution at this subset of attitudes, some are likely to make this attribution less robust because they might lead to violations of requirements at bigger subsets (e.g. dropping the means-end belief). And we can also maintain that there is a difference between our two agents, one who is almost ideally rational, and one who is almost entirely irrational, despite them satisfying the same requirement – the rationality attribution that is warranted in both cases would carry a different robustness-value, which can account for the difference.

At this point, let me caution against understanding robustness as a measure of degrees of rationality, despite what might be suggested by locutions like 'more rational' – at least not in the standard way. The account of rationality attributions I propose here sticks to the binary rational/not rational divide. Robustness only enters the picture once we can make a (positive) rationality attribution and hence cannot be understood as modelling degrees of rationality. The robustness of someone's rationality presupposes them being rational, so to say. What robustness does is to give us a way of understanding the way in which one way of compliance is 'better' than another without creating the need to introduce complicated and somewhat controversial degrees of rationality.[3] So the only way in which we could understand robustness as measuring

[3] At this point, the sympathiser of gradual rationality might wonder whether we could not use the size of subsets to model degrees of rationality. The thought might be that the bigger the subset that is the basis for a rationality attribution, the more rational the agent and vice versa. But again, I think we should resist this temptation.

degrees of rationality is at the level of attitudes, not agents – they could be seen as representing the varying degrees in which ways of satisfying requirements (and hence attitudes) can be more or less rational. For what it is worth, I prefer sticking to the language of 'robustness' rather than the language of 'degrees of rationality' because it has less potential for misleading us. But if one wanted to understand robustness in this way, this would be in keeping with my general commitment to focussing on attitudes as the bearers of rationality – as long as we are clear on the fact that we are not talking about the degrees of overall rationality of agents.[4]

4.3 Which Set? The QUD mechanism

Until now, we have discussed two main features of the account of rationality attributions proposed here: reference to a subset of attitudes and robustness. A question I have ignored so far is how we choose the particular subset that then forms the basis for the rationality attributions. In short, I take this to be the result of different questions about rationality that arise in different situations and correspond to different sets of attitudes to be evaluated. To further cash out this idea, I draw on the notion of 'questions under discussion' as it is used in semantics and linguistics to describe the way in which questions can partition informational space.

4.3.1 Questions Under Discussion

The notion of 'questions under discussion' (QUD) can help to shed light on how different sets of attitudes ultimately lead to different rationality attributions. In particular, different sets of attitudes under consideration correspond to different questions under discussion. But first, let me introduce questions under discussion in a bit more detail.

Questions under discussion define the goal of inquiry, or the topic of a given conversation. In doing so, they narrow down logical space to a smaller partition, namely the one that is of interest in a given context. For example, if we are discussing which house is the nicest in the neighbourhood, the area of logical space that we are focussing on is an area that contains information about the houses in the neighbourhood, about aesthetic value, personal preference etc. Other areas of logical space, e.g. areas containing information about the nature of free will, are not relevant for this particular question under discussion.

One problem of this approach is that attitude subsets are very arbitrary. It might be possible to add additional attitudes to the set in question. This would make the set bigger, which, on the suggestion explored here, would then mean a higher degree of rationality. But this strikes me as odd, since these attitudes could be more or less trivial, and might not trigger additional rational requirements. So the rational performance of the agent would remain unchanged; yet, they would obtain a higher degree of rationality for free. A similar problem would arise for the opposite case, if we remove some (trivial) attitudes from the initial attitude set. If this does not change the applicable requirements, the agent's rational performance would remain unchanged and yet result in a diminished degree of rationality. In addition to these problems, different people might just have different amounts of attitudes overall. On the suggested approach, this would mean that they would have different maximal degrees of rationality they could possibly obtain. This strikes me as odd.

[4] For a recent treatment of non-ideal rationality and degrees of rationality in a Bayesian framework, see Staffel (2019).

The way in which questions are thought to narrow down logical space is by picking out subsets of propositions which would be adequate answers to the question. As Hamblin (1973: 48) puts it, "[p]ragmatically speaking a question sets up a choice-situation between a set of propositions, namely, those propositions that count as answers to it." Roberts (2012: 6) calls this set of propositions the *q-alternative set* or *q-alt* set. In what follows, I rely on her account.[5] In order to actually answer the question, we need to check the propositions in the *q-alt set* for their respective truth and falsity. These truth values are provided by the facts.

4.3.2 QUD applied to Rationality Attributions

After this brief excursus into the semantics of questions, let us see how the notion of QUD can be put to work in our account of rationality attributions. I have already suggested that the different sets of attitudes that we are interested in, which ultimately form the basis of different rationality attributions, correspond to different questions under discussion. More specifically, the questions under discussion we are dealing with here are polar questions of the kind "Is S rational with regard to subset a_{n} of her attitudes?". This means that the *q-alt* set for this question only consists of the proposition we are inquiring about, i.e. "S is rational with regard to subset a_{n} of her attitudes". A complete answer to the question then assigns a truth-value to this proposition: it is either true or false that S is rational with regard to subset a_{n} of her attitudes.

Normally, we just have to look at the facts to determine the truth-values of the propositions in the *q-alt set*. In the case of rationality, we now have to look at the rational requirements to determine the truth-value of the proposition(s) in the *q-alt set*, since rationality is a matter of conforming to rational requirements (see Chapter 2). Which requirements we have to check depends on the set a_{n} of the agent's attitudes. For example, a means-end coherence requirement would not need to be checked if the set of salient attitudes only contained beliefs. Compliance or non-compliance with the applicable requirements then determines the truth-value of the proposition in the *q-alt set*, i.e. whether S is rational with regard to subset a_{n} of her attitudes.

I have already mentioned that I take different questions under discussion to correspond to different sets of attitudes. Depending on the question under discussion, a different set of attitudes is salient, which then leads to different applicable rational requirements. Depending on satisfaction of these requirements, we can then attribute rationality to the agent with regards to the particular set of attitudes. To sum up, the mechanism takes us from questions under discussion and their corresponding attitude sets to rational requirements, and finally to rationality attributions.

4.3.2.1 Selecting QUDs

As for the first step – questions under discussion and their corresponding sets of attitudes – these can be as varied as our interests when we evaluate agents for their rationality. We are free to

[5] This is not the first work on questions in semantics and linguistics. For earlier accounts that Roberts references throughout, see e.g. Hamblin (1958; 1973), Higginbotham (1996) and Groenendijk & Stokhof (1984).

draw the set a_{n} in whichever way we want when we consider the QUD "Is S rational with regard to subset a_{n} of her attitudes?". This flexibility is a major advantage of my account. But for what it is worth, I think that there at least two ways in which we can be primed to focus on certain QUDs and the corresponding attitude sets: explicit mentioning and markers.

Explicit Mentioning Often, the way in which a case or an agent's situation is presented primes us to consider a specific question under discussion and the corresponding attitude set. This priming can happen directly through markers, such as signal words or phrases, or indirectly, by the fact that only some of the agent's attitudes are explicitly mentioned. To illustrate this sort of priming, consider again the following example, found in Brunero (2012: 236):

> "Candice decides to go to the post office this afternoon to send out some mailings, but on the way there, she gives up on this end and decides to go buy groceries instead. But on the way to the market, she yet again trades in this end for another: going to hang out with her friend David. But on the way to David's house, she once more changes her mind and intends to spend a relaxing afternoon at home, but by the time she gets home the afternoon is gone and she's accomplished nothing. We're inclined to accuse Candice of some kind of irrationality. And, since Candice failed to follow through on any of her ends, we're inclined to accuse her of being instrumentally irrational."

I am assuming that the attitudes that are explicitly mentioned in this example do not exhaust the entirety of Candice's attitudes. Surely, Candice's mental life is a lot richer as could be captured by a mere four attitudes. But the fact that these ones are being explicitly mentioned primes the reader to think that they are the ones of interest for this particular evaluation. This suggests the QUD "Is Candice rational with regard to subset a_{n} of her attitudes?", where a_{n} includes the explicitly mentioned attitudes, i.e. an intention to go to the post office, an intention to buy groceries, an intention to visit David, an intention to spend the afternoon at home, and their respective absences. When deciding whether Candice is rational with regard to this subset of her attitudes, we would then need to check whether she satisfies the rational requirements that apply to her at this subset. Presumably, she would violate some Persistence of Intention requirement, and so it is likely that we would not attribute rationality to her in this instance.[6]

To see how the explicitly mentioned attitudes can prime different QUDs, consider the following, more fleshed out version of Candice's case:

> Candice decides to go to the post office this afternoon to send out some mailings, but on the way there, she remembers that her partner had promised her this morning to

[6] For an example, see Broome (2013: Ch. 10).

> take care of the mailings. She then gives up on this end. Nevertheless, she decides to go buy groceries, because the market is right around the corner from the post office and she has a habit of combining these two trips. But on the way to the market, she realises that it is Monday and that the market is closed on Mondays. So she trades in this end for another: going to hang out with her friend David, who lives close by and with whom she wanted to catch up for a while. On the way to David's house, she decides to send him a message to let him know she is coming, and upon seeing their earlier conversation, remembers that David is out of town this week. She once more changes her mind and intends to spend a relaxing afternoon at home, but by the time she gets home the afternoon is gone, but she is not too unhappy about this, because she also believes that sunny days like today should not be spent inside.

This more complete story suggests a different QUD, whose attitude set includes not only the four initial intentions, but also some beliefs (e.g. that her partner is sending the mailings, that the market is closed, that David is out of town, that sunny days should be spent outside). This in turn makes different rational requirements applicable, e.g. means-end coherence and reasons-responsiveness requirements, which she presumably satisfies. And so we might be inclined to attribute rationality to Candice, given this different QUD. Note that the different QUDs were not caused by a difference in Candice's situation, but rather by a difference in which attitudes the presentation of the case primed us to focus on.

Markers For an example of the other sort of priming that uses *markers*, consider Kolodny (2005: 554, my emphasis), who asks us to imagine the following exchange:

> "Look: I think you're nuts to believe in God, let alone that He created anyone. But *given that* you believe that God created all people equal, and given that you agree that people whose skin is a different colour from yours are people, you ought to believe that He created them equal too. It would be irrational of you not to."

This presentation primes us to only take into account the explicitly mentioned attitudes of the agent: a belief that God created all people equal, and a belief that people whose skin is a different colour from yours are people. Moreover, the locution 'given that' marks these attitudes as particularly important in this case. Now consider a different presentation, in which the 'given that' marker is absent and in which we are given more information about the believer's attitudes:

> Look: I think you're nuts to believe in God, let alone that He created anyone. And in fact, you seem to agree with me, because you have told me that you think that evolutionary theory is well-supported by the evidence and that you also find it hard to find fault with the big bang theory. But you also believe that God created all

> people equal, and you agree that people whose skin is a different colour from yours are people.

On this second presentation, we are primed to consider two additional attitudes of the agent as relevant, because they are being explicitly mentioned. This leads to a different QUD, which leads to different applicable requirements, which in turn might motivate a different rationality attribution, leaving it open whether it would be rational for the agent to believe that God created people with a different skin colour equal (given their beliefs that question the reliability of their belief in God).

Sometimes, the markers used will correspond to a sub-domain of rationality. Consider the familiar passage from Brunero (2012: 129-130, my emphasis):

> "[T]o say that giving up an end, intending the means, and giving up an instrumental belief are all on a par *as far as instrumental rationality goes* doesn't commit us to thinking that they are all on a par overall. Once we expand our focus beyond those attitudes *relevant to instrumental rationality*, one way of proceeding may be more rational than another. (For instance, if one believes that one has conclusive reason to intend the end, and believes that one has conclusive reason to hold the instrumental belief—and, in case it matters, one also believes that these two beliefs are themselves beliefs that one has conclusive reason to hold, and so on—and there are no other relevant attitudes bearing on this case, then intending the means is more rational than giving up the end or giving up the belief.) There is no inconsistency between saying that the two ways of proceeding are on a par as far as instrumental rationality goes, but, *overall*, one way of proceeding is more rational than another."

Phrases like 'as far as instrumental rationality goes' and 'relevant to instrumental rationality' suggest QUDs of the sort "Is S instrumentally rational?". The mechanism works slightly different in these cases. Such QUDs suggest a set of rational requirements as salient (here: requirements of instrumental rationality). But we still need the attitude set suggested by the presentation of the case, e.g. the explicitly mentioned ones, to determine which requirements are applicable. The mechanism then works in the familiar way. QUDs can attach to all sub-categories of rationality, e.g. "Are S's beliefs closed under implication?" or "Is S reasons-responsive?". These correspondences between QUDs and sub-domains of rationality should be thought of as short-hand for the more precise QUD which features the specific subset. This is convenient but should not trick us into thinking that these familiar questions are the more important questions. Instead, we should think of them as lucky coincidences and should embrace the more general template for questions under discussions, since it allows us to make much more specific rationality attributions, not all of which correspond neatly to sub-domains of rationality.[7]

[7] One might wonder why I, in keeping with the literature, focus on so few requirements, when there is obviously

The last marker I want to point out here is 'overall'.[8] The QUD "Is S rational overall?" corresponds to the entire set of her attitudes, hence creating the need to check all rational requirements that apply to S. Note that this marker interacts with the explicitly mentioned attitudes, because if we do not have any additional information about the agent, the overall (or global) set will just correspond to the set of attitudes we know exist, even though the agent might have additional attitudes that we simply do not know about. In this way, rationality attributions are contingent on the available information.

The preceding exploration shows that the account can also meet the third desideratum. The notion of 'question under discussion' and how these can arise through our interests or through primers in the presentation of the cases explains how we arrive at the attitude sets that our rationality attributions attach to.

Finally, let me illustrate how the QUD mechanism interacts with one of the other main features of this account, namely the explicit reference to a subset. For ease of exposition, I am leaving out the robustness-value, which has already been discussed at length in the previous section.

Take the initial Sara-case as an example. The case was presented in a way that marked as salient the following set of her attitudes: {intention to send off paper today, belief that skipping lunch is necessary for this, absence of intention to skip lunch}. Call this a_1. This corresponds to the question under discussion "Is Sara rational with regard to subset a_1 of her attitudes?", which in turn gives us the *q-alt set* {Sara is rational with regard to subset a_1 of her attitudes}. In order to determine whether this proposition is true, we need to check whether Sara complies with all rational requirements that apply to her with regard to this subset. As a guiding principle, the rational requirements that apply to one at a given subset are those which we can assess for compliance, because the subset features the attitudes that the requirement governs. In Sara's case, this plausibly includes a means-end coherence requirement but excludes, for example, a closure requirement on belief. It would also exclude a requirement about sufficient support by

more to coherence rationality than means-end coherence and closure. I think that there is a pragmatic explanation for this. These requirements and the corresponding conflict states they govern represent clear-cut and relatable cases. We have all been weak-willed before (thereby making us means-end incoherent) and we have all struggled to adopt a perhaps unwelcome belief that followed from our other beliefs (thereby making us violate closure). The intuitiveness of these requirements and corresponding conflict states becomes even more clear if we compare them to less intuitive coherence requirements like Bayesian updating requirements. In order to make sense of these, one first needs to master the concepts of conditional probabilities, which does not translate to everyday experiences as straight-forwardly as, say, the concept of a means-end belief. If I am correct in conjecturing that this explains the strong focus on these requirements, this also means that we should be careful not to interpret their ubiquity as indicative of some sort of special or privileged standing. These requirements, with their corresponding conflicts, sets of governed attitudes and QUDs, are just some amongst many others. At the very least, if we want to grant them special standing, we should not do so for the sole reason that they have been the focus of inquiry.

[8] This marker obviously interacts with the question whether rationality attributions are local or global in extent. Whenever we encounter the 'overall' marker, the result will be a global rationality attribution, i.e. one that takes into account *all* of an agent's attitudes, regardless of the particular understanding of 'global' one adopts. This is another instance in which the flexibility of my account is an advantage – we do not have to commit ourselves to either a local or a global view of rationality attributions, and can also accommodate different understandings of 'global'.

reasons, since the relevant attitudes (e.g. a belief about having sufficient reason to submit the paper today) do not feature in a_1. Since Sara is means-end incoherent, and hence violates this requirement, we can answer the question under discussion negatively, thereby not attributing Sara rationality with regard to subset a_1 of her attitudes.

If Sara responds to this by giving up her intention to submit the paper today, we face a different question under discussion. That is because a different set of attitudes has been marked as salient: {absence of intention to skip lunch, belief that skipping lunch is necessary for sending off the paper today, absence of intention to send off the paper today}. Call this set a_2. The corresponding question under discussion now is "Is Sara rational with regard to subset a_2 of her attitudes?". In order to answer this, we need to check whether Sara complies with the rational requirements that apply to her at a_2, i.e. a means-end coherence requirement. Sara satisfies this requirement, and so we can answer "Sara is rational with regard to subset a_2 of her attitudes".

In yet another version of the example, a different set of attitudes was marked as salient, namely: {absence of intention to skip lunch, belief that skipping lunch is necessary for sending off the paper today, absence of intention to send off the paper today, belief that one has conclusive reason to intend to send off the paper today, belief that one has conclusive reason for the instrumental belief} – call it a_3. This again corresponds to a different question under discussion, namely "Is Sara rational with regard to subset a_3 of her attitudes?". In order to answer this, we need to check for compliance with the rational requirements that apply to her at a_3. Because a_3 includes additional attitudes, an additional requirement applies to her, namely one that demands conforming one's attitudes to one's reasons. And by giving up a belief that she takes herself to have conclusive reason for, Sara violates this requirement. So it is not the case that she satisfies all rational requirements that apply to her at a_3, which leads us to answer "Sara is not rational with regard to subset a_3 of her attitudes."

In the same way, the account can help us to make sense of the other introductory examples that highlighted the need for an account of rationality attributions in the first place. We have just seen that whether Sara is rational in taking a lunch break will depend on the set of salient attitudes and the corresponding QUD. If the salient attitudes only include her intention to send off the paper, her means-end belief and the lack of her intention to skip lunch, the answer to the corresponding question under discussion is going to be no, since she violates a (means-end coherence) rational requirement that applies to her. If the set of salient attitudes consists of the absence of an intention to skip lunch and the belief that not skipping lunch is a good habit to adopt, then the answer might be yes, given the different question under discussion and the different applicable rational requirement about reasons-responsiveness. The different possible evaluations of Sara in the second case and of Muriel, the aspiring axe-murderer, can equally be explained by different sets of salient attitudes, corresponding to different questions under discussion, which lead to different salient requirements and hence different rationality attributions.

With this in place, let us briefly reconsider the desiderata I stated at the beginning of this chapter and how the account I developed meets them. Turning to the first one, flexibility, my account makes no restrictions on the size of the attitude set that a rationality attribution is relative to. This set could be as big as the entire set of an agent's attitudes (making the rationality attribution into a global one), it could be a very small set (making the attribution into a (strictly) local one), or anything in-between. In this way, my account is an improvement on the competing understandings of attributions as either global or local. By replacing this dichotomy with the much more flexible explicit reference to particular subsets, we avoid the problems of the global and the local view respectively (see Chapter 3), whilst maintaining their advantages and also accommodating the symmetry objection.

With regards to the second desideratum of informativeness, the addition of a robustness measure enables my account to make the kinds of intra- and interpersonal comparisons that were missing on the competing views of rationality attributions explored in the previous chapter. The robustness measure is able to account for the perceived differences between multiple ways of satisfying rational requirements. This allows us to make nuanced comparisons of various options available to a single agent, or of the options taken by different agents.

The third desideratum of delimitation is met by the interplay of questions under discussion, primers and their respective sets of attitudes. As a result, the selection of attitude sets, which then become the basis for a rationality attribution, becomes non-arbitrary. This is another advantage of my account when compared to the liberal view, which as we saw, had no resources to limit the set of attitudes in question.

4.4 Closing the Loop: Back to Rational Requirements

Despite all these advantages, one might wonder at this point how much progress we have actually made. Whilst it is true that my account allows for more flexible and informative attributions, it still relies on requirements in a way that might be deemed problematic. I have laid out how different QUDs correspond to different sets of attitudes which results in different applicable requirements. And so it seems like this step still requires the authoritative list of correct rational requirements. But if the discussion in Chapter 2 has shown us anything, it is that this search is most likely to be futile. This pessimism motivated switching strategies and turning straight to rationality attributions in the first place. So have we gone through all this trouble only to end up back where we started?

This would be too extreme a reaction. We can still accept my account of rationality attributions whilst deferring the search for the authoritative list of requirements to first order theorists. This does not take away from the insights that my account provides regarding the relationship between rational requirements, attitudes, and rationality attributions. And in fact, this might even be a strength of the account: we can treat the phrase 'applicable requirements' as a placeholder for

whichever theory of rational requirements one deems correct. The more general claims about how requirements feed into attributions then hold in a theory-neutral way.

In a similar vein, it should be noted that my account is compatible with all options that the various choice points explored. Deciding on one option rather than another obviously has consequences for the resulting rationality attribution. But other than this unsurprising interaction, my account does not entail further commitments regarding the choice points. For example, whether rational requirements take narrow or wide scope is a question that my account of rationality attributions does not take a stand on. In both cases, the same requirements would be applicable, given the same set of attitudes in question, since narrow and wide scope requirements govern the same attitudes. Scope only makes a difference when we decide whether the agent satisfies or violates the applicable requirements – for example, in dropping a means-end belief they would satisfy a wide-scope version of *(Means-End Coherence)* but not a narrow scope version. So even though scope might have an impact on which rationality attributions would be warranted, there is nothing in my account that would commit us to a narrow or wide scope reading of rational requirements.

Turning to the choice point of local vs global requirements, my account is again compatible with both options. If we were to adopt a global understanding of rational requirements, this would have the consequence that nothing short of the entire set of an agent's attitudes would trigger them. If we were to adopt a more local understanding of requirements, then smaller subsets would also be able to trigger rational requirements (namely the ones that govern the attitudes that the subset contains).

The situation looks very similar when we turn to the choice point of process vs state requirements. This is not surprising, since this choice point is related to the previous one. I have pointed out at various points that a commitment to local requirements might be difficult to square with a process understanding of requirements, since the latter necessitates taking into account more attitudes (past and present) which can be difficult to accommodate, given the restrictions on attitude numbers that a (strictly) local understanding of requirements entails. But this is a relation that exists within the debate about the correct *requirements*. And since my account is neutral about whether the local or global understanding is correct, there is no principled reason why it could not allow for both state or process requirements. We might only encounter a restriction here if we had previously committed ourselves to a (strictly) local understanding of requirements, which might entail a commitment to state requirements. But again, the source of this commitment would be the first-order theory about correct rational requirements, and not my account of rationality attributions.

However, turning to the question of coherence vs substantive requirements, there is one limitation of my account that I want to acknowledge: because I identify attitudes as the bearers of rationality, I can only allow for rational requirements that can be expressed in terms of attitudes. This might be difficult to square with a substantive view of rationality that includes requirements

on how to respond to objective reasons. That is because these reasons need not be reflected in an agent's attitudes at all. Agents might be entirely unaware of the relevant objective reasons. For example, consider an agent who finds herself in a gin-and-petrol scenario (Williams, 1981), where she is unaware of the objective reason not to drink the liquid in front of her because it is in fact petrol. The mechanism I suggest for making rationality attributions starts with a set of attitudes and then designates those requirements as relevant that govern the set of attitudes in question. But in a gin-and-petrol scenario, there is no attitude that would reflect the objective reason not to drink the liquid (since the agent is unaware of it being petrol), and so a potential rational requirement that requires us to respond to objective reasons would not be designated as relevant. This means that we would not be able to make rationality attributions if we adopted a view of rational requirements as concerned with responding to objective reasons.

However, even once we acknowledge this limitation, there still is a way for my account to accommodate more substantive requirements, namely if we can express them in terms of attitudes, e.g. beliefs about reasons. In this way, the connection between an agent's attitudes and the rational requirements that govern them could be re-established. What is more, this limitation of my account to propositional attitudes as input actually goes hand in hand with my proposed understanding of the Normativity of Rationality (see Chapter 8), which is also deeply rooted in an agent's attitudes. So there is no worrisome mismatch between the descriptive part of the account and the claimed normative status of rationality to which it is connected.

Chapter Five

Saying No to Contextualism

We have just seen that my proposed account of rationality attributions lets us reconcile some seemingly contradictory rationality verdicts – e.g. when it seemed both rational and not rational for an agent to not intend to skip lunch, or to drop a means-end belief. The fact that it allows us to accommodate seeming disagreements might suggest that the account actually uncovers something fundamental about the property of rationality. In particular, one might feel reminded of how appeals to *Epistemic Contextualism* are often used in similar ways to resolve disagreements about whether a given property, in that case knowledge, obtains. Given this parallel, one might be tempted by the idea that rationality is context-sensitive, too. For the purposes of this chapter, I will understand Epistemic Contextualism as the view that the truth-conditions of knowledge attributions shift with context, with the shifting parameter being the required epistemic standard. In what follows, I explore the possibility of adopting a contextualist view of rationality but conclude that we can ultimately resist the pull of context-sensitivity.

There are a number of reasons for why resisting contextualism might be desirable. Even fairly established and well-developed versions like Epistemic Contextualism are controversial. For example, it has been pointed out that existing linguistic models for other uncontroversially contextual terms like 'tall' cannot be straightforwardly applied to 'knows' (Stanley, 2004). A contextual view of rationality would most likely inherit this problem. So if my view need not commit itself to contextualism, it could avoid this problem. What is more, it could also avoid the substantial claims about the very nature of the property of rationality that a contextual view would entail, which are notoriously difficult to defend. Nevertheless, discussing the option of contextualism about rationality will prove useful. It will allow us to further highlight one of the main advantages of the account, namely its ability to accommodate some disagreements and to distinguish between actual and only seeming disagreements.

5.1 The Parallel with Epistemic Contextualism

In laying out my account, I rely on cases that seem to license incompatible rationality attributions, such as 'Sara is rational in dropping her means-end belief' and 'Sara is not rational in dropping her means-end belief'. This might seem reminiscent of the data used in defences of Epistemic Contextualism, which often rely on cases that seem to justify both a knowledge attribution and a knowledge denial, too.[1] As an example, consider the famous bank cases (DeRose, 2009: 1-2):

> *Bank Case A.* My wife and I are driving home on a Friday afternoon. We plan to stop at the bank on the way home to deposit our paychecks. But as we drive past the bank, we notice that the lines inside are very long, as they often are on Friday afternoons. Although we generally like to deposit our paychecks as soon as possible, it is not especially important in this case that they be deposited right away, so I suggest that we drive straight home and deposit our paychecks on Saturday morning. My wife says, 'Maybe the bank won't be open tomorrow. Lots of banks are closed on Saturdays.' I reply, 'No, I know it'll be open. I was just there two weeks ago on Saturday. It's open until noon.'
>
> *Bank Case B.* My wife and I drive past the bank on a Friday afternoon, as in Case A, and notice the long lines. I again suggest that we deposit our paychecks on Saturday morning, explaining that I was at the bank on Saturday morning only two weeks ago and discovered that it was open until noon. But in this case, we have just written a very large and very important check. If our paychecks are not deposited into our checking account before Monday morning, the important check we wrote will bounce, leaving us in a very bad situation. And, of course, the bank is not open on Sunday. My wife reminds me of these facts. She then says, 'Banks do change their hours. Do you know the bank will be open tomorrow?' Remaining as confident as I was before that the bank will be open then, still, I reply, 'Well, no, I don't know. I'd better go in and make sure.'

DeRose (ibid.) argues that these cases license a knowledge attribution of the form 'S knows that the bank will be open on Saturday' for case A, and a knowledge denial of the form 'S does not know that the bank will be open on Saturday' for case B, despite S's epistemic position being the same in both cases. Cases of this sort are used to motivate Epistemic Contextualism. When S utters 'I know that the bank will be open on Saturday' in Case A, they do so in a context with low epistemic standards, making it the case that the corresponding knowledge attribution would be true. But if they utter the same sentence in Case B, they do so in a context with high epistemic standards, meaning that they need to be in a "*very* strong epistemic position with

[1] I only deal with 'traditional' Epistemic Contextualism here, leaving variants like Relativism and related views like Subject-Sensitive Invariantism, or Contrastivism largely to the side.

regard to [the proposition in question]" (ibid.: 3). Arguably, this condition is not met in Case B, which then renders the knowledge attribution 'S knows that the bank will be open on Saturday' false for Case B. Importantly, the epistemic position of the agent is the same in both cases – they have access to the same evidence, the same reasons to trust or doubt the reliability of their evidence etc. This is why the two statements seem incompatible initially. But by claiming that knowledge attributions are contextual in this way, one can maintain both the knowledge attribution and the denial, since they have to meet different truth conditions.

I have claimed that my account can do similar work in cases of apparently incompatible rationality attributions. To recall one of the examples, it seemed reasonable to say that 'Sara is rational in dropping her means-end belief' and also that 'Sara is not rational in dropping her means-end belief'. I proposed to reconcile these two attributions by making explicit the different sets of attitudes that underlie the respective attributions, which correspond to different questions under discussion. But given everything we have just heard about Epistemic Contextualism, one might ask whether I should endorse contextualism about rationality instead? Following this line of thought, we could then explain and maintain the seemingly incompatible rationality attributions by appealing to differences in context, which correspond to different rational standards that Sara would need to meet in order to count as rational, which in turn would create different truth conditions for each attribution.

Moreover, I have also argued that features of the case, rather than a difference in rational status on the part of the agent, play an important role in whether a rationality attribution is warranted (namely because it leads to different questions under discussion). This looks like another parallel to Epistemic Contextualism, since this too postulates sameness of (epistemic) position of the agent, and appeals to external differences in the cases. Given these parallels, why not instead opt for contextualism about 'is rational' to explain the data? This is the question I investigate in the following sections. I ultimately conclude that we need not endorse contextualism here because one, the evidence for the context-sensitivity of 'is rational' is not conclusive and two, there are alternative non-contextual explanations for the parallels.

5.2 Testing for Contextualism – The Disagreement Test

Further support for taking a contextual approach to 'is rational' might be derived from context-sensitivity tests, which test for the context-sensitivity of an expression. One such example is the Disagreement Test, as proposed by Cappelen & Hawthorne (2009: 56-60). So let us see whether applying this test to rationality pushes us further towards endorsing contextualism about rationality.

The test focusses on whether speakers in different contexts could be seen as agreeing or disagreeing in yet another context. Cappelen & Hawthorne propose three different formulations of this test, but I will only be concerned with the first one, which is phrased in terms of agreement:

> "Let u be a sincere utterance of S by A in C and u' a sincere utterance of 'not-S' by B in C'. If from a third context C'' they cannot be correctly reported by 'A and B disagree whether S', then S is semantically context sensitive." (Cappelen and Hawthorne, 2009: 55)

As an example, consider an indexical, like 'here'. Take the sentence 'Elsa is here' (S). One person utters this sentence in their kitchen (C), with Elsa standing next to them. Another person utters its negation 'Elsa is not here' in their office (C'), without Elsa being present. Plausibly, it would be false to report for a third person in the comfort of her living room (C'') that the other two disagree about whether Elsa is here. If they shared the same context, e.g. were both in the kitchen with Elsa, they would agree that Elsa is here. And in the context of the office that Elsa has never set foot in, they would agree that 'Elsa is not here' is true. Since the two people cannot be correctly reported to disagree by someone in a third context, the test provides evidence for the context-sensitivity of the expression 'is here'.

Running the test for a context-insensitive expression like 'is wet' can provide further illustration. Imagine someone utters 'Water is wet' in the context of taking a walk in the rain. Someone else utters 'Water isn't wet' in the context of swimming the English Channel. Now, it seems correct for a third person, in the context of, say, doing the washing up, to say that the other two do in fact disagree about whether water is wet. Since 'is wet' fails the Disagreement Test, the test does not provide any evidence for its context-sensitivity.

When we apply this test to 'is rational', it seems like rationality might actually pass the Disagreement Test, which in turn would further motivate its context-sensitivity. Consider the following conversation, inspired by a case presented in Brunero (2010: 48):

> "So I'll see you at the lecture this afternoon?" – "I don't think so; I think I'll stay at home." – "But Annie, you just told me that you think it is important that you attend the lecture. I remember you saying that you had all the reasons to attend the lecture!"

In this context, a speaker is primed by the explicitly mentioned attitudes (an intention to attend the lecture and a belief about conclusive reasons). This makes it likely that they will conclude that Annie is not rational. However, another speaker might come to conclude the opposite in a different context, e.g. one where additional attitudes of Annie's are brought to their attention. Consider the following continuation of the conversation:

> "I know, but I also told you that I am not so sure about this belief – in fact I worry that it is only my puritanical upbringing that makes me believe that I should always do what I'm told and attend the lectures."

This context makes an additional attitude salient, suggesting that a speaker in this context might conclude 'Annie is rational' when she does not intend to attend the lecture. After all, she has doubts whether this intention is sufficiently backed up by reasons after all.

Is this a case of disagreement? It definitely is the case that one speaker asserts S and another asserts not-S. In a third context, where Annie's attitude-states as a whole are marked as salient, i.e. where we focus on all of her attitudes, we might think that the two persons now cannot be said to disagree. That is because their rationality attributions were made in different contexts, which marked different sets of attitudes (of the same person in the same case) as salient.

We can find another example in Kolodny (2005: 509):

> " 'Fine. Since that's what you intend, you ought to intend to open a new carton of cigarettes; it would be irrational of you not to.' "

Kolodny suggests that 'S is rational' in this case. This seems plausible in this context, which marks as salient the attitude of intending to open a new carton by explicitly mentioning it and by drawing additional attention to it via the phrase "since that's what you intend". Now consider the following truncated version:

> "Fine. You intend to open a new carton of cigarettes. You ought to intend to open a new carton of cigarettes; it would be irrational of you not to."

If presented with this, a speaker might plausibly utter 'S is not rational'. In this context, the same attitudes are explicitly mentioned but without the "since that's what you intend"-phrase, this attitude is not marked as the only salient one. The speaker might therefore attribute additional attitudes to the agent, like a belief that smoking is harmful, which then motivates uttering that S is not rational. Can these two speakers be said to disagree in a third context? Again, one might think that this would be an accurate representation of the situation. Their utterances were made in different contexts, which marked different attitudes as salient. This leaves open the possibility of them agreeing in, say, a third context which marks as salient the entirety of an agent's attitudes.

The previous examples have shown that 'is rational' might pass the Disagreement Test. We can now clearly see that two speakers who utter these respective attributions do not disagree – the explicit reference to the respective subsets lets us see that they are in fact saying different things.

5.3 Resisting the Lure of Contextualism

Despite the apparent parallels between my account and contextualism, there are multiple reasons to be sceptical of contextualism about rationality.[2] I now turn to discussing two ways of pushing back on the contextualist analysis: reconsidering whether context-sensitivity tests actually provide evidence for the context-sensitivity of 'is rational', and the existence of an alternative, non-contextual explanation of the data that is compatible with my account.

5.3.1 Distinguishing Contexts

To start, it is unclear whether 'is rational' really passes the Disagreement Test (or other context-sensitivity tests). This depends on whether the way we ran the test in the previous section distinguishes contexts correctly. As it stands, I have applied the test as if a difference in information constituted a different context. For example, the difference between what I treated as the first context, where Annie expressed her intentions about attending the lecture, and what I treated as the second context, was an additional attitude that was mentioned (i.e. her belief that she might only believe she should go to the lecture because of her puritanical upbringing). But are these really two different contexts?

The question of how to individuate contexts is by no means a trivial matter and also one that is often overlooked in debates about contextualism. For the purposes of my discussion here, I limit myself to two different notions of context, as distinguished by Stalnaker (2014). One way of understanding 'context' is to treat it as a "centred possible world [. . .] [which] includes all that is the case [. . .] [, where this] include[s] any information that is relevant to determining what is said in an utterance" (ibid.: 24).[3] This would mean that a difference in context would amount to a different possible world, which includes all information. With this in mind, we can see that what I have treated as different contexts when I ran the Disagreement Test actually would not be different contexts, since the only difference between them are the pieces of information that are somehow marked as relevant. Putting the point differently, merely mentioning an additional attitude that the agent had all along, like a belief about how her upbringing impacts her intentions, does not amount to a different possible world. So if we understand contexts as possible worlds which already include all information, then we would actually be dealing with just one single context throughout. This means that 'is rational' does not pass the Disagreement Test after all – at least not in the way I presented it, which would take away some of the motivation for the context-sensitivity of 'is rational'.

[2] The most obvious, but least interesting reason to resist the draw of contextualism would be to find fault with the proposed context-sensitivity tests. For an example of such criticism, see Hawthorne (2006) (and DeRose's (2009) endorsement of Hawthorne's criticism). I do not engage with this option here, mainly because it does not help to shed light on how my account differs from contextualism, whilst being able to explain similar data about rationality attributions.

[3] Stalnaker calls this "the Kaplan/Lewis notion of context" (Stalnaker, 2014: 24).

Stalnaker does discuss another notion of context, which understands context as common ground, i.e. as what is presumed to be common knowledge by the speakers in that context. This marks a difference to the previous notion, in that we are not simply understanding context as a possible world including all information, but rather as all information that is deemed *relevant*. Understanding context this way puts the possibility of 'is rational' passing the Disagreement Test back on the table. Whether it does will depend on whether we think that the difference constituted by e.g. mentioning Annie's belief about her upbringing amounts to a different common ground. At present, I am not taking a stance on this. Even if 'is rational' passed the Disagreement Test on this understanding of context, the fact that it does not do so on another popular understanding of context is reason enough to treat with caution the (potential) support for the context-sensitivity of 'is rational' that can be derived from the Disagreement Test.

5.3.2 More Testing: Indirect Reports

Interestingly, this result can be re-encountered for other context-sensitivity tests, too. Cappelen & Lepore (2005) claim that evidence for the context-sensitivity of an expression can also be obtained if an indirect report of an utterance containing this expression in a (relevantly) different context comes out as false.[4]

As an example, consider the expression 'now', whose context-sensitivity is uncontroversial.[5] Now imagine that an agent, Emma, makes an utterance, which contains this expression, such as 'It is three o'clock now', where the context is such that it is indeed three o'clock at the time of utterance. To run the test, we now need to consider an indirect report of this utterance, such as 'Emma says that it is three o'clock now', uttered in a relevantly different context, e.g. one in which it is 5 o'clock at the time of utterance. This report of Emma's utterance is false. Emma never meant to express that is three o'clock when it is five o'clock. And so we have evidence that 'now' is context-sensitive.

Contrast this with an expression that is most likely not context-sensitive, like 'has three sides'. Imagine that Emma utters 'Triangles have three sides' in a maths class. An indirect report of this would be 'Emma says that triangles have three sides'. This is true, no matter how different the context of utterance of the report. For example, it would be true to say 'Emma says that triangles have three sides' in a literature class, on top of a mountain, or the next day. The fact that the indirect report is not blocked provides evidence that 'has three sides' is not context-sensitive.

Let us apply this to rationality attributions. Are indirect reports of sentences containing 'is rational' get blocked? Take again Brunero's (2012: 236) example of Candice:

[4] They summarise this test under the slogan "An Expression is Context Sensitive Only if it Typically Blocks Inter-Contextual Disquotational Indirect Reports", where 'disquotational' "just means you can adjust the semantic values of components of S that are generally recognized as context sensitive, i.e., we just test for the controversial components" (Cappelen and Lepore, 2005: 88, 89). I gloss over this for the sake of simplicity.

[5] Capellen and Lepore consider 'now' to belong to the "basic set of context sensitive expressions" (ibid.: 1-2), in reference to Kaplan (1989).

> "Candice decides to go to the post office this afternoon to send out some mailings, but on the way there, she gives up on this end and decides to go buy groceries instead. But on the way to the market, she yet again trades in this end for another: going to hang out with her friend David. But on the way to David's house, she once more changes her mind and intends to spend a relaxing afternoon at home, but by the time she gets home the afternoon is gone and she's accomplished nothing. We're inclined to accuse Candice of some kind of irrationality. And, since Candice failed to follow through on any of her ends, we're inclined to accuse her of being instrumentally irrational."

If someone is asked to evaluate the rationality of Candice and is presented with this information about her, the following rationality attribution seems plausible and is suggested by Brunero:

> 'Candice is not rational.'[6]

In order for this test to provide evidence for context-sensitivity, we would need an indirect report of the form

> '*A* says that Candice is not rational.'

to be blocked in a context of utterance that it is relevantly different from the previous one. The more fleshed out version of Candice's case that we have previously considered might suggest that this is in fact what happens. To recall, the more fleshed out story goes as follows:

> Candice decides to go to the post office this afternoon to send out some mailings, but on the way there, she remembers that her partner had promised her this morning to take care of the mailings. She then gives up on this end. Nevertheless, she decides to go buy groceries, because the market is right around the corner from the post office and she has a habit of combining these two trips. But on the way to the market, she realises that it is Monday and that the market is closed on Mondays. So she trades in this end for another: going to hang out with her friend David, who lives close by and with whom she wanted to catch up for a while. On the way to David's house, she decides to send him a message to let him know she is coming, and upon seeing their earlier conversation, remembers that David is out of town this week. She once more changes her mind and intends to spend a relaxing afternoon at home, but by the time she gets home the afternoon is gone, but she is not too unhappy about this, because she also believes that sunny days like today should not be spent inside.

[6] I'm assuming that 'irrational' entails 'not rational'.

Given this more complete story, it might seem false to report that '*A* says that Candice is not rational'. That is because in light of not only four intentions of hers (going to the post office, going to buy groceries, going to hang out with David, spending the afternoon at home), but also in light of some of her beliefs (that her partner is sending the mailings, that the market is closed, that David is out of town, that sunny days should be spent outside), her behaviour does not seem irrational. In fact, the initial utterer *A* might even think that Candice's reactive behaviour is displaying rationality, since she repeatedly saves herself from potentially irrational intentions. Alternatively, *A* might accuse Candice of other things, like forgetfulness, but that does not amount to an accusation of irrationality and does not justify reporting *A*'s utterance in this way.

Just like for the Disagreement Test, whether this is the correct way of running the test depends on our understanding of context. If we adopt the first understanding of context as a possible world including all information, then it makes no difference whether we are dealing with the shorter or longer version of Candice's case. And if we are not dealing with different contexts, we cannot run the test in the way I did. But if we adopt an understanding of context as common ground, then it is as least possible that the more fleshed out story of Candice presents a difference in common ground, and hence a different context, which would mean that we can indeed run the test in this way, and thereby potentially derive some support for the context-sensitivity of 'is rational'. However, as the following section will show, even if we think that some support can be provided by this test, it might eventually be undermined by yet another test, which I introduce in the next section: the Retraction Test.

5.3.3 Even More Testing: The Retraction Test

This brings me to a different, but related reason to resist contextualism about 'is rational', namely that 'is rational' arguably passes the Retraction Test. Unlike the previous tests, passing the Retraction Test can be seen as providing evidence *against* the contextualism of an expression.[7]

The basic idea is as follows: if a speaker is willing to retract a statement they made in a different context, the expression in question is probably not contextual, since there would be no need to retract otherwise. Take MacFarlane's (2014) example. A speaker utters 'Fish sticks are tasty' as a child and is reminded of that utterance as an adult. Since they no longer find fish sticks tasty, it seems natural to expect them to retract this utterance, rather than saying something along the lines of 'Yes and I stand by that. But fish sticks aren't tasty'. If 'is tasty' were context-sensitive, its truth conditions in the childhood context and in the adulthood context would be different, and so the statements would be compatible and the initial one would not need to be retracted. The adult version of our speaker could stand by the truth of 'Fish sticks are tasty' as uttered as a

[7] The test is usually used within discussion of relativism, which is a view in the contextualist family, but separate from Epistemic Contextualism. MacFarlane (2014) uses the test to motivate one version of relativism (assessment-sensitive relativism) over another (use-sensitive relativism).

child, and also by the truth of 'Fish sticks aren't tasty' as uttered now. The fact that this seems like an odd reaction, and that retraction seems much more natural, can be seen as evidence against the context-sensitivity of 'is tasty'. Putting the point differently, if we think that not retracting seems unnatural, then this might indicate that the truth conditions of the proposition are not determined by context after all.

Applying this to 'is rational', we get a similar result. Even if we grant the contextualist that the difference in information constituted by the long and short version of Candice's amounts to different contexts, 'is rational' arguably passes the Retraction Test. If the speaker who stated that Candice is not rational in response to the short version is presented with the longer version (and hence finds herself in a different context), they might say something like 'Now that all these additional facts about Candice have been brought to my attention, I take back what I said – she actually is rational'. But if 'is rational' were context-sensitive, our speaker should not feel any pressure to retract their previous statement, since its truth-conditions would be different, as determined by a different context.

These considerations obviously gloss over a number of issues that arise when we carry over a test from one debate to another. But even if these observations do not amount to a knockdown argument against the context-sensitivity of 'is rational', they should at least caution against jumping to contextualist conclusions, based on what might be merely superficial parallels.

5.3.4 Alternative Explanations

The case against a contextualist understanding of 'is rational' can be bolstered by appealing to alternative explanations of these perceived parallels. Let me reiterate the contextualist thought: in cases like Sara's, it seems like the statements 'Sara is rational' and 'Sara isn't rational' in dropping her means-end belief are both warranted, despite the fact that her situation remains unchanged. One possible explanation of this compatibility is the appeal to contextualism. The thought would be the that the statements are made in different contexts, which require different rational standards to be met in order to license a rationality attribution.

However, in debates of Epistemic Contextualism, it has been observed that the compatibility of such attributions (be it about rationality or knowledge) can be explained non-contextually. Let us start with Turri's (2017: 153) observation that "philosophers should remember that judgments about cases are influenced not only by explicitly stated details, but also by the inferences that people draw from them." In a similar vein, Dinges (2016) contends that our intuitions about knowledge attributions are influenced by the way the cases are presented to us. Almost always, these cases are underspecified, and how we fill them in makes a difference to the knowledge attributions we deem warranted. These points are entirely compatible with my account. In fact, I have made very similar observations in my discussion of the Strictly Local View and symmetry, where I claimed that readers fill in the attitude profiles of the agents in the cases they are being presented with attitudes they think the agent is likely to have (see Section 3.2.1).

Once we acknowledge the incompleteness of hypothetical cases, and our need to fill them in, it only makes sense to have a closer look at *how* we fill them in. Here, Gerken (2013: 48) reminds us that

> "our intuitive epistemic judgments, including those about knowledge ascriptions, are similar to other intuitive judgments [...] they are normally generated automatically, pre-reflectively, and effortlessly. [...] Intuitive judgments are, in general, governed by cognitive heuristics, many of which are fairly reliable in normal circumstances [...] [but also] systematically fallible, and again, there is little reason to suspect that intuitive epistemic judgments differ in this regard."

Because epistemic judgments rely on heuristics, they never take into account the entire complexity of a case, but focus on specific features. Which features we focus on then obviously impacts the outcome, i.e. the judgment (or knowledge ascription). Gerken calls this "epistemic focal bias" (ibid.: 49). Because we tend to aim at keeping cognitive effort as low as possible, we only process the smallest amount of information necessary for making a judgment, which intensifies the importance of the explicitly stated features in a case. This can mean that we sometimes disregard alternatives that are epistemically relevant, which can result in a (false) knowledge ascription. At other times, we might mistake epistemically irrelevant alternatives for relevant ones, which can result in missing out on a (true) knowledge ascription.

In much the same spirit, Dinges (2016: 223) argues that when we fill in cases that are supposed to support contextualism, such as the bank cases, we do so "in accordance with what [we] think is normal". He calls this the "Normalcy Projection". This way, what is or not explicitly mentioned in a given case directly impacts our judgment. As Dinges (ibid.: 223) argues, "if the author of a scenario fails to specify a relevant bit of the scenario, then, presumably, this part of the scenario is not supposed to be special in any way", and is hence disregarded. This is also a familiar thought, since I have argued that the question under discussion, which corresponds to a set of attitudes, which in turn determines the applicable rational requirements, can be the result of the attitudes that are *explicitly mentioned*, or otherwise marked as salient (see Section 4.3.2.1). In his attack on Epistemic Contextualism, Dinges then goes on to argue that bank cases are guilty of not mentioning the possibility of error in their low stakes versions, which means that readers do not consider it in their judgments, which in turn explains the differing knowledge attributions.

The points made by Turri, Dinges and Gerken allow us to push back further on understanding 'is rational' as context-sensitive. We can explain how rationality attributions like 'Sara is rational' and 'Sara isn't rational' in dropping her means-end belief are compatible without having to endorse contextualism. As developed in the previous chapter, my account explains this by appealing to and explicitly stating the different sets of attitudes that the attributions are attached to, which correspond to different questions under discussion. The way in which a case

is presented plays a crucial role here, as it impacts the cognitive process a reader undergoes, which leads to different attitudes being considered.

5.4 Conclusion

This discussion has drawn out some interesting insights. One, we have seen that the mere fact that my account holds that rationality attributions are relativised to sets of attitudes, which are at least in part determined by features that might be deemed contextual (such as differences in information), is not enough to push us all the way to endorse contextualism about 'is rational'. This has the additional advantage that my view need not inherit problems of Epistemic Contextualism.

Two, we now have an improved understanding of a main advantage of my account, namely to reconcile apparent disagreements, or apparently incompatible rationality attributions, by explicitly stating the set of attitudes in question.

And three, I take it that this discussion also provides a methodological insight: we should pay close attention to how we present the cases we rely on when arguing for a given position, since what is and is not explicitly stated can have an impact on the data we then go on to explain.

Chapter Six

Transmission Principles and the Normativity of Wide Scope Rational Requirements

At this point, we have achieved one of the two main goals of this thesis. We have reached a nuanced and thorough understanding of rationality attributions. I have developed an account that makes these attributions more flexible, informative, and accurate, without having to claim context-sensitivity for 'is rational'. We are now ready to move on to the second theme of the thesis – the Normativity of Rationality. Before I develop my argument in favour in Chapter 8, let us spend some more time on why the claim that rationality is normative has become a point of contention. As we will see, the findings of the first half of this thesis will continue to inform my discussion of this second main theme as well.

As a first step, I quickly recall the problem for the Normativity of Rationality. If this is understood along the lines of bridge principles like *(BPR)* that connect the requirements of rationality to normative notions, such as reasons, then we face a detachment problem. Because *(BPR)* is formulated as a material conditional, it allows for detachment via modus ponens, thereby giving us a reason for whatever is rationally required of us. This becomes problematic in cases where there is no reason to do what is rationally required, such as intending to kill your competitor for a promotion in *(The Ruthless Careerist)*, or believing a proposition on the mere basis that you believe it (in the case of bootstrapping). We have already seen in Section 2.1.3.1 how wide-scoping can provide part of a solution to this problem. When we accept wide-scope formulations of requirements, all that is required of us is to make a disjunction true, rather than adopting a specific attitude. And in turn, what we have a reason for, via *(BPR)*, is also a disjunction. For example, we would have a reason to either intend to kill our competitor, or to drop our belief that this is necessary for achieving our end, or to give up our intention of getting the promotion. In the belief case, we would have a reason to either believe p or to drop the initial belief. This

result is more acceptable. However, wide scope requirements are not entirely immune to the detachment problem. If we accept certain independently plausible principles, the detachment problem returns. This also means that Normativity of Rationality is still contestable. This chapter is concerned with exploring these independently plausible principles. These so-called 'transmission principles' make claims about how reasons are transmitted from ends to means. I will first focus on transmission principles in various forms and how they reintroduce the detachment problem even for wide-scope rational requirements. I then explore two ways of maintaining all elements of the problem – requirements, transmission principles and *(BPR)* – while still upholding the Normativity of Rationality as a possibility.[1] As we will see, one of these ways makes use of insights from Chapter 4. This paves the way for my positive argument in Chapter 8.

6.1 The Trouble from Transmission Principles

To see how detachment is still possible even for wide scope requirements, we have to take a closer look at transmission principles. Such principles come in very different forms and strengths. A weak version, which I call *(NT)*, stipulates that reasons are transmitted from ends to necessary means. A stronger version, *(LT)*, holds that reasons are transmitted from ends to sufficient means.[2] In what follows, I introduce examples of such principles in order to see how they function and how they reintroduce detachment.

6.1.1 Liberal Transmission Principles

Let us start with a very intuitive and straightforward formulation:

> *(LT1)* "If you have a reason to do A and doing B is a sufficient means to doing A, you have a reason to do B."[3]

As an example, consider *(Bus Stop)*. Say that I have a reason to meet my friend who is waiting for me at the bus stop. Maybe I have made a promise to meet her there. Getting off the bus at said stop is a sufficient means for meeting my friend at the bus stop. By the above principle, I then have a reason to get off the bus at that particular stop. The reason that I had for my end of meeting my friend has been transmitted to the means, i.e. getting off the bus. In cases like this, the principle seems to get things right.

But once we take a more fine-grained look at means, we can see that a principle as broad as *(LT1)* is liable to the so-called "problem of subversion" (Bedke, 2009: 679). This problem can

[1] As previously mentioned, I focus my discussion here on the reasons-version of bridge principles. Since *(BPR)* is weaker than *(BPO)*, problems for *(BPR)* are also problems for *(BPO)*.
[2] See Kiesewetter (2015) and Rippon (2011).
[3] (Way, 2010b: 224).

occur once we allow for reasons to be transmitted not only from ends to means, but also from means to means-to-means. If I have the end of meeting my friend at the bus stop, and getting off the bus is a means to that, surely pushing the stop button is a means to getting off the bus. One might think that reasons are transmitted "down the chain" (ibid.: 679): I have a reason to get off the bus because this is a means to my end and I also have a reason to push the stop button because this is a means to my means to my end. Whilst this seems plausible for *(Bus Stop)*, it can lead to problematic scenarios. Imagine I have two lottery tickets (Kiesewetter, 2017: 92). Arguably, I have a reason to increase my chances of winning. Buying another ticket is a means to increasing my chances of winning. *(LT1)* now holds that I also have a reason to buy another ticket. Moving further down the chain, we see that selling the tickets I already have is a means to buying another ticket. And if reasons were transmitted down the chain as seemed plausible in *(Bus Stop)*, it would now be the case that I have a reason to sell my lottery tickets. This clearly defeats my purpose of increasing my chances of winning. In order to block such cases of reasons-transmission down the chain, Bedke proposes a modification of the transmission principle:[4]

(LT2) "One has reason to take the means to what one has ultimate reason to do."[5]

By specifying that transmission has to start with an *ultimate* reason, where this is understood as a "non-means based reason", we block transmission down the chain (Bedke, 2009: 678).[6] The reason I have to sell my tickets is a means-based one: it is only a reason because selling the tickets is a means to something else I have a reason for, namely buying more tickets. But I do not have ultimate reason to buy more tickets; I only have a means-based reason to buy more tickets. So *(LT2)* does not give me a reason to sell my tickets. It only gives me a reason to buy more tickets, as this is a means to do what I have ultimate reason for: increasing my chances of winning.

The literature provides many more versions and restrictions of liberal transmission principles. For example, Raz's famous *facilitative principle* can be understood as a liberal transmission principle that is restricted to *undefeated* reasons: "When we have undefeated reason to take an action, we have reason to perform any one (but only one) of the possible (for us) alternative plans that facilitate its performance" (Raz, 2005: 5).[7] I do not discuss these more restrictive principles because, as Kieswetter (2017: 92-93) points out, adding too many restrictions goes

[4] Let me note that there are cases in which transmission down the chain seems unproblematic, like the bus stop example. These are cases where the means to the means is also a means to the end.

[5] (Bedke, 2009: 678).

[6] Kiesewetter (2017: 92) proposes a similar modification that is supposed to avoid the same problem. He uses the term "intrinsic reason" rather than "ultimate reason", see *(LT4)* below.

[7] Raz does not distinguish between "facilitating steps", "means", "preconditions" and "facilitating conditions" (Raz, 2005: 5).

against the very idea of *liberal* transmission.[8] He points to an argument put forward by Raz (2005: 8-9) himself, which locates the appeal of liberal transmission principles in the idea that taking means to ends we have amounts to "acting for a reason" which makes our actions intelligible, independently of "whether or not [agents] decide to follow it, and whether or not [agents] should conform to it". In this spirit, and going forward, I will rely on the following principle put forward by Way (2010b: 224), which is a combination of *(LT1)* and *(LT2)*:

> *(LT3)* "If you have an ultimate reason to do A and doing B is a sufficient means to doing A, you have a reason to do B."[9]

6.1.1.1 Liberal Transmission and Detachment

The previous section motivated liberal transmission principles, given the necessary restrictions. This section shows how accepting liberal transmission principles can pose problems for the normativity of rational requirements – even on the wide scope reading. The problem is that while wide scope requirements avoid detachment via modus ponens, they still allow for detachment when combined with *(LT3)*.

Take *(Means-End Coherence)*. This requirement, in combination with *(BPR)*, gives us a reason to [intend to p, if you intend q and you believe that p is a necessary means for q-ing]. In other words, it gives us a reason to [intend p or not believe that p is a necessary means for q-ing or not intend q]. Presumably, this reason is also ultimate in that it cannot be explained by being a means to something else that we have reason to do. Rather, this reason is directly provided by *(BPR)*. If this is the case, then we end up with the first part of the antecedent of *(LT3)*: A has an ultimate reason to [intend p or not believe that p is a necessary means for q-ing or not intend q].

Obtaining the second part of the antecedent is straight-forward.[10] Either of the above disjuncts makes the disjunction true. For example, intending to p is a sufficient means for A to [intend p or not believe that p is a necessary means for q-ing or not intend q]. Now that we have the entire antecedent, we can apply *(LT3)* and end up with the consequent: (i) A has a reason to intend to p. By iterating the same process, we can also conclude that (ii) A has a reason to not believe that p is a necessary means for q-ing and that (iii) A has a reason to not intend q. That is, *(LT3)* in combination with *(Means-End Coherence)* and *(BPR)* allows us to detach any of the above disjuncts and gives us reasons for each one.

We now face the same problem cases as the narrow scope-version of *(Means-End Coherence)*. Our careerist now has a reason to intend to kill her competitor. Of course, she also has a reason for all the other options, for example, the 'better' option of not believing that killing her competitor

[8] For what it is worth, as we will see, detachment can arise on more restrictive liberal transmission principles too. Since that is the crucial result here, not much hangs on this choice.

[9] (Way, 2010b: 224).

[10] In his 2005 paper, Raz goes through an argument of this form.

is a necessary means for getting the promotion in the case of immoral means-end beliefs. But this does not mitigate the problem: it is still the case that we can detach one option of complying with the requirement and obtain a reason for it – be it an unproblematic or a problematic option.

The same worry arises in the domain of theoretical rationality. Kolodny's entailment requirement in its wide scope version requires the agent to [not believe p, or believe what p entails]. In combination with *(BPR)*, we get a reason to [not believe p, or believe what p entails]. Obviously, either of the disjuncts is a sufficient means to make the disjunction true. If we also accept that this is an ultimate reason by the same reasoning as above, we can again apply *(LT3)*. It follows that (i) A has a reason to not believe p and (ii) A has a reason to believe what p entails. But now Kolodny's bootstrapping objection regains its force. Since p entails p, A has a reason to believe p. We can detach this option and also obtain a reason for it – regardless of the evidence or how plausible p is.

Adopting a liberal transmission principle like *(LT3)* reintroduces the detachment problem into the wide scope account of rational requirements. While the requirements themselves do not require a single attitude but rather a certain pattern of attitudes; *(LT3)* transmits the reason we have to instantiate this pattern to the individual attitudes that make up the disjunctive pattern. And this means that we are back where we started: rational requirements, this time in combination with *(BPR)* and *(LT3)*, over-generate reasons.

6.1.1.2 Against Liberal Transmission

A possible reply to the reoccurring detachment problem is to simply deny liberal transmission principles. This seems like an attractive option. After all, we might still not be ready to give up *(BPR)*. And we might also want to hold on to wide scope requirements. This seemed like a plausible reading of rational requirements and allowed us to avoid at least part of the detachment problem: it is no longer the case that the attitudes featuring in the required disjunctive pattern are individually required. The remaining part of the detachment problem is that we obtain a reason for each individual attitude – and this reason is transmitted via *(LT3)*. Therefore, this seems like a natural starting point.

Indeed, this strategy has been fairly popular. In many instances, counter-examples have been raised that cast doubt on the plausibility of liberal transmission principles, independently of the role they play in the debate about the normativity of wide scope rational requirements.

In all the counter-examples, the means is the 'flawed' element. On the more drastic counter-examples, we are confronted with immoral or harmful means; on the more moderate counter-examples, the means are in some way not ideal, for example inefficient. An example of the latter sort is provided by Bedke (2009: 683) who attributes it to John Broome:

(The Inefficient Snacker)

> "While at work you become hungry for a snack, and the snack machine is in the basement, one floor down from your office. Let us assume the only relevant ultimate reason is to satisfy your hunger, which you will do if you make it to the snack machine. One way to get a snack is to first walk two floors up before heading down to the basement to buy a snack. Under the instrumental principle, if this 'way' of getting a snack is a means, we have the odd conclusion that you have a reason to walk two floors up the building and then head to the basement for a snack. You don't have reason to do that, or so Broome thinks."

If we accept that one has ultimate reason to avoid being hungry, then *(LT3)* gives us a reason for everything that is a sufficient means for avoiding being hungry. And since going up two floors and then back down to the vending machine is sufficient for avoiding being hungry, the reason is transmitted to going up and down. But obviously, going up two floors only to go back down again seems like a pointless exercise – especially since exercising is not the goal here – satisfying hunger is. Moreover, a better option is available to the agent: she can just go down to the basement directly. Bedke (and Broome) think that principles like *(LT3)* are still too liberal: reasons should not be transmitted to just any sufficient means, as that would include very inefficient means like in the example above. Whilst it is true that an agent in the above example has a reason to avoid being hungry, it is not true that she has a reason to first walk up two floors and then go down to the basement to buy a snack.

More drastic counter-examples are provided by Broome (2005) and Rippon (2011):

> *(The Suicidal Snacker)*
>
> Plausibly, you have an ultimate reason to avoid being hungry in the afternoon. A sufficient means for avoiding being hungry later on is to kill yourself. And so by *(LT3)*, you have a reason to kill yourself.

Broome does not think that this is plausible. Whilst the means here would be efficient, it seems too drastic a result to have a reason to kill yourself – especially since there is a less drastic available option, namely eating your packed lunch. So again, *(LT3)* seems to over-generate reason statements.

> *(The Sexist Joker)*
>
> Plausibly, you have ultimate reason to tell a joke at dinner, as jokes generally contribute to everyone's merriment and good mood. One way of telling a joke at dinner is to tell a sexist joke. And so *(LT3)* gives you a reason to tell a sexist joke at dinner.

In this case, the sufficient means is an immoral one. Whereas this seems to matter with regard to what we think we do and do not have reason for, it does not matter with regard to whether a means is sufficient for an end. And since *(LT3)* only focusses on the latter aspect, a reason is transmitted from an unobjectionable end to an immoral, yet sufficient means. This seems false.

Independent counter-examples like the ones above have caused authors to reject principles like *(LT3)*. Being a sufficient means, albeit for an end that we have ultimate reason for, just does not seem a strict enough condition to warrant reasons-transmission in all cases.

6.1.2 Necessary Means Transmission Principles

One way of maintaining the intuitive idea behind *(LT3)* while avoiding the counter-examples is to restrict reasons transmission to necessary means *(NT)*. But even within this category of restricted transmission principles, there is a spectrum from more to less restrictive.

A fairly strict version is put forward by Setiya:

> *(NT1)* If you should do E, all things considered, and doing M is a necessary means to doing E, you should do M, all things considered, too.[11]

Note that *(NT1)* is phrased in terms of 'should' rather than 'reason'. But insofar as 'should' is the stronger normative notion, I take it to imply 'reason': whenever it is the case that you should A, it is also the case that you have a reason to do A.[12] Now, *(NT1)* transmits not only reasons but also all-things-considered should's from ends to necessary means.

To see why *(NT1)* is (at least somewhat) plausible, consider *(Bus Stop)*. Since your friend is waiting for you and you promised her to be there, you should arguably meet your friend at the bus stop, all things considered. Forming the intention to get off the bus is a necessary means to meeting your friend at the bus stop. By *(NT1)*, it follows that you should intend to get off the bus, all things considered. This seems to get things right. Furthermore, *(NT1)* also avoids the above counter-examples. Whilst we might accept that you should avoid being hungry, all things considered, walking up two floors before going down to the basement or killing yourself are not necessary means for avoiding being hungry. And so the all-things-considered 'should' is not transmitted to these means (and hence, no reason is transmitted either). Similarly, telling a sexist joke is not a necessary means for telling a joke and so no transmission takes place. In this case, one might even drop out of the principle earlier by denying that you should tell a joke at dinner, all things considered, since jokes always run the risk of upsetting someone.

This hints at a problem for *(NT1)*. The principle only applies to ends that one should have, all things considered. This strongly restricts the ends that can possibly transmit should's or

[11] (Setiya, 2007: 652).
[12] Horty (2012) and Nair (2016) reject this.

reasons. Presumably, we do not have that many ends E for which it is true that we should E, all things considered, because many of our ends are merely permissible. But the motivation behind transmission principles is a more general one, as laid out in Section 6.1. So *(NT1)* seems to be a bit too restrictive. A weaker version would be formulated in terms of reasons:

> *(NT2)* If you have a reason to do A and doing B is a necessary means to doing A, you have a reason to do B which is at least as strong as your reason to do A.[13]

Again, this sits well with *(Bus Stop)*. You have a reason to meet your friend at the bus stop, intending to get off the bus is a necessary means to meeting your friend at the bus stop, and so you have an equally strong or stronger reason to intend to get off the bus. *(NT2)* also allows for reasons-transmission for a greater variety of cases, as we have weakened the applicability condition from (all things considered) ought to having a reason. Nevertheless, we still avoid the problem cases noted above because none of those means meet the necessary means condition. The formulation that I will be using going forward is

> *(NT3)* If you have a reason to do A and doing B is a necessary means to doing A, you have a reason to do B.

This is basically Way's *(NT2)*, but I have dropped Way's specification regarding the weight of the transmitted reason, as this anticipates a possible reply to the detachment problem, which I will discuss later on.

So far, we have seen that *(NT3)* and its ilk do well in capturing unproblematic cases of reason transmission like *(Bus Stop)* while avoiding the above counter-examples featuring immoral or harmful means. If they are to be preferred to their liberal counterparts, we want them to also avoid the problematic consequences for the normativity of rational wide scope requirements that the latter engendered.

6.1.2.1 Necessary Means Transmission and Detachment

(NT3) and its ilk are not susceptible to detachment problems for rational requirements in the same way that their liberal counterparts are. To see this, again consider *(Means-End Coherence)*. In combination with *(BPR)*, we get a reason to [intend p, or not believe that p is a necessary means for q-ing, or not intend q]. This is the first part of *(NT3)*'s antecedent. So far, the argument runs parallel to the one for liberal transmission. But in the case of *(NT3)*, we do not get the second part of the antecedent straight away. Neither of the disjuncts is individually *necessary* for making the entire disjunction true. And so *(NT3)* does not transmit a reason from the disjunction to the individual disjuncts – we have avoided the detachment problem. The

[13] (Way, 2010b: 225).

situation is much the same for Kolodny's entailment requirement in the realm of theoretical rationality. *(BPR)* still gives us a reason to [not believe p, or believe what p entails], but neither of the two options is individually necessary for making the disjunction true and so no single one can be detached.

Unfortunately, the detachment problem persists in other ways, even with necessary means transmission. Setiya (2007) gives us a case that follows the by now familiar pattern. Call it *(Cigarettes)*. Say I intend to smoke and I believe that buying cigarettes is a necessary means to smoking. *(Means-End Coherence)* then requires me to [intend to buy cigarettes, if I intend to smoke and I believe that buying cigarettes is necessary for smoking] or, in the disjunctive version, to [not intend to smoke, or not believe that buying cigarettes is a necessary means for smoking, or intend to buy cigarettes]. *(BPR)* then also gives me a reason for it. But there is a catch. In Setiya's example, "there is nothing I can do to change my intention to smoke or my belief about the necessary means: these attitudes are not under my control" (ibid.: 660). In effect, this means that intending to buy cigarettes is the only way in which I can satisfy the requirement, making it a *necessary* means to satisfying the requirement. Now, we can get to the consequent of *(NT3)*: I have a reason to [not intend to smoke, or not believe that buying cigarettes is a necessary means for smoking, or intend to buy cigarettes] and [intending to buy cigarettes] is a necessary means for this. Therefore, I have a reason to [intend to buy cigarettes]. Detachment has returned, albeit in a slightly different guise. This time, we are not requiring the wrong things. But we still end up with reasons where we do not want them.

It should be noted that *(Cigarettes)* has been met with criticism. Bratman (2009) denies that reasons are transmitted here. He locates the appeal that rational requirements like *(Means-End Coherence)* have within the notion self-governance. Self-governance is crucial for cross-temporal agency: in order to have agency over a period of time, it is important that we are able to make plans and to follow through with them *ourselves*, rather than being at the mercy of fate. Put more simply, "we have an intrinsic reason to govern our own lives" (ibid.: 432). Practical rational requirements like *(Means-End Coherence)* provide a way of spelling out this notion of self-governance in terms of specific individual demands. So we have a reason for self-governance. This reason is then transmitted, via a principle like *(NT3)*, to the "necessary constitutive element of self-governance" (ibid.: 433). In Setiya's example, the only way in which the agent can comply with the requirement and hence display self-governance is to intend to buy cigarettes. But Bratman denies that a reason is transmitted to this intention. That is because the transmission presupposes that self-governance is possible. But in Setiya's example, it is not. Self-governance involves being able to change our attitudes upon reflection. This is not possible here, as the attitudes are stipulated to be out of the control of the agent. The cigarette case then is one where self-governance simply does not apply and hence no reason can be transmitted from it – without transmission, no detachment and no problem for *(BPR)*.

It is tempting to accept Bratman's criticism as a way of cashing out the impression that the

restrictions in Setiya's example seem ad hoc, and thereby deny that *(NT3)* gives rise to the detachment problem. But we can underpin Setiya's example by considering what I take to be its underlying view. For the relevant background, consider what Greenspan calls a "time-bound view" (Greenspan, 1975: 260) of ought-statements. She claims that this view does justice to the time-sensitivity of ought. What an agent ought to do changes over time and in particular with the options that are still available to the agent. She argues against modus ponens-like detachment rules that detach Oq from p and $O(p \rightarrow q)$. We can only ever detach an ought if it is the only live option remaining. There are two ways in which this could happen. One, all the other options are cancelled out by another ought. Two, there is an unalterable fact which makes it the case that only one option is still viable.

She uses the example of a parking ticket. I ought to [if I get a parking ticket, pay the fine]. This ought-statement gives me two options: I either not get a parking ticket, or I pay the fine. At the initial point in time, one of the options (that I ought to pay the fine) is cancelled out by another ought, namely that I ought to "avoid paying a fine, since my money is needed elsewhere" (ibid.: 265). So the only live option that I have is to not get a parking ticket and this is what I ought to do. At a later point in time, I have now gotten the parking ticket. This is an unalterable fact, there is nothing I can do to change this. Out of the initial two options – not getting a ticket and paying the fine – only one is a live option. Therefore, I ought to pay the fine.

The general lesson to take away from this is that what you ought to do depends on your live options. If there are two options to start with, but you cannot do one of them, then you ought to do the remaining one. The application of Greenspan's account of ought to reasons is straightforward. If I have a reason to not-p or q, and only one of those options is viable (be it because one is cancelled by another (stronger) reason or because one is an unalterable fact), I have a reason for the remaining option.

This matches the structure of *(Cigarettes)* perfectly. By stipulating that I cannot change my intention to smoke and my means-end belief, Setiya introduces two unalterable facts and thereby restricts the live options of the agent to one: intending to buy cigarettes. It then follows that we can detach a reason for this intention. I take it that the discussion of Greenspan's account shows that the view about transmission of normative statements (be it ought or reasons) that underlies Setiya's example is well-motivated. Criticism of his example then should be aimed at his seemingly ad hoc introduction of unalterable facts, thereby excluding the possibility of self-governance. But if this the case, Schroeder (2009: 227) shows how live options could be restricted to one without relying on out of the blue unalterable facts.

That is because it is a lot harder to change beliefs than it is to act on them. And even if we think that it is possible to change beliefs at will, it will take time. Schroeder's point is that, during this intervening time, changing beliefs is not a live option for the agent. In this sense, Setiya's stipulation that the agent cannot change her belief that buying cigarettes is necessary for smoking seems less ad hoc. Admittedly, this still leaves Setiya's agent with two options: intending

to buy cigarettes or no longer intending to smoke. But we can easily think of examples where the live options would indeed be restricted to a single one. Schroeder talks about a different rational requirement, namely *Enkrasia*: Rationality requires that [if you believe you ought to p, you intend to p].[14] The agent that is subject to this requirement only has two options to start out with: no longer believing that she ought to p, or intending to p. By *(BPR)*, she has a reason for that disjunction and by *(NT3)*, she has a reason for whatever is necessary to make that disjunction true. If we accept Schroeder's claims that changing beliefs takes time, thereby restricting the live options, and Greenspan's claims about detachment when there is only one live option, it is now the case that the agent has a reason to intend to p, since this is a necessary means for complying with *Enkrasia* – no matter how irrational or immoral p might be.

Moreover, this problem generalises because "with respect to all of our beliefs that we are not already trying to change, we are *always* in this intervening time" (ibid.: 227). This implies that a case where we are left with only one live option is not an exception but rather the norm. Bratman's self-governance ability now seems overly demanding.

To sum up, unlike *(LT3)*, *(NT3)* does not in and of itself transmit reasons from disjunctions to disjuncts. But given certain circumstances that are a lot more common than previously thought, it is still possible that a reason is transmitted to an individual disjunct, again making it the case that we are given a reason-statement that we are unwilling to accept. And just like before, this casts doubt on the Normativity of Rationality.

6.1.2.2 Against Necessary Means Transmission

Does this mean that we are finally being pushed to give up *(BPR)*? Not necessarily. Just like before, authors have put forward independent arguments to show that necessary means transmission principles are false. But unlike liberal transmission principles, it is not easy to come up with counter-examples to *(NT3)*.

To see this, consider two cases in Schroeder (2005), one of which features the now familiar case of Muriel:

> *Grass Blade Counters and Axe Murderers*
>
> "In cases in which someone's ends are deeply immoral or irrational – for example, that of counting the blades of grass on the White House lawn, or of becoming the world's most prolific axe-murderer – it does not seem to follow from the fact that the agent has such an end that there is a reason for her to take the means to it. It seems that there is no reason for aspiring grass-blade-counters to count blades of grass, and it

[14] This is a plausible requirement. Moreover, detachment is supposed to be a problem for the normativity of rational requirements in general. So if there is a problem with the normativity of one of them, there is a problem with the normativity of all of them. I therefore think it is warranted to appeal to a different requirement here.

> seems that there is no reason for aspiring axe-murderers to sharpen their blades and practice their swings."[15]

These cases seem convincing. The mere fact that an agent has irrational, immoral or otherwise objectionable ends does not seem to license the transmission of a reason to a means. But as it stands, *Grass Blade Counters and Axe Murderers* does not apply to *(NT3)*, since *(NT3)* postulates that the agent *already has a reason* for their end. But this is not stipulated in the above cases. If they are to serve as a counter-example against *(NT3)*, we would need to change them accordingly.

Let us assume then that someone's ends are irrational and immoral in the same way, but that they also have a reason for their ends: counting the blades of grass on the White House lawn would instil in them a great sense of achievement and becoming a prolific axe-murder would bring with it the attention of the public eye. Now, we can apply *(NT3)* and transmit this reason, however shaky it may be, to the necessary means. The agent then has a reason to sharpen their axe or to start counting.

One might think that this is indeed a counter-example to *(NT3)*. The fact that it allows for such transmission, regardless of the irrationality or immorality of one's ends, and provides one with some sort of normative encouragement in form of a reason for the means to one's irrational or immoral ends seems to speak against *(NT3)*. But still, this might be too quick. If we accept that the agent has a reason for their end, as obscure or weak as it may be, then we should probably also accept that they have a reason for the means, as obscure or weak as that one might be. So I do not think that we can reject necessary means transmission principles using the same strategy as we did with liberal transmission principles: the former seem to be immune against counter-examples.

This suggests that a rejection of *(NT3)* would need to be more principled. Kolodny (forthcoming) argues that *(NT3)* does not meet one of the desiderata for instrumental transmission, namely:[16]

> *Means Probabilize (MP):* If there is reason to E and there is a positive probability, conditional on one's M-ing, that one's M-ing, or some part of one's M-ing, helps to bring it about that one E's, then that is a reason to M, whose strength depends on the reason to E and on the probability.[17]

This means that the cases that cause trouble for necessary means transmission principles are the ones where the means, while being necessary, do not make obtaining the end more likely, or

[15] (Schroeder, 2005: 6).

[16] In his paper, this refers to "Strong Necessity (SN): If there is a reason for one to E, and M-ing is a necessary means to E-ing, then that is a reason, at least as weighty, for one to M" (Kolodny, forthcoming: 2).

[17] (ibid.: 4).

sufficiently more likely. Applying this to the modified *Grass Blade Counters and Axe Murderers* shows that these cases do not fall in this category. The means in question make achieving the end more likely – starting to count helps to bring about a final count of all the grass blades. Similarly, if there is a reason to become a prolific axe-murderer, it certainly holds that sharpening your axe will help to bring about the reputation of being a prolific axe-murderer. So we would need to look for other cases to put pressure on *(NT3)* on the basis of it not accounting for *(MP)*.

Presumably, these are cases where achieving the end is somehow impossible, even if we take the necessary means. In other words, these cases are a sub-class of the cases where the means are necessary but not sufficient for the end. Raz (2005: 7) gives us such an example – *(The Good Grandchild)*.

Plausibly, I have a reason to visit my grandmother on Easter Island, namely that it would make her happy. A necessary means for visiting her is to buy a plane ticket. But this means might fail to probabilise my end. If there is an air strike, buying a plane ticket will not help to bring about that I visit my grandmother.

What can we conclude from this? Does *(NT3)* really violate *Means Probabilize* in the above example? One might want to deny this because there is no positive probability, conditional on my buying the ticket, that buying the ticket will help to bring about that I visit my grandmother, given the air strike. One could also argue that we need to add some sort of proviso to *Means Probabilize,* stating that *M*-ing *normally* helps to bring about that one *E*'s, or make *Means Probabilize* apply to subjective rather than objective probability.

But even if we did all of that, there would still be another problem with *(NT3)* and *Means Probabilize*. This problem is related to the strength of the transmitted reason. *Means Probabilize* claims that the strength of the reason for the means depends on the strength of the reason for the end and the probability of bringing about the end, given the means. *(NT3)* does not specify anything about the strength of the transmitted reason – it is simply transmitted.

To illustrate this, consider Kolodny's (forthcoming) example. Two agents, Lucky and Unlucky, both have an antique sitting on their front porch and both have a reason to save it from the coming rain. By *(NT3)*, they both have a reason to take a taxi home as this is a necessary means for saving the antique. But Lucky stands a much better chance of actually saving the antique because the rain arrives slower in her area, Unlucky happens to live in an area where the rain arrives a lot faster. Nevertheless, the probability of saving the antique conditional on getting a taxi is still positive for both – it is just a lot higher for Lucky. *Means Probabilize* now claims that Lucky has a stronger reason to get a taxi than Unlucky. But *(NT3)* does not reflect that. It simply transmits *a reason* to both of them. So if we think that reflecting *Means Probabilize* is indeed a desideratum of instrumental transmission, then *(NT3)* falls short in this sense.

However, it might be too quick to reject *(NT3)* on this basis. Maybe this only shows the need

for more sophistication. We can imagine a necessary means transmission principle that includes a specification of the strength of the transmitted reason, along the lines of

> *(NT4)* If you have a reason to do A and doing B is a necessary means to doing A, you have a reason to do B whose strength depends on the strength of the reason to do A and on the likelihood that B-ing brings about A.

Now, *(NT4)* would yield the correct verdict in the case of Unlucky and Lucky. Both have a reason to take a taxi home, but Lucky has a stronger reason to take a taxi home because she is more likely to get there in time and save the antique.

6.1.2.3 Transmission Principles and Facilitating Means

So it seems like necessary means transmission principles have an answer to criticism stemming from *(MP)*. However, there might be a related criticism available. We have seen that *(NT)* principles can be rendered more plausible by integrating parts of *(MP)*. But one might have a more general worry: by fine-tuning and hence restricting *(NT)* principles more and more so that they avoid the problem cases, we run the risk of losing the essence of reasons transmission that the principles were supposed to capture in the first place. In particular, one might worry that *(NT)* principles generally (and their more restrictive versions especially) miss out on an entire category of reason transmission – namely cases of facilitating means.

Many cases of instrumental reason involve means that are neither necessary, nor sufficient, but simply means that make obtaining an end more likely, that facilitate it, that get us closer to our ends, so to speak. For example, agreeing to meet my friend at the gym is neither necessary nor sufficient for me actually going to the gym, but makes it more likely that I go to the gym, given the social pressure that this creates. *(NT)* principles only apply to cases in which the means that reasons are transmitted to are necessary for the end, and so would remain silent on cases like this one. But not being able to capture these means-end relations can be seen as a serious short-coming of any account of instrumental rationality. Schroeder (2009: 247) even goes so far as to claim that "it should be a constraint on an adequate account of [...] means-end coherence—that it [...] generalize to explain what is going on in cases where [the agent] has beliefs in non-necessary but facilitative means", such as the belief that agreeing to meet my friend there would facilitate me going to the gym. Necessary means transmission principles generally, and especially their more sophisticated versions like *(NT4)*, cannot account for facilitating means, and so we might reject them on that basis (thereby also eliminating the threat they pose for the Normativity of Rationality, as laid out in Section 6.1.2.1).[18]

[18] To be clear, the same criticism would apply to versions of *(LT)*, which are limited to *sufficient* means and hence also cannot account for merely facilitating means. But since *(LT)* is difficult to uphold, even without this additional objection, I do not discuss it here. Now, one might also wonder whether this means that we should turn to facilitating means transmission principles altogether. But turning to facilitating means, rather

These observations motivate a change in strategy – instead of tweaking individual transmission principles in order to avoid the problem cases, maybe we have to approach the issues from an entirely different angle. I explore this option in the next section.

6.2 Understanding Transmission

The counter-examples and arguments given in the previous sections certainly provide a challenge for transmission principles. But they do not amount to a knock-down argument against them. While liberal transmission principles face serious counter-examples, necessary means transmission principles have a whole collection of tools they can use to produce more refined variants. And there is also the option of facilitating means transmission principles.

This makes it difficult to cast a general verdict on transmission principles. Are liberal, necessary and facilitating means transmission correct or only one of them? And if at least one is correct, do we inevitably reencounter the detachment problem? Are we then forced to give up *(BPR)*, if we accept (some) transmission principles?

In this section, I want to take a step back from the individual principles and have a closer look at the general claim behind transmission principles. What exactly are we claiming when we say that reasons are transmitted from ends to necessary or sufficient means, and is this an accurate description of what is really going on? In turn, this will cast light on how pressing the detachment problem really is, as caused by transmission principles.

than necessary or sufficient means, leaves us in largely the same position with regard to the overall plausibility of reason transmission principles in general. For example, consider the following proposal (see Schroeder, 2009: 246):

> *general reason transmit (GRT)*
> If X has a reason to do A and X's doing B would facilitate her doing A, then X has a reason to do B of weight at least proportional to X's reason to do A, and to how well her doing B would facilitate her doing A.

When applying *(GRT)* to our example, this would mean that X's reason to make arrangements to meet her friend at the gym would depend on how strong her reason is to go to the gym, and how likely it is that making the arrangement will actually result in her going to the gym. And when applying it to our problem cases, *(GRT)* unsurprisingly yields the same verdicts. Grass-blade counters and axe-murderers end up with reasons to count blades and sharpen their axes, if we assume that they have a reason to pursue those ends in the first place, where the strength of the transmitted reason depends on the strength of the initial reason and the likelihood of the means achieving the end. In the case of *(The Good Grandchild)*, the grandchild's reason to buy a plane ticket is certainly weakened by the fact that this does not facilitate their end of visiting their grandmother, given the air strike, if it is transmitted at all (which would depend on how exactly the two provisos in *(GRT)* about the strength of the transmitted reason depending on i) the strength of the reason for the end, and on ii) how facilitating the means is, interact). Whether we think that principles along the lines of *(GRT)* are plausible will likely track whether we think that sophisticated versions of necessary means transmission like *(NT4)* are plausible, which in turn depends on whether we deem it an acceptable result that grass blade counters and axe murderers still have reasons for their means, even if we assume that they have reasons for their ends in the first place. Given the immoral or impractical nature of their ends, one might be hesitant to accept this. And so the move to facilitating means yields no real improvement.

6.2.1 Taking Transmission Seriously

Transmission principles gain much of their initial appeal from their name. Calling these principles 'transmission' principles suggests a picture on which a reason is simply passed on from an end to a means. Nothing new is added; reasons do not miraculously appear; the reason for the end just jumps over to the means: if you have a reason r_1 for an end, this reason is transmitted to the necessary or sufficient means. And so you have a reason r_1 for the means. The simplicity and elegance of this picture makes it compelling.

However, once we look closer, we can see that 'transmission' is not as plausible as previously thought. To see this, consider some of the previous examples and cases. If transmission principles really transmit reasons from ends to means on the literal meaning of these terms, then I should have the same reason r_1 for the means and for the end. But the counter-examples we encountered to liberal transmission principles raise doubts about the plausibility of transmission principles taken literally. These counter-examples often start out with an end that seems reasonable to have and then invoke a sufficient means that is objectionable in some way, e.g. due to moral concerns or inefficiency.

> *(Bus Stop)* In this example, we stipulated that the reason I have to meet my friend at the bus stop is that I promised to meet her there. It seems plausible to say that the fact that I have promised my friend to meet her at the bus stop is also a reason for me to get off the bus. So the transmission picture seems plausible here. But this should not be too surprising, given that this example was introduced to motivate transmission principles in the first place. Things look different if we turn to the counter-examples.

> *(The Inefficient Snacker)* The reason I have to get a snack is that I want to satisfy my hunger.[19] Would it be correct to claim that wanting to satisfy my hunger is a reason for me to first go up two levels and then down to the basement? This seems a lot less plausible already. It is not obvious that wanting to satisfy my hunger favours taking the detour, if considered in isolation of the end of getting a snack. But if we are to take transmission seriously, this favouring-relation (as a very general characterisation of the reason-relation) should hold in isolation.

The situation is similar for necessary transmission principles. Here, the counter-examples also

[19] To be precise, Bedke's initial example does not explicitly state the reason one has to buy a snack. Wanting to satisfy your hunger is presented as an end, for which buying a snack is a means. Taking a detour to the vending machine is presented as a means to the means of buying a snack. In order to maintain the same structure and to avoid 'transmission down the chain'-like scenarios, I choose to present the case as a classic two-step means-end scenario, where we have a reason (wanting to avoid being hungry) to an end (buying a snack), for which taking a detour to the vending machine is a means.

exploit the lacuna between the end and its reason on the one side and the means on the other side.

> *(Cigarettes)* The case for transmission gets worse if we consider Setiya's case. The reason I have to [not intend to smoke, or not believe that buying cigarettes is a necessary means for smoking, or intend to buy cigarettes] is provided by *(BPR)* – it is the fact that I am rationally required to be means-end coherent. If the proposed transmission mechanism is to be taken at face value, then it would need to be the case that the fact that I am rationally required to be means-end coherent is a reason for me to intend to buy cigarettes. Again, the required favouring-relation does not hold between the reason and the means, in isolation of the end. The fact that I am rationally required to be means-end coherent is not itself related to an intention to buy cigarettes, let alone via a favouring relation.

> *(The Sexist Joker)* If the literal reading of transmission principles is correct, I should have the same reason I have to tell a joke over dinner – that jokes generally contribute to everyone's merriment and good mood – as I have to tell a sexist joke. But the claim that jokes generally contribute to everyone's merriment and good mood is a reason for me to tell a sexist joke seems outright false. After all, telling a sexist joke might contribute to everyone's discomfort instead.

> *(The Good Grandchild)* Assume that the fact that visiting my grandmother on Easter Island would make her happy is a reason for me to visit her. And buying a plane ticket is a necessary means for visiting her. By transmission, it would then need to be true that the fact that visiting my grandmother on Easter Island would make her happy is a reason for me to buy a plane ticket. But this seems false, given the additional feature of Raz's example that there is an air strike and buying a plane ticket gets me nowhere near Easter Island.

The lesson to be drawn from this is that when we spell out the very idea of 'transmission' in this literal and precise way, it becomes clear that this picture is not the correct one for all cases. Once we fill in the reason r_1 that is supposedly transmitted, the transmission picture is a lot less plausible. Are these observations sufficient to reject transmission principles altogether?

That might be too quick. One might worry that what I have presented so far is a straw man argument. If we refer back to the proposed principles, we can see that, despite their names, none of them are actually formulated in terms of 'transmission'. Instead, they claim that reasons follow certain patterns. Generally speaking, these principles claim that if you have *a* reason for

an end, you have *a* reason for the (necessary or sufficient) means.[20] These two reasons need not be the same. Instead of reasons being transmitted from ends to means, these principles actually claim that reasons follow a pattern: whenever you have a reason r_1 for an end, it will be accompanied by a reason r_2 to the means, where r_1 and r_2 need not be identical. Instead of speaking of transmission principles, it might therefore be more helpful and charitable to think of these principles as reason-detection principles: whenever you have a reason for an end, there will be a reason for you for the means in the vicinity.

Support for understanding transmission principles this way comes from a popular view, or family of views, that understands reasons as tied to *promotion*. These views hold that the existence of a reason is tied to whether that reason would promote some kind of objective, which is captured by the following template:

> *(Reason-Promote)*
>
> There is a reason for an agent, A, to ϕ iff A's ϕ -ing promotes a K of the relevant kind.[21]

There are many different ways of filling in this template. For example, the kinds of things that are 'of the relevant kind' to be promoted are understood as objective values (Wedgwood, 2009; Parfit, 2011), or desires of the agent (Schroeder, 2007). The same is the case for the question of how to understand promotion. For my purposes, it will suffice to understand promotion along the lines of probabilising or facilitating the objective, as featured in *(MP)*. If we understand reasons in that way, then the literal understanding of transmission principles looks a lot less appealing from the outset. That is because the cases that cause trouble for these principles are precisely cases where promotion breaks down – the transmitted reason no longer promotes the objective in question. In what follows, I will go through some of these cases to see how promotion-based views can provide a more unified diagnosis of the problems that transmission principles face.

Before I turn to some of the problem cases, let me briefly consider *(Bus Stop)* for illustrative purposes, as it is a case that actually is not problematic. Since I have promised my friend to meet her at the bus stop, I have a reason to meet her at the bus stop, because this would promote something of the relevant kind, namely keeping my promise. By the same token, I also have a reason to get off the bus, since this would still promote the objective of me keeping my promise. This result is in line with transmission principles, which would transmit a reason from the end of meeting my fried at the bus stop to the (necessary and/or sufficient means) of getting off the bus. Appealing to promotion-based views allows us to provide a more precise explanation of why

[20] For example, take *(LT1)*: If you have a reason to do A and doing B is a sufficient means to doing A, you have a reason to do B; and *(NT2)*: If you have a reason to do A and doing B is a necessary means to doing A, you have a reason to do B which is at least as strong as your reason to do A.

[21] Sharadin (2015: 102), cf. Snedegar (2014: 46).

(Bus Stop) is not a problem case for literal reason transmission – because the transmitted reason upholds promotion.

Things look different when we move on to some problematic cases. Take *(The Inefficient Snacker)*. If literal reasons transmission was true, it would need to be the case that the fact that I want to avoid being hungry is a reason for me to go up two levels and then down to the vending machine in the basement. We have already noted that this seems like an odd claim, and now we can explain why: the promotion-relation between the objective K (avoiding hunger) and the means (taking the detour) does not hold, or is weak at best.[22] Appealing to promotion thereby lets us better diagnose why *(The Inefficient Snacker)* is a problem case: it is a case in which promotion between the means and K breaks down.

We can tell a very similar story for *(The Sexist Joker)*. It seemed false to claim that the fact that jokes generally contribute to everyone's merriment and good mood is a reason for me to tell a sexist joke over dinner, as would be the result of literal transmission principles. And again, by appealing to promotion, we can explain why: telling a sexist joke would not promote K – contributing to everyone's merriment and good mood. In fact, it would do the opposite. The situation is very much the same for *(The Good Grandchild)*. It is not the case that the fact that visiting my grandmother on Easter Island would make her happy is a reason for me to buy a plane ticket, because this would not promote making my grandmother happy (K), given the air strike.

Revisiting these cases in light of promotion-based views of reasons teaches us further lessons about why literal transmission can fail, namely because the promotion-relation between the means that a reason is supposedly transmitted to, and the objective K that is to be promoted, breaks down. In turn, this gives us an indication as to how transmission principles would need to be modified in order to avoid the counter-examples. What could go some way towards accommodating these additional requirements is to take on the previous suggestion of understanding them more as reason-detection principles that describe certain patterns, rather than literal transmission down the line. Maybe the more plausible transmission principles would specify patterns of reasons that also feature the presence of the relevant promotion-relation.

At this point, let me note that these more general conclusions do not require accepting a promotion-based view of reasons. There might well be other available diagnoses and solutions to the problems faced by literal transmission. But what this discussion shows is that there exists a respected conception of reasons on which literal transmission does not seem very plausible from the outset, which should at least caution us against accepting transmission principles too quickly, and highlight the importance of resisting their temptation, given their elegance and simplicity.

[22] Whether we are willing to speak of promotion in this case will depend on how exactly we spell out promotion. For example, if we are happy to count even marginal increases in likelihood as promotion, then this would presumably be a case where promotion still holds.

Let us take a step back and consider what these insights mean for the general dialectic of this chapter. On the positive side, the suggested principles need not necessarily be abandoned just because we have seen that the idea of literal transmission is misguided. Maybe all we need to do is understand them as more complex principles that specify more complex reason patterns. This might allow the principles to either avoid or accommodate the raised counter-examples. On the negative side, this also means that the principles lose at least some of their initial plausibility. Their simplicity, elegance and non-mysteriousness played a big role in motivating them. But now, arguing for such principles becomes a much harder task, because we need to explain how it is that all these elements (reasons for ends, potentially different reasons for means, and promotion relations) fall into place in the right way when we are faced with means-end relations. We now need an argument for why reasons would display these patterns. And presumably, the range of cases that would fall under these more complex pattern principles would be much more limited.

In the greater context of this chapter, this is a welcome result. Transmission principles caused trouble for the truth of *(BPR)* by reintroducing detachment. The threat to *(BPR)* is significantly weakened if transmission principles lose plausibility, or prove to be restricted in their application. This leaves open the possibility that we can avoid the problematic detachment cases, if these can be identified as instances where the principles are either violated or not applicable.

6.3 Understanding the Normativity Problem

Leaving the remarks of the previous section aside, is there another way to defend the Normativity of Rationality? Let us take a step back and remind ourselves why transmission principles posed a problem for the Normativity of Rationality in the first place. The combination of three plausible principles led to an undesirable result. The three principles in question are *(BPR)*, which links rational requirements to reasons, a wide-scope means-end coherence requirement and a transmission principle, linking reasons for ends to reasons for means. The undesirable result was produced in the following way: we start with the rational requirement, which requires us to either intend p, to not intend q, or to not believe that p is a necessary means for q-ing. Adding *(BPR)* gives us a reason for this disjunction. Once we add a transmission principle, say *(LT3)*, we get a reason for each of the disjuncts, because each disjunct is a sufficient means for making the disjunction true. Since there are no restrictions on the contents of our attitudes, we now have a reason to, say, intend to kill our competitors or to tell sexist jokes. This is the undesirable result that we want to avoid: ending up with reasons for contentious attitudes through an uncontentious process.

If we want to avoid this result, there are four different available strategies. We can either reject one of the three principles, or find a way to endorse the result. The majority of this chapter was dedicated to exploring the strategy of rejecting transmission principles. The previous sections have shown that we might be justified in doing so, either because the transmission principles are prone to counter-examples or because the underlying idea of 'transmission' is flawed. Still, these

conclusions still leave open the possibility of transmission principles (or better versions thereof) sometimes being correct. This possibility might be enough to make us doubt whether this strategy is successful – after all, we would not have a guarantee that the undesirable result is *always* avoided. Given the overwhelming support for wide scope rational requirements like *(Means-End Coherence)* and the crucial role that some version of *(BPR)* plays for our practices of criticising people for violating rational requirements, these two strategies would incur significant costs. Could we avoid these costs by developing the remaining strategy – endorsing the undesirable result?

6.3.1 The Weak Reasons Strategy

The Weak Reasons Strategy is one way of following this strategy. The general idea is that once we realise that the reasons for contentious attitudes that we end up with might just be very weak reasons, the undesirable result seems less undesirable.

Schroeder (2005) proposes one such strategy. He suggests that we often interpret claims like "X is a reason for you to p" as "X is a strong, weighty reason for you to p".[23] This implicit assumption that the reason in such a statement is a strong one explains why we might want to deny that you have a reason to intend to kill your competitor (as liberal transmission principles would have it) or to intend to buy cigarettes (as necessary transmission principles would have it in Setiya's case). It does seem false to claim that the agents in these examples have a strong reason to intend to kill their competitors or to buy cigarettes. But this is not what the transmission principles say. They merely state that the agent has *a* reason for the means – without making claims about the strength of that reason. It could be a very weak one. Now, it sounds a lot more plausible to hold that the agents indeed have a reason (namely a weak one) to intend to kill their competitors or to buy cigarettes.[24] Schroeder suspects that critics who reject transmission principles actually misread them as stronger principles, which claim weighty reasons for the means. Once we make explicit that the reason in question could be of any strength, from weak to strong, the counter-examples lose some of their force.

Bedke's (2009) argument for *(LT3)* proceeds in a similar vein. In response to the counter-example of *(The Inefficient Snacker)*, he holds that we do in fact have *a reason* to take a detour before going to the basement. To draw this out, he appeals to a possible alternative: not moving at all. When we compare 'taking a detour' with 'not moving', it is obvious that we have more reason

[23] Schroeder (2005: 7-8) provides two independent reasons to support this claim. One, because reasons play an important part in deliberation, whenever we talk about a reason, we assume that it must at least be weighty enough to be considered in deliberation. Two, because presumably there are so many weak reasons, whenever we claim that *there is* a reason without naming it, the assumption is that it is weighty or important enough to deserve the mention – it would almost be platitudinous to say 'there is a reason for you to p' if that reason referred to one of the many weak ones.

[24] Note that even on this more charitable interpretation, we can still question the plausibility of transmission principles on their literal reading. It is possible to accept the Weak Reasons Strategy while rejecting literal transmission.

to take a detour than we have to not move. And, so Bedke argues, if you have more reason for X than you have for Y, you must have *some* reason for X, with X being 'taking a detour'. Of course, this reason can easily be outweighed by other available alternatives like 'going directly to the basement'. The important thing is that we still have *a reason*, albeit a weak one, to take a detour, just like *(LT3)* predicts. Bedke thinks that our initial puzzlement about this result can be explained by appealing to the "pragmatic oddity and practical irrelevancy" involved: "[...] it would be pragmatically odd to say that one has a reason to get to the snack machine by first taking a circuit on the stairs. [Since] [t]hat is a poor way to get to the snack machine as compared to obvious alternatives" (Bedke, 2009: 684). *(The Suicidal Snacker)* might seem like a trickier case, but Bedke holds that it can be explained in the same way. We can hold that you do have a reason to kill yourself, even though it will be massively outweighed by the reasons you have for the other available alternatives – like eating your lunch.

To further motivate his defence, Bedke also puts forward the claim that adding a better alternative does not amount to providing an undercutting defeater. Undercutting defeaters are considerations that 'undo' a reason. An example would be me checking my bank account and seeing that my balance is 0. This seems to give me a reason for believing that there is no money in bank account. But that reason can be undercut by me receiving an email from my bank, informing me that due to technical maintenance works, all accounts will display a balance of 0 for 24 hours. I no longer have a reason to believe that there is no money in my bank account.

In *(The Suicidal Snacker)*, we seem to treat an attractive alternative (eating lunch) like such an undercutting defeater for the reason we have for another alternative (killing yourself). Bedke holds that it is more plausible to treat the attractive alternative as a "competing reason, one that massively outweighs, or perhaps trumps, the reason it competes with" instead of "eliminating the normative force of the reason it competes with" (ibid.: 685). With that clarification in mind, we should be less inclined to deny that you have a reason to kill yourself, just because you could eat lunch instead in order to avoid being hungry. Of course, this still assumes that you have a reason to kill yourself to start with, which we have established by comparing the course of action to worse alternatives, like we did for *(The Inefficient Snacker)*.

I summarise both Bedke's and Schroeder's replies under the 'Weak Reasons Strategy' because their defences of liberal and necessary means transmission respectively rely on claiming that the transmitted reason need not be a strong one and indeed is more likely to be a fairly weak one. This makes it much easier to endorse the undesirable result – we still end up with reasons for contentious attitudes through an uncontentious process, but these reasons are merely weak ones.

6.3.2 The 'Weak Rationality' Strategy

But one might still be inclined to worry about the undesirable result. After all, we still have reasons where we think we should not, even if they are only weak ones. To address this worry, I think we can supplement the Weak Reasons Strategy by offering a somewhat similar diagnosis

for a second aspect of the undesirability of the undesirable result. Once this diagnosis is on the table, the undesirable result might be even easier to endorse.

Why is the mere fact that agents obtain weak reasons for contentious attitudes via the previously described mechanism seen as such a big problem for the Normativity of Rationality? I conjecture that part of this answer is that it seems like an agent who acts on the resulting (and potentially weak) reason is subject to a two-fold positive appraisal. Not only do they act in accordance with one of their reasons, they also satisfy a rational requirement. This connects back to how we rely on rationality in our practices of criticising or praising agents as described in Chapter 1.

Behaviour of this kind normally seems praiseworthy to us. An agent who intends to read a logic textbook, because (i) they intend to become better at reasoning and believe that reading a logic textbook is necessary for that and (ii) because the fact that this book is required reading for their logic class is a reason for them to read the logic textbook, is definitely doing many things right. We are naturally inclined to call such a person rational.[25] After all, their attitudes are in perfect order and so they can be seen as rational from a coherence-point of view. Moreover, they can be seen as rational in an additional sense: they are also reasons-responsive.

Since situations like the sexist joker or the ruthless careerist share the same structure with the situation of the exemplary logic student, we might be pushed to think that the same positive appraisal and rationality attribution is warranted. They are satisfying a rational requirement that applies to them, namely *(Means-End Coherence)*, and they are also acting on a reason they have, namely the one that is transmitted via the mechanism described in Section 6.1.1.1. So they seem to satisfy an additional requirement, namely *(Reasons-Responsiveness).* But this preponderance of positive appraisal seems out of place in the cases of sexist jokers or ruthless careerists, especially when compared to rational exemplars like the exemplary logic student. This mismatch might explain, at least in part, why the undesirable result is perceived as so undesirable.

We have already encountered one way to lessen the cost of accepting the undesirable result. The Weak Reasons Strategy pointed out that the reason involved might be a very weak one. So an agent who acts on a weak reason might actually only be minimally reasons-responsive. The positive appraisal that they should be awarded is limited, since they might fail to act on stronger reasons they have. And to be precise, they might not even satisfy *(Reasons-Responsiveness)* after all, since this requirement is phrased in terms of intending what one has *decisive* reason to intend. The Weak Reasons strategy casts doubt on exactly this – presumably, the potentially very weak reason that the mechanism provides to the agent is not a decisive one. And so on this understanding, we might not even be able to call them reasons-responsive at all.

This weakens the positive appraisal available to sexist jokers and ruthless careerists, which might

[25] This is in keeping with the account of rationality attributions I have developed in this thesis. I am leaving the more precise attribution aside for now, but will come back to it shortly.

already soften the blow of the undesirable result. But we can do more. The second form of positive appraisal – the rationality attribution – can also be modified accordingly, once we remind ourselves of the more nuanced picture of rationality attributions that I develop in Chapter 4. I suspected that the reasoning that led to this unrestricted positive appraisal went as follows: ruthless careerists satisfy a rational requirement that applies to them and since rationality consists in satisfying rational requirements, they are rational. An unqualified rationality attribution of this kind can easily be understood as making a general claim about the agent's overall rationality, which is then met with resistance.

But this would amount to an impoverished and imprecise understanding of the rationality attribution. I have argued that making explicit the set of attitudes that underlies the attribution can alleviate this (see Section 4.1). Ruthless careerists are a great application for the account that bring out its advantages. They are rational *with regard to a subset of their attitudes*, where this subset only includes their intention to get the promotion, their belief that this necessitates killing their competitor and their intention to kill their competitor. This subset also corresponds to a particular question under discussion. They satisfy the rational requirement that applies to them at this subset. So far, so good. But notice that once we have made explicit the subset in question, the resulting rationality attribution is obviously restricted in its focus – it is far from making a general, all-attitudes-considered statement about the agent's rationality. This statement would only be warranted if the subset a_{n} that we considered included *all* of their attitudes and if they then satisfied all rational requirements that apply to the entirety of their attitudes. Bringing to the forefront the specific underlying subset should make it easier to accept the (restricted) rationality attribution as a form of positive appraisal for sexist jokers and ruthless careerists – they are rational, but only with regard to a very small subset of their attitudes. Importantly, this leaves open the possibility of criticising them. Once we take into account larger sets of their attitudes or all of their attitudes, they plausibly violate other rational requirements that apply to them, given all of their attitudes, which means that we would not attribute overall rationality to them.[26] Note also that this would be reflected in the robustness values of the rationality attributions in question. Since sexist jokers and ruthless careerists are likely to violate rational requirements at only slightly larger subsets, the rationality attributions available for them at the subset of evaluation would not be very robust.

This should restrict the second form of positive appraisal to be awarded to agents in the problematic cases even further. Once we realise that the positive appraisal given to sexist jokers and ruthless careerists does not need to match the positive appraisal given to the exemplary logic student, despite the structural similarity, it is easier to accept the undesirable result, without having to reject any of the plausible principles. We only need to be comfortable in claiming that a ruthless careerist is minimally reasons-responsive and not very robustly rational with regard to

[26] For example, the ruthless careerist might also have a belief that one ought not kill people for one's own benefit. If we add this belief to the subset of attitudes, an additional rational requirement applies to them, namely *(Enkrasia)*, which they would violate.

a small subset of their attitudes.[27] I take it that this is not such a bitter pill to swallow.

These two strategies then present a way of softening the blow of the undesirable result – we might be able to maintain all elements of the mechanism, namely *(BPR)*, transmission principles (on their appropriate reading) and the involved rational requirements – by endorsing the positive appraisal of sexist jokers and ruthless careerist on its improved, more nuanced understanding.

6.4 Conclusion

In this chapter, I surveyed various transmission principles and settled on two versions, *(NT3)* and *(LT3)* as most plausible. I explored how these principles are standardly taken to create a problem for the Normativity of Rationality. In order to maintain this normativity, I looked at ways to reject the principles. I concluded that while many of the replies to this criticism have their merit, what we should really do is to understand such principles as detection- rather than transmission-principles for reasons. This created the need for additional argument for the plausibility of these principles, which in turn weakened the problem they presented for the Normativity of Rationality. In the final section, I approached the issue from a different angle and argued that even if we maintain transmission principles as they are, we do not necessarily run into problems for the Normativity of Rationality, if we adopt my preferred way of making rationality attributions.

In the overall narrative of this thesis, this means that we are now a step closer to achieving the second aim of this investigation: providing support for the Normativity of Rationality. The above arguments explored various ways in which a commonly held knock-down argument for the Normativity of Rationality might not be so problematic after all. I remain neutral with regard to which of the explored options should be preferred, and I take it that this neutrality, and the variety of options, support my case. What matters is that there *are* ways of accommodating the detachment problem caused by transmission principles, and not so much which exact way we settle on.

[27] And given the account developed in Chapter 4, we are already comfortable with making rationality attributions that seem incompatible but actually are not, since they concern different sets of attitudes with their corresponding QUDs.

Chapter Seven

Minimal Disturbance: In Defence of Pragmatic Reasons of the Right Kind

7.1 Introduction

Before turning to my positive argument in the final chapter, this chapter brings to the forefront the methodological current that has been underlying much of this thesis. I have mentioned at various points that adding more complexity, more fine-graininess and more explicitness can lead to progress with persisting problems in debates about rationality. This was most prominent in my discussion of rationality attributions and also of transmission principles. In this context, this chapter can be seen as a case study illustrating the approach. In particular, it draws attention to an important methodological shortcoming in debates about what counts as a reason for belief. An extremely influential distinction in this literature is between reasons of the 'right kind' and the 'wrong kind'. However, as I will demonstrate, arguments making use of this distinction often rely on a particular conception of epistemic rationality, that is *not explicitly stated*. Shifting focus to a reasonable alternative can lead to surprising consequences. The alternative I explore is a coherence conception of rationality, which should be unsurprising, since my discussions so far have made no secret of where my sympathies lie. Nevertheless, many of my points hold regardless of whether one shares this endorsement of coherence rationality. For the length of this chapter however, I will explicitly assume this conception to show that pragmatic reasons can, against orthodoxy, indeed be reasons of the right kind for belief.

This finding also interacts with the previous chapter in two ways. One, we can treat the problem created by transmission principles as evidence for the plausibility of treating pragmatic reasons as reasons of the right kind for belief. This would explain why the detached reasons, e.g. for believing a bootstrapped belief, are deemed problematic – and in fact, so problematic that

we are ready to call into question the truth of transmission principles or the Normativity of Rationality. These detached reasons are most plausibly understood as pragmatic reasons, at least on a standard understanding that takes pragmatic reasons for belief to be non-evidential, and non-pragmatic reasons for belief to be evidential (see Section 7.2.1). In any case, detached reasons are certainly not evidential – whether or not the favoured belief is supported by any kind of evidence did not play a role in the generation of the reason. If pragmatic reasons were only reasons of the wrong kind for belief, the result explored in the previous chapter should not be worrisome. Admittedly, the combination of transmission principles, rational coherence requirements and the Normativity of Rationality leads to odd reasons, but these are of the wrong kind, and so could easily be brushed aside. But the fact that the opposite is the case – most people tend to be very worried about this result – indicates that treating pragmatic reasons as reasons of the right kind for belief might not be such an absurd idea after all.

Two, the findings developed in this chapter can then be used to supplement the Weak Reasons Strategy in response to the problematic result discussed in the previous chapter. By allowing pragmatic reasons into the realm of reasons of the right kind for belief, we can increase the likelihood of the problematic detached reasons being outweighed. This means that there is another way available to soften the blow that the detachment problem presents for the Normativity of Rationality.

Here is the plan of the chapter: I first spell out the distinction between reasons of the right and wrong kind for belief and relate it to an evidentialist conception of epistemic rationality. This conception can be seen as representing what I called the reasons-responsiveness conception of rationality and seems to suggest itself quite naturally (Section 7.2). I then show that it is not obvious that this is the only plausible conception – the distinction leaves room for others. My diagnosis of why this conception has become the default is that people generally and implicitly assume that the rationality and the correctness of a belief converge. The coherentist conception of epistemic rationality, which I explore and motivate some more, challenges precisely this convergence and also provides us with an example of a pragmatic reason of the right kind (Section 7.3). In this section, I also briefly outline how this finding can supplement the existing Weak Reasons Strategy in its defence of the Normativity of Rationality. I finish by considering some objections (Section 7.4) and by noting the importance of specifying standards of correctness and rationality conceptions for debates about rationality in general (Section 7.5).

7.2 Reasons: The Right and Wrong Kind

My aim in this chapter is to highlight the importance of specifying the conception of epistemic rationality underlying philosophical discussions before drawing substantial conclusions. Whilst this might sound like nit-picky methodology, there is a real danger attached to not doing so. In particular, one runs the risk of begging the question against subscribers of competing conceptions, thereby weakening the strength of the overall argument. If whatever conclusions are drawn

non-trivially depends on which conception of, in this case, epistemic rationality, is adopted, these conclusions can hardly be seen as universal. Critics can easily reject them by merely pointing to competing conceptions, instead of engaging with the merits of the argument. Next to disadvantages in the philosophical debating arena, there is the more theoretical danger of drawing misleading, if not outright false conclusions. We might, for example, come up with a false characterisation of a given phenomenon.

I aim to illustrate these points by focussing on a particular debate that has been plagued by this methodological shortcoming, namely the wrong kind of reason debate. The terms *reasons of the right kind* (RKR) and *reasons of the wrong kind* (WKR) have been used to draw a distinction amongst the *normative reasons* for belief in a proposition p, namely reasons that count in favour of the belief that p.[1] This distinction is motivated by the intuition that even though many considerations count in favour of belief in a proposition, only some provide the right kinds of reasons for believing it. Take an example: I set myself the goal of running a marathon and started to go on regular long-distance runs. That I have completed this training counts in favour of the belief that I can run a marathon. Now compare the fact that I have trained appropriately to the fact that confidence in my physical ability increases the chances of me actually finishing a marathon. This also counts in favour of my belief that I can run a marathon: if I have this belief, I am more likely to achieve my goal, all else being equal. So both considerations, that I have trained appropriately and that confidence increases chances of success, count in favour of believing that I can run a marathon. But only the first one seems to do so *in the right way.*[2]

Several attempts have been made to characterise reasons of the right kind.[3] A popular strategy that I focus on here is the appeal to earmarks. This strategy consists in two steps. First, a list of earmarks is developed that is supposed to capture features of WKR and RKR respectively. Second, token reasons are categorised as WKR or RKR depending on whether they display these earmarks. Schroeder (2012) provides an example of such an earmark strategy. He proposes three 'earmarks' of RKR:[4]

[1] These terms have their origin in discussions of fitting-attitudes accounts of value and, more recently, in buck-passing accounts of value (e.g. Brentano, 1969 [1889]; Scanlon, 1998). Rabinowicz and Rønnow-Rasmussen (2004) name it the "WKR problem", which takes a slightly different form here: there can be reasons that count in favour of adopting a pro-attitude towards a particular action, even though this action does not actually seem to be valuable (see Crisp's (2000) famous example of preferring a saucer of mud if an evil demon would otherwise inflict pain on you). So there can be a mismatch between what is valuable and what we have reasons to adopt a pro-attitude towards. This creates a problem for fitting-attitudes analyses unless they can rule out such mismatches by identifying these reasons as reasons of the wrong kind. But providing a general rationale for telling apart right from wrong kinds of reasons in this context has proven difficult. This part of the problem is similar to the one that I am concerned with here.

[2] At this point, one might wonder whether the fact that confidence increases my chances of success really counts in favour of believing that I can run a marathon *in the wrong way.* This might be the case if what we are concerned with is adopting true beliefs or beliefs that are good candidates for knowledge; but maybe not if we are concerned with actually running a marathon. This 'wrong for what?'-line of thought is compatible with the methodological shortcoming in the debate that this chapter draws attention to.

[3] For overviews, see Gertken and Kiesewetter (2017) and Sylvan (2016).

[4] Schroeder proposes a fourth earmark: "pragmatic reasons for belief have a recognizable 'flavor' that makes them feel intuitively like reasons for other attitudes [...]". He himself thinks that this is "arguably not a proper

(1) Right-kind reasons motivate belief on their basis.[5]

(2) Right-kind reasons bear on the rationality of a belief.

(3) Right-kind reasons bear on the correctness of a belief.[6]

These earmarks are often used to exclude pragmatic reasons from the realm of RKR by showing that they do not display some or all of the earmarks and to motivate the view that only *evidential* reasons can be RKR for belief.[7] Before discussing this position, commonly known as *Evidentialism* in the next section, note that the second earmark explicitly mentions the rationality of a belief. It does so without further specifying the underlying conception of (epistemic) rationality. This is an instance of the practice that I am cautioning against. If we decide that a token reason (or indeed a whole sub-class of reasons) are WKR because they do not bear on the rationality of a belief, without specifying which conception of rationality is at work, this categorisation is conditional on endorsing this particular conception. Failure to state this clearly can result in categorisations that appear to be universal when they are in fact contingent, and as a result, we may prematurely dismiss token reasons (or entire sub-classes of reasons) from the realm of RKR.

Before moving on to my discussion of Evidentialism and the potential of pragmatic reasons being RKR, let me briefly consider some related but different methodological concerns that have been raised in the literature.

Hieronymi (2013) takes issue with excessive reliance on earmarks in general. She argues that we should not treat first characterisations of phenomena as actually capturing the nature of these phenomena. These first characterisations should be treated as nothing more than initial guides that allow us to build an account, which we can then use to revise the first characterisation – rather than treating this characterisation as set in stone. In other words, we should not treat pre-theory as theory.[8] Schroeder's use of earmarks does not always steer clear of this danger. Whilst he initially describes the earmarks as "striking features, which do, pretheoretically, characterize the difference [...]" (Schroeder, 2012: 458) between the two kinds of reasons, he later seems to rely on them as individually necessary and jointly sufficient conditions for being a RKR: "if

earmark in its own right" and does not spell it out further (Schroeder, 2012: 460).

[5] This is connected to the debate about doxastic voluntarism, which I bracket. For arguments for and against, see Ginet (2001) and Williams (1970). Moreover, the earmark assumes a strong link between evidential reasons and our ability to form beliefs on their basis. This link has been questioned by e.g. McCormick (2014) and Reisner (2009a).

[6] Whilst this is often treated as a truism, there are accounts which question whether truth is the standard of correctness for belief (e.g. Gibbons, 2013: 90).

[7] Occasionally, the expression 'epistemic reasons' is used as well. I understand evidential reasons as a specification of epistemic reasons, where 'epistemic reasons' just is a placeholder for the category of reasons that are reasons of the right kind for belief.

[8] In a similar vein, Rabinowicz and Rønnow-Rasmussen (*Ethics Discussions at PEA Soup: Rabinowicz and Ronnow-Rasmusson on Schroeder*) wonder whether Schroeder intends the earmarks to function as a definition of the RKR/WKR-distinction.

these considerations bear all of the marks of right-kind reasons, they *are* right-kind reasons [...]" (ibid.: 466).

Whilst being sympathetic to Hieronymi's words of caution, the point I want to make is a more general one. Hieronymi's criticisms derive their force from an assumed mismatch between pre-theory and theory. We only have to worry about excessive reliance on pre-theoretical earmarks if there is a danger of pre-theory and theory coming apart. So in an ideal case, where our pre-theoretic conceptions perfectly track the correct theory, we could happily rely on earmarks to demarcate phenomena. My point applies even in the ideal case: whenever we use theoretically laden features to characterise or demarcate certain phenomena, it is crucial to specify which conception of these features is employed in order to avoid begging the question. Schroeder's earmark strategy faces this problem, but not uniquely.

There are a number of criticisms in the literature that share my more general methodological concern about implicit reliance on conceptions in the investigation of a subject matter in cases where there is disagreement about the correct conception. In her argument for unity of theoretical and practical norms for belief and against the distinction between WKR and RKR, McCormick (2014; 2018) argues that the underlying conception of 'reason' matters greatly for the outcome of this investigation.[9] For example, if the underlying conception of 'reason' already includes a commitment to only evidential reasons being RKR, it is not surprising that pragmatic (or in her terminology, practical) reasons qualify as WKR. McCormick argues that such an underlying conception of 'reason' is not the only option available, hence opening up the space of RKR. This is what we agree on. But McCormick is doubtful of the distinction as such and thinks that the already familiar distinction between *good* and *bad* reasons is sufficient. It can do all the work that the WKR/RKR distinction is supposed to do.[10] In particular, McCormick's view can allow for pragmatic reasons to be reasons for belief, albeit bad ones. In contrast, the example of a pragmatic reason of the right kind that I develop in this chapter is intended to not only be a reason for belief, but also a potentially *good* reason for belief. What is more, I do not take my points to question the distinction between RKR and WKR per se. Instead, I only intend to cast doubt on one particular characterisation of it.

Further traces of my point can be found in Reisner (2013). Again, his target is a slightly different one. He is concerned with the problem that plagued us at the very beginning of this investigation, namely finding the set of genuine rational requirements (see Chapter 2). In the course of this, Reisner argues that the enkratic principle is not one of them. But he shares with McCormick and myself the methodological concern that previous commitments substantially impact the

[9] A similar point can be found in Gibbons (2009; 2010). Whilst Gibbons is not directly concerned with the RKR/WKR distinction itself, he also acknowledges that the underlying conception of 'reason' impacts the conclusions drawn from an argument. In his case, he argues against a dualist view of reasons that regards normative and motivating reasons as separate by pointing out how such dualism is motivated by a previously accepted conception of 'reason' as perspective-dependent.

[10] Against this, D'Arms and Jacobson (2014) argue that the WKR/RKR and good reasons/bad reasons distinctions cut across each other.

investigation of a given subject matter. Whilst McCormick focusses on the impact of previously accepted conceptions of 'reason', Reisner and I focus on the underlying conception of 'rationality' that theorists bring to the table. What matters most for Reisner's project is whether our starting point is 'intuitionism' or 'algorithmic systematicity' about rationality (Reisner, 2013: 437).[11] On the algorithmic systematicity-conception, rationality is "principally concerned with consistency amongst an agent's mental states" (ibid.: 438). This mirrors what I have called the coherence conception in Chapter 2. He goes on to argue that the enkratic principle cannot be constructed as a genuine rational requirement on this conception of rationality. Reisner and I have different subject matters as targets. But we agree that the underlying conception of rationality can impact the resulting conclusion, resulting in our shared aim of showing the importance of specifying these conceptions. Reisner's paper does not endorse the algorithmic systematicity conception of rationality. It merely puts forward a conditional conclusion: if one accepts the algorithmic systematicity conception, then the enkratic principle is not a genuine rational requirement. This parallels my exploration of a coherence conception of epistemic rationality: if one accepts such a coherence conception, then pragmatic reasons can be reasons of the right kind for belief.

7.2.1 Which Reasons are RKR?

With the methodological points on the table, let us turn to the WKR/RKR debate to trace the shortcomings I raised and to see how this results in a premature dismissal of pragmatic reasons from the realm of reasons of the right kind for belief.

A striking difference between the two reasons in our example is that one provides evidence for the truth of the belief whereas the other does not. This motivates *Evidentialism*, which I will ultimately reject. Evidentialism in its general version accepts that there can be reasons of the right and wrong kind for belief (e.g Hieronymi, 2005), but holds that only evidential reasons can be reasons of the right kind for belief. Reisner (2009a: 258) provides a more modest characterisation of Evidentialism as the view that "there is something especially important or central about evidential reasons for belief".

It should be noted that there exists a different brand of Evidentialism, sometimes called *Strict Evidentialism* (e.g. Shah, 2006), which is committed to WKR-scepticism, i.e. the view that there are no reasons of the wrong kind for belief. On this view, a consideration either is or is not a reason for belief; it cannot count in favour of belief in the wrong way. In this chapter, I limit my discussion to the weaker version, or Evidentialism proper, so to speak.

The earmark strategy can be used to provide support for Evidentialism. By showing that evidential reasons display the earmarks of reasons of the right kind, whereas pragmatic reasons do not, it is concluded that only evidential reasons can be RKR.

[11] To be exact, Reisner takes these to be stances towards *rational requirements* rather than *rationality* simpliciter. But since he takes rationality to "comprise[s] a set of rational requirements" (Reisner, 2013: 461), I take this shortcut in my presentation to be unproblematic.

As an example, take the fact that I have completed my training. This is an evidential reason that also displays the earmarks: having completed a training plan motivates my belief that I can run a marathon. It also bears on the rationality of this belief because it is conducive to a conventional epistemically important property, namely truth. And since evidential reasons bear on the truth of a belief, it is almost a truism that they also bear on the correctness of belief.

With this in place, reasons of the wrong kind would include all non-evidential reasons. Pragmatic reasons belong to this category.[12] Pragmatic reasons are similar to evidential reasons in that both count in favour of belief in some way. But instead of being conducive to believing truly, pragmatic reasons are conducive to various other interests, such as maximising utility, producing desirable consequences or prudence. Since the consequence-based reason in the example is non-evidential, we should expect it to lack the earmarks: it does not motivate belief as easily; it is not necessarily conducive to believing truly and hence does not bear on the belief's rationality and also not on its correctness.[13] And so, following Schroeder's strategy of identifying RKR with reasons that display the earmarks, our non-evidential reason is a reason of the wrong kind for belief because it lacks the earmarks.

Note at this point that this line of argument simply uses 'rationality' as it features in the second earmark to draw a conclusion about the realm of RKR without specifying which conception of rationality is used here. Argumentative moves like this display precisely the methodological shortcoming that this chapter is concerned with.

7.3 Broadening Right-Kind Reasons: Pragmatic Reasons as RKR?

Since I want to motivate the possibility of pragmatic reasons being reasons of the right kind, I am generally sympathetic to Schroeder's argument, as he also construes the category of RKR more broadly to include some pragmatic reasons, albeit for withholding belief.[14] Unlike Schroeder, I think that there can be room for pragmatic RKR for belief, not just for its absence. This is

[12] I am inclined to think that pragmatic reasons exhaust the category of non-evidential reasons but I will not argue for this here. It does not matter for my purposes whether there are other non-evidential reasons that are not also pragmatic reasons.

[13] Strict Evidentialists like Shah (2006) will consider a lack of the first earmark decisive. For Shah, it is a conceptual truth about belief that it is transparent, i.e. that deliberating *whether to believe p* immediately gives way to the question *whether p*. He adds a deliberative constraint which holds that in order for something to be a reason for belief, it must be able to motivate belief. The combination of transparency and this constraint leads him to conclude that only evidential reasons can be reasons for belief. On this account, if pragmatic reasons cannot motivate belief, they could not be reasons for belief at all – not even of the wrong kind. Regarding the second earmark, Kelly (2002) and Shah (2006) share the view that pragmatic reasons do not bear on belief's rationality. Both argue that it is part of the nature of belief that only evidential reasons can make belief rational.

[14] More precisely, Schroeder argues against the idea that the distinction between state- and object-given reasons tracks the distinction between RKR and WKR. His argument relies on showing that there are state-given reasons that have the earmarks of RKR. The state-/object-given distinction is not my concern here.

where we disagree: Schroeder thinks that there is "a distinctive dimension of rational assessment of beliefs—sometimes called *epistemic* rationality—that is affected by the epistemic reasons of which the subject is aware but not affected by the pragmatic reasons of which the subject is aware" (Schroeder, 2012: 459). But I do not think that it is necessarily the case that there is such a distinct domain of epistemic rationality on which pragmatic reasons have no bearing. Quite the opposite: I want to show that there *is* a plausible and venerable conception of the rationality of belief which allows for pragmatic reasons. If this is the case, then I have shown that pragmatic reasons can display one of the earmarks, namely the second one. Moreover, this highlights the importance of specifying the conception of epistemic rationality at play, because it leads to different substantial conclusions. So let us turn to the second earmark and explore an alternative conception of epistemic rationality.

7.3.1 A Coherentist Conception of Rationality

The conception of rationality that I will motivate is broadly coherentist. I now provide a short characterisation of 'Coherentism' in general. I then introduce two popular ways of cashing out 'rationality as coherence', in order to lend additional support to the coherentist approach. In the next section, I look closer at the origins of these views in order to draw some parallels that will help me formulate my example of a RKR pragmatic reason for belief.

'Coherentism' can apply to a number of theories. Prominent varieties include coherentism about justification, truth and, what is most important for this chapter, rationality. Their basic claim can be summarised as 'A justified/true/rational belief is a coherent belief'. Spelling out what 'coherence' amounts to is a research field in its own right. But BonJour's (1985) five principles of coherence provide a nice overview, as they combine various different approaches:

> (1) A system of beliefs is coherent only if it is logically consistent,
>
> (2) A system of beliefs is coherent in proportion to its degree of probabilistic consistency,
>
> (3) The coherence of a system of beliefs is increased by the presence of inferential connections between its component beliefs and increased in proportion to the number and strength of such connections,
>
> (4) The coherence of a system of beliefs is diminished to the extent to which it is divided into subsystems of beliefs which are relatively unconnected to each other by inferential connections,
>
> (5) The coherence of a system of beliefs is decreased in proportion to the presence of unexplained anomalies in the believed content of the system (ibid.: 95, 98).

(1) and (4) relate to the view that some coherence relations are basically deductive relations.

In contrast, (2) and (3) are connected to the idea that some coherence relations are inductive relations; (5) is closer to the thought that coherence is in part about explanatory relations.[15]

The conception of rationality as coherence amongst attitudes corresponds to one of the two main views within the debate about the nature of rationality (see Hinchman (2013) and Section 2.1; the other one being rationality as responsiveness to reasons). Within this conception, Reisner (2013) draws a further distinction between understanding rationality as "strict consistency" on the one side and rationality as "unity" on the other side.

On the strict consistency understanding, rationality is concerned with constraining the "relations amongst the contents of an agent's mental states" (ibid.: 437), where the relevant constraints are provided by rigid formal systems like logic or probability calculus.[16] Hence, this conception reflects BonJour's first and second principle. In the realm of theoretical rationality, some of the closure requirements we have encountered are good examples for the kinds of demands strict consistency conceptions of rationality make. Take, for example:

> *(Modus Ponens Closure)* Rationality requires that [if you believe p, and you believe that p implies q, then you believe q.]

All this requirement does is to demand that an agent's beliefs conform to modus ponens. The content of the requirement is provided by a rule of inference in classical logic. In the practical realm, our familiar means-end coherence requirement is an example of a strict consistency requirement:

> *(Means-End Coherence)* Rationality requires that [if you intend to e, and you believe that intending to m is necessary for e-ing, then you intend to m.]

Reisner (ibid.) understands requirements like this as the practical analogue to the previous theoretical requirement. But as Reisner notes, rationality is not only concerned with strict consistency demands, such as can be derived from rigid formal systems like logic. Often, we are concerned with "a certain kind of psychological consistency" that cannot be captured by strict consistency alone (ibid.: 440). This requires a broader understanding of rationality as coherence, that I have not touched on until now. Reisner captures this as "unity": "Consistency between one's normative beliefs, which are a form of normative self-ascription, and one's related attitudes constitutes a rationally successful unification of an agent's psychology" (ibid.: 445). Reisner (ibid.: 444-445) proposes so-called "matching attitude requirements" as examples for

[15] For a helpful overview that cuts the debate up in this way, see Murphy (*Coherentism in Epistemology*).

[16] Most notably, Bayesianism can be seen as a coherence conception of rationality that relies on probability calculus. It considers beliefs (or rather credences) rational when they are coherent, where coherence is understood as compliance with with the axioms of probability theory in combination with a conditionalisation constraint along the lines of Bayes's Theorem.

rational coherence requirements along the lines of the unity understanding. They follow a general template:

> You are rationally required that: if you believe that you ought to have attitude A towards contents C, then you have attitude A towards contents C.

Depending on what we fill in for attitude A, these requirements are theoretical or practical. For the purposes of this chapter, it does not matter whether we subscribe to a strict consistency or a broader unity understanding of rationality. Both can be captured as coherentist conceptions of rationality. What they have in common is that they require or prohibit certain patterns of attitudes, e.g. the belief that p, the belief that if p then q, and the absence of the belief that q. So if we want to assess the rationality of, say, the belief that q, looking at that belief alone is not enough. We also have to take into account its relation to other beliefs – in this case its deductive relations – to be able to assess whether the prohibited pattern has been instantiated. Similarly, evaluating the rationality of an individual attitude A is not possible. In order to decide on its rationality, we have to consider whether it would instantiate the prohibited pattern of believing that you ought to have attitude A, but not actually having A. In other words, we have to factor in how attitude A is related to other attitudes.

7.3.2 Quinean Inspirations

These features of coherentist conceptions of rationality possess striking parallels to some claims about the nature of scientific theories as put forward by W.V.O. Quine, whose holism has been one of the most influential accounts of coherentist thought, albeit primarily within the philosophy of science. Given my focus on the rationality of belief, in what follows, I draw parallels between two features of scientific theory-building and belief-formation.

The first parallel concerns the impossibility of individual evaluation. Quine (1951) holds that since individual scientific sentences are underdetermined by evidence, it is impossible to evaluate a single sentence individually for its truth. If I want to evaluate the sentence 'The pond freezes at temperatures below 0°C', I also have to evaluate previous data points and the reliability of my equipment and my reasoning. In the more recent Quine and Ullian (1970: 16), this idea is explicitly applied to belief: "Often in assessing beliefs we do best to assess several in combination. [...] It is in light of our full body of beliefs that candidates gain acceptance or rejection; any independent merits of a candidate [belief] tend to be less decisive". For example, if I want to evaluate the rationality of my belief that my roommate cooked dinner for me, I have to take into account many other related beliefs: beliefs about my roommate's habits but also about the reliability of my sensory experience. This extends to deductive relations. With rationality requirements like deductive closure requirements, evaluating a single belief is equally impossible – we have to consider other beliefs as well as the (deductive) relations between them.

This brings us to the second parallel. Structurally, the impossibility of individual evaluation naturally suggests that an individual belief is embedded in a network of other beliefs, which are themselves connected. The parallel to theory-building would be to recognise that the sentences within a theory are members of a complex system whose individual elements bear manifold connections to the others, as opposed to a clear-cut tier system in which a few sentences provide the foundation for others. Thinking of scientific theories and belief systems as structured in this way recommends taking every element – a belief or a scientific sentence – as potentially revisable. If I were to encounter a conflict between my theory and the evidence, I could not identify an individual element of the theory as the exact source of the conflict. Hence, any of the sentences in the system that makes up a theory can be revised to make theory and evidence fit. As Quine and Ullian (ibid.: 16, 22) put it, "[w]hen a set of beliefs is inconsistent, at least one of the beliefs must be rejected as false; but a question may remain open as to which to reject. Evidence must then be assessed, with a view to rejecting the least firmly supported of the conflicting beliefs. [...] When an observation shows that a system of beliefs must be overhauled, it leaves us to choose which of those interlocking beliefs to revise."

For example, if my thermometer indicates an outside temperature of -5°C but I also see that the water in the pond is not frozen at all, I could revise any of the sentences in my system to avoid this conflict. I might give up the belief in the correct functioning of my thermometer, I might start to doubt my eyesight or I might postulate a different freezing point of water. According to Quine, even the "profoundest laws of atomic physics or even of pure mathematics and logic" could potentially be revised to preserve fit with experience. They are no exception to "[t]he totality of our so-called knowledge or beliefs [which is] a man-made fabric which impinges on experience only along the edges" (Quine, 1951: 116).

These parallels between belief formation and scientific theory-building naturally suggest themselves on a coherentist approach to rationality that relies on requirements like deductive closure and means-end coherence. Given this, I think we can go a step further and use some of the criteria that guide the process of theory formation to also guide the process of belief formation and revision.

7.3.3 Minimal Disturbance: A Pragmatic Right-Kind Reason

If all sentences in a scientific theory have to be treated as part of a system rather than as single elements, and no single one is immune to revision, which part of the system should we change on encountering a conflict with evidence? Which sentences to drop, alter or add? Here, Quine endorses a "thorough pragmatism" (ibid.: 126). The process of adapting our system of scientific sentences is guided by *simplicity* and *conservatism*: the system is to remain as simple as possible and the less invasive way of accommodating the evidence is preferred to alternatives which would require a more drastic revision. We have a "natural tendency to disturb the total system as little as possible [...]" (ibid.: 120). In his later work, Quine explicitly states conservatism as a virtue

of hypotheses:

> "In order to explain the happenings that we are inventing it to explain, the hypothesis may have to conflict with some of our previous beliefs; but the fewer the better. [...] The less rejection of prior beliefs required, the more plausible the hypothesis–other things being equal" (Quine and Ullian, 1970: 66-67).

Given the parallels, I suggest to adopt the same strategy for belief-change. There is additional reason to think that this is a sensible strategy in the belief case. Beliefs about the world and about what we should do or believe play a crucial role for agency. We rely on them all the time, for example when we make plans, judgments or decisions.[17] Without such beliefs, we would have a much harder time navigating the world and going about our lives. So there is a need for a fairly stable and continuous belief set. Moreover, our computational capacities are limited. To constantly question and revise our beliefs about the world would be difficult for us, as agents with limited computational capacities. Given this, and assuming that our belief-formation method mostly is not entirely and systematically flawed, it makes sense to adopt a conservative approach to revising our belief system. Otherwise, we risk to unnecessarily overthrow large parts of our belief set. Assuming that it takes time to form new beliefs and to restore order in our belief set, this would create a state of abeyance that, in extreme cases, might amount to a temporary loss of agency.[18]

Since conservatism allows us to maintain agency and continuity in our belief system, I will now explore how it could be put to work in the context of changing a belief system, which includes adding, dropping or altering beliefs.[19] Conservatism demands that the system in its entirety be changed as little as possible in order to accommodate incompatible evidence. The consequences that changing one belief has on other beliefs is something that bears on the rationality of belief. To see how this criterion of Conservatism could be transformed into a decision rule, consider the following toy example of a belief system (Figure 7.1).

The boxes b_1 - b_5 denote individual beliefs. The whole net of beliefs, including their connections, is the belief system S. Note that the beliefs differ with regard to their connections. Some of them are more interconnected than others. An outgoing arrow describes a support-connection, for example b_1 supports b_2 and b_3 and is supported by b_2 and b_4. In turn, b_4 has more outgoing

[17] See Bratman (1987) on the importance of plans: they "facilitate coordination both socially and within our lives, [and] help enable prior deliberation to shape later conduct". Plans are needed because "we are not frictionless deliberators" (ibid.: 28) and plausibly require a stable belief set, too.

[18] For example, see Schroeder (2009).

[19] I do not intend to develop a full theory of belief-formation here. For example, the notions of 'importance' and 'centrality' of beliefs are a lot more complex than I present them. I take the points I make to provide just the necessary background for my example of a pragmatic RKR.

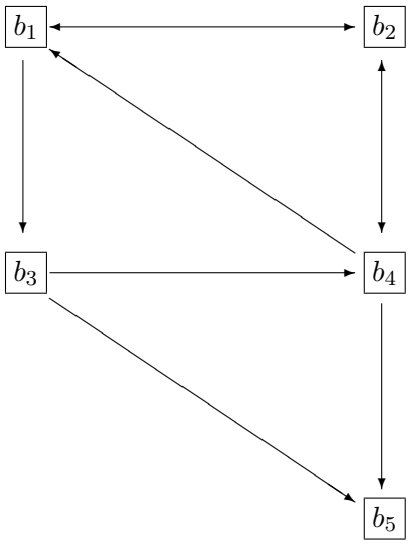


Figure 7.1: Belief System

connections than b_1 and hence plays a more important role in S.[20] Similarly, b_1 can be considered more important than b_5. Now assume that an agent with belief system S encounters a new piece of information that is incompatible with S. Now, she is required to make changes in S in order to accommodate this new piece of information.

On our coherentist picture, every single one of her beliefs is a potential candidate for revision. So which belief should the agent alter, drop or add? This is where conservatism steps in as a guiding principle for her decision. It demands that S be changed as little as possible. This suggests a decision rule along the along the lines of *Minimal Disturbance (MD)*:

(MD) Minimise disturbance of S when resolving the conflict.

This rule favours minimal disturbance by recommending the change that leads to the fewest changes in S and solves the conflict. It provides the example of a pragmatic RKR: that a particular change would minimise disturbance of S is a reason for this change. But in order to apply this decision rule, we need a way of measuring and comparing the overall changes caused by changing an element of S. An intuitive indicator for this is the number of outgoing support connections of beliefs. Call this C. Then,

[20] The notion of 'support connection' is weaker than entailment. Support connections also should not be understood as evidential relations (see Section 7.4 for why even an evidential understanding of support connections does not threaten my argument). Support connections can take a number of forms, similar to the counting-in-favour relations featured in the initial example about running a marathon: the belief that I have stuck to a training plan supports the belief that I can run a marathon; so does believing that confidence in my physical ability increases my chances of success.

$Cb_1 = Cb_2 = Cb_3 = 2;$

$Cb_4 = 3$

$Cb_5 = 0.$

This allows us to specify *(MD)* a bit more:

> *(MD*)* Change S such that (i) C is minimised for all b involved and (ii) the conflict is resolved.

Now, *(MD*)* can provide specific reasons for the change which minimises disturbance and resolves the conflict.

(MD)* places the number of outgoing support connections at the centre of the deliberation about how to change the belief system in order to accommodate the conflict. Whether an agent has a reason to add or drop a belief (or multiple ones) will depend on which of these changes minimises the overall C-score.[21]

For illustrative purposes, I assume for now that that dropping one belief would already resolve the conflict. So we do not have to consider the options of changing multiple beliefs or belief-addition for the moment.[22] The question then becomes: which belief to drop? Given these assumptions, *(MD*)* provides a reason for dropping $b_5$because Cb_5 is the lowest in S. Dropping b_5 does not require any additional change in S, as it has no outgoing support connection. Compare this to, say, altering b_4. This could potentially require three other belief alterations which, in a more complex S, could set off a chain reaction and require even more changes of beliefs. Not so for dropping b_5: this would not remove any support for other beliefs. But dropping b_4 instead would make three other beliefs less supported.

Unlike in my toy example, *(MD*)* might recommend changing a number of beliefs over changing a single one. These will be cases where changing a single belief either does not suffice to resolve the conflict or does not result in overall C-minimisation. According to *(MD*)*, the agent should opt for the change that resolves the conflict *and* minimises the C-score of the beliefs involved, even if that requires multiple belief changes.

One might want to object at this point that *(MD*)* does not actually minimise disturbance. Would it not disturb the system less if we first identified the belief whose change would resolve the conflict and then directly moved towards changing that one? There are two things I want to say about this putative alternative to *(MD*)*. One, on the coherentist picture, we cannot establish

[21] For present purposes, it will suffice to think of the overall C-score as the simple addition of the individual beliefs' C-scores. This does not exclude the possibility of a more sophisticated mode of aggregation in a more detailed version of the toy example.

[22] I discuss an example featuring belief addition at the end of this section.

a one-to-one correspondence between a single belief and a piece of incompatible information, because beliefs are embedded in a system and impossible to be evaluated individually. There might not be a single belief in the system that is responsible, so to say, for the conflict and even if there was, we would have no principled way of finding out which one it is. And so the alternative to *(MD*)* outlined above would fail at its first stage. Two, it is too quick to think that changing multiple beliefs causes more disturbance than changing a single one. When measuring disturbance, we have to take into account the beliefs' support connections. It might be less disturbing to change a couple of beliefs with low C, as this would require only few additional changes down the line, than changing only one belief with a high C, as that would potentially require many changes down the line. For these reasons, I do not consider the above suggestion a plausible alternative to *(MD*)*.

My picture has the further advantage that C is also an indicator for the importance of a belief in the system, in addition to measuring the consequences of a particular belief alteration: the higher the number of outgoing support connections, the more important a role the belief plays. This has the desirable consequence of accommodating an important objection against coherentist approaches in general. Because these approaches hold that all elements in a system are potentially subject to revision, laws of nature or logical truths do not enjoy any kind of special standing. If required, they can be changed just like any other element. Introducing a decision rule like *(MD*)* does not eliminate this possibility, but makes it the case that beliefs which take logical truths or laws of nature as their content are in fact much less likely to be revised or given up. That is because their C will be one of the highest in the whole system. A belief b_{g} about the law of gravitation, for example, will support other beliefs about the interplay between physical forces, all the way to the belief that the apple will fall on the floor if I drop it. Hence, it is very unlikely that b_{g} will ever be revised because there will be many other beliefs with a lower C that will be changed first. And if the encountered conflict with experience is so profound that revising all the other beliefs with lower C did not resolve it and b_{g} would be the next on the list, then the conflict must be of such fundamental nature that it deserves this kind of radical revision. But again, conflicts of this kind are rather unlikely to occur.

To see how *(MD*)* could provide a reason for adding instead of dropping a belief, consider a historical example: Dr Ignaz Semmelweis's famous observations about the connections between childbed fever and handwashing (*The Dirty History of Doctors' Hands*). In the 1840s, maternal mortality rates after giving birth in one of Vienna's two public hospitals were a lot higher than maternal mortality rates after having given birth outside of these hospitals, such as at home, in the streets or in the fields. There also was a significant difference between the two hospitals, with approximately 1 in 25 women dying in the public hospital that trained midwives, as opposed to approximately 1 in 10 women dying in the public hospital that trained medical students. Faced with this evidence, chief resident Semmelweis famously postulated a connection between having contact with cadavers and risk of infections. This would explain why mortality rates were a lot higher in the hospital that trained medical students, who had regular contact with

cadavers during autopsies. As a result, he decided to make chlorine handwashing mandatory before assisting with a birth, resulting in drastically decreasing maternal mortality rates.

This example illustrates how *(MD*)* could provide a reason for adding a belief. Semmelweis had to reconcile a number of his beliefs with incompatible information, for example the belief that midwives are better trained to deal with birth-related complications than medical students, that medical students are better trained to deal with birth-related complications than persons without medical training and that receiving medical attention is safer than not receiving it. The incompatible information is constituted by the drastically higher mortality rates connected to births within rather than outside hospitals.

Dropping or changing any of these beliefs would likely have demanded many more changes to Semmelweis's belief system. Presumably, he had good reason to believe that receiving medical attention is better than not receiving it, such as sick people being more likely to survive in hospitals than on the streets. Changing his belief about the competences of midwives and medical students, respectively, also would have required many more revisions, for example of beliefs about how the content of medical training impacts skill.

But adding the belief that contact with cadavers increases the risk of infection does not require any of those revisions. It accommodates the incompatible information whilst maintaining the other beliefs. This is an example of how *(MD*)* could favour adding a belief over altering or dropping one, thereby giving us a pro tanto reason *for* belief.[23]

7.3.4 *(MD*)* and AGM

At this point, the reader might feel reminded of the AGM framework for belief revision. And indeed, there exist a number of parallels between the view I sketched and the framework. But they also differ in important ways. In this section, I explore some of these similarities and differences.[24] The most important point to take away from this discussion is that the similarities with an established and well-worked out framework like AGM only lend more support to my methodological claim. Even if one was not convinced by the example of a pragmatic RKR that I sketched via *(MD*)*, there exist alternative examples, similar in spirit, which can also give rise to pragmatic RKR by exploiting the same methodological lacuna in adopting a broadly coherentist conception of rationality.

[23] My remarks here should not be understood as saying that adding a new belief *always* is the best way of accommodating the conflict. For example, a newly added belief might be incompatible with existing beliefs and cause more disturbance in that way. In many cases, changing or dropping a belief would be the option favoured by *(MD*)*, as illustrated by the toy example I gave earlier. Whether *(MD*)* recommends dropping or adding a particular belief (or multiple ones) will depend on its final specification, which in turn will depend on how we specify the toy example to obtain a fully fledged model of a coherentist belief system. This is a task that I cannot take up in this chapter.

[24] My presentation of the AGM framework relies mostly on Kyburg and Teng (2001), which largely follows Gärdenfors (1988). The main tenents of the AGM framework have their origins in Alchourrón and Makinson (1982), Alchourrón and Gärdenfors and Makinson (1985), and Gärdenfors and Makinson (1988).

The AGM framework represents belief systems as logical theories, in which each individual belief is represented as a sentence. This feature entails one of the two main similarities between AGM and my picture: both share a concern for coherence. More precisely, AGM requires that the belief system remain logically consistent. In this regard, it can be seen as a specification of what Reisner called 'strict consistency' conceptions of rationality, which is not equivalent to but compatible with broader coherentist conceptions, such as the *(MD*)* picture.

Like in my toy example, belief systems, or theories, can be changed in three different ways in the framework. We can either add a belief (or sentence), we can drop a belief or we can change a theory. But these three options take on a slightly different form in the AGM framework. To see this, consider first another feature that sets AGM apart from my picture. Since AGM represents belief systems as logical theories, these systems are deductively closed, i.e. they also contain all the logical consequences of the individual beliefs. The belief system on my picture only contains the beliefs the agent actually holds and not, for example, beliefs in all disjunctive statements that follow from the other beliefs. Now, if we gain new information and need to add a belief to our belief system, AGM has us add not only this belief but also all its consequences to the system, in order to maintain deductive closure. On my picture, we would only need to add the new belief itself. Whenever a belief is dropped, AGM requires that we remove not only this particular belief but also other beliefs from which the dropped belief could be derived. And in the case of changing the system, AGM assumes that a specific belief in the system has been contradicted. In order to accommodate this and to maintain consistency, we have to remove the contradicted belief and add a belief containing its negation. Further changes might be required to ensure that there are no inconsistencies between the consequences of the newly added belief in the negation and its consequences or the consequences of the existing beliefs.

This leads us to the second main similarity between my picture and AGM. They share a concern for conservatism. On my picture, this takes the form of the decision rule *(MD*)*. On AGM, this concern becomes relevant when we drop beliefs or change theories. In both cases, there might be multiple options available that would yield the desired change and preserve consistency and deductive closure. AGM demands that we choose the option which yields minimal change, where this is understood as maximising closeness to the original belief system. It is recognised that this choice cannot be made with the logical structure of the system alone but needs to appeal to "extralogical structure"instead (Kyburg and Teng, 2001: 167). This is further spelled out as "epistemic entrenchment", which tracks the reliability of the existing beliefs (ibid.: 167). The higher the epistemic entrenchment of a belief, the more reliable it is. When choosing amongst the options, less epistemically entrenched beliefs have to be preferred.

(MD)* also demands conservatism, but it does so in terms of the outgoing support connections of beliefs. It is not concerned with their reliability or epistemic entrenchedness. However, it does care about how a particular belief is connected to others. The more connected, the less likely it is to be recommended for change by *(MD*)*. So interconnectedness and epistemic entrenchedness perform

similar functions. In contrast, the number of outgoing support connections is a meaningless measure in AGM because every belief has infinite outgoing support connections, given that it entails infinite further (disjunctive) beliefs.[25] So whilst both pictures share a concern for conservatism, they do so in very different ways.

In general, AGM and my picture are similar in spirit but AGM is obviously the more restrictive framework. This is mainly because it represents belief systems as deductively closed logical systems and understands the connections between beliefs as governed by logical laws. The *(MD*)* picture, despite already being highly idealised, is closer to an actual representation of a belief system, which need not be deductively closed (we rarely believe all the consequences of our beliefs) and in which beliefs can be connected in all kinds of ways (our reasoning usually does not strictly follow the laws of logic). In turn, AGM is best understood as a model of belief revision for fully rational agents. Given the similarities, it might be possible to achieve similar precision for my view by embedding *(MD*)* in the AGM framework. Alternatively, one could embrace the latitude of my picture. After all, our subject matter here are reasons, which are notoriously messy. The support that reasons provide for belief comes in many different ways and is not exhausted by the the logical connections that exist between the believed propositions. This is part of their appeal, which might be lost if we could only apply the notion of 'reason' within very neat but restrictive frameworks like AGM.

7.3.5 Help with Detachment

Before I move on to consider some objections to the picture I have developed, let me make good on the promise that the findings in this chapter also inform the Normativity Problem as discussed in the previous chapter. Recall the problem: if we understand the Normativity of Rationality along the lines of bridge principles like *(BPR)*, we are claiming that we have reason(s) to do as is rationally required of us. Since *(BPR)* is formulated as a material conditional, we can detach the consequent, i.e. the reason statement. This seemed unproblematic on wide scope readings of rational requirements which usually require that we make a disjunction true. However, once we also accept transmission principles, which claim that reasons for ends are transmitted to means, we ended up with problematic results. For example, it now seemed to be the case that we have a reason to intend to buy cigarettes, or to kill our competitor, or to tell a sexist joke (to use just some of the examples). In other words, this combination of claims and principles licensed seemingly false reason statements. This is the challenge for the Normativity of Rationality: since rational requirements and transmission principles seem well supported, *(BPR)* seems like the weakest link and should be rejected. One of the strategies to defend *(BPR)* against this challenge was the Weak Reasons Strategy. Roughly, this strategy claimed that the seemingly false reason statements are actually unproblematic if we understand them as potentially very weak reasons

[25] This assumes that outgoing support connections could be represented as entailments on AGM, which is a possible but very restrictive reading of 'support connection', as I mentioned earlier.

that are easily outweighed. At this point, my discussion in this chapter can provide additional support to this strategy. Here is how.

Once we allow for pragmatic reasons to bear on the rationality on belief, we can thereby provide additional competing reasons to the reason we deemed the problematic result of the detachment mechanism. Take Sara's case as an example. One option available to her is to drop her belief that taking a lunch break is necessary for submitting today. And the detachment mechanism gives her a reason to drop this belief, even if it is well-supported by evidence. We have already seen a number of criticisms of this option, e.g. stemming from asymmetry worries. So this might be a prime example of why detachment is so challenging for the Normativity of Rationality. Now, the Weak Reasons Strategy already provides a way of making this result less problematic: the reason Sara has for dropping her means-end belief might only be very weak. And now that we can also draw on pragmatic reasons to bear on the rationality of belief, the likelihood of this (already weak) reason being outweighed might increase. Presumably, a decision rule like *(MD*)* would not recommend that Sara drop her means-end belief because this would not be the minimally invasive option. Arguably, this belief is well-supported by her other beliefs and she relies on it in other instances of practical reasoning. And the fact that *(MD*)* advises against dropping this belief, on the grounds of minimising disturbance, and hence on ultimately pragmatic grounds, can be seen as providing a reason *against* dropping this belief that could potentially outweigh but certainly competes with the (weak) reason *to* drop the belief that was provided by the detachment mechanism. In this way, allowing pragmatic reasons into the realm of RKR for belief can help to make the seemingly problematic reasons that are the result of the detachment mechanism less problematic, thereby ultimately bolstering the case for the Normativity of Rationality.[26]

I should add at this point that my investigation of potential RKR for belief is by no means complete. *(MD*)* is only one example of how such reasons could come about. There might be other sources for pragmatic RKR for belief. This would only increase their potential to supplement the Weak Reasons Strategy, because we could draw on a larger range of cases where pragmatic considerations might provide reasons that can compete with or outweigh the weak detached reasons.[27]

[26] One might worry that is is unclear whether we can weigh pragmatic reasons against non-pragmatic (e.g. evidential) reasons, and even if we can, how we should go about it. But note that what I propose here does not require a full-blown commitment to the commensurability of pragmatic and non-pragmatic reasons. All we need to accept is that pragmatic reasons (e.g. a reason to drop a means-end belief, provided by a coherence rational requirement and a transmission principle) can compete with other pragmatic reasons (e.g. a reason not to drop a means-end belief, provided by *(MD*)*).

[27] For what it is worth, I think that it is also possible for pragmatic RKR to provide attenuators for the problematic detached weak reasons, although I do not have space to engage with this possibility here. Another way of broadening the support that the findings of this chapter can lend to the Weak Reasons Strategy would be to extend the focus to intentions, rather than just beliefs. This should be an easier task since pragmatic considerations are more often seen as RKR for intentions already.

7.4 Objections & Replies

Now that we have seen how the findings of this chapter inform and supplement the beginnings of my defence of the Normativity of Rationality, one might be even more worried about potential objections to the case for pragmatic RKR for belief that I have put forward. In this section, I will try to calm these worries.

One might object that *Minimal Disturbance (MD*)* does not provide pragmatic reasons. In the case of altering or dropping a belief, the objection would have it that the fact that changing b_5 causes only minimal disturbance to the belief system really is just an evidential reason. This might be a tempting line of thought since *(MD*)* relies on the support connection between beliefs. If these are interpreted as evidential relations, it seems plausible to consider the reasons provided by *(MD*)* to be evidential, since they pertain to evidential relations. But on closer inspection, this does not entail that *(MD*)* provides evidential reasons. On an intuitive understanding of 'evidential reasons' for belief, these reasons bear on the truth of belief. But the reasons provided by *(MD*)* do not. To see this, note that b_5 is well supported in the system. If we assume that the belief system is well adjusted to the truth, then it would be surprising, to say the least, if evidential reasons recommended changing or altering a belief that is evidentially well-supported. Of course, belief systems are not always well-adjusted to the truth. But even on a somewhat distorted belief system, *(MD*)*'s reasons cannot be interpreted as evidential reasons. That is because there is no direct correspondence between the reason it provides (namely that changing b_5 would minimally disturb the system) and the truth of b_5. Instead, this reason is entirely pragmatic – it is grounded in the consequences of changing b_5.

The objection would run similarly for adding the belief that contact with cadavers increases the risk of infection. Again, it would be argued that *(MD*)* actually provides an evidential reason. This is a very tempting conclusion to draw in this case, because the progress of science has provided additional support for Semmelweis's postulation. We now know for a fact that cadavers carry bacteria that can cause infection. But this is just a coincidence in which a pragmatic reason produced a true belief and cannot be taken as proof that the reason was evidential in the first place. *(MD*)* did not count in favour of adding the belief about risk of infection because of its truth, likely or not, but only because this addition would lead to minimal revisions in the belief system. It would have equally recommended a false belief, if this would have had the consequences of minimally disturbing the belief system.

Nevertheless, it is likely that following *(MD*)* will generally produce true beliefs. If we accept my earlier stipulation that our belief formation method is not systematically flawed, then it is reasonable to assume that most of the beliefs in our belief system will be true, or at least evidentially supported. Following *(MD*)* has the effect of preserving the coherence of our belief set, which presumably contains many true beliefs. So, the changes that it recommends will be conducive to the belief system's coherence, and, as a likely side effect, lead to more true beliefs

as well. But importantly, this is only a side-effect: *(MD*)*'s reasons are ultimately pragmatic and coherence-based.

As it stands, *(MD*)* does not provide evidential reasons. It does not pertain to the truth of the belief changes it counts in favour of. Instead, it only takes into account the consequences of the potential change and recommends the minimally invasive change. I do not deny that there could be conceptions which construct the very notion of 'evidence' in a coherentist way. I take it that this would involve understanding evidence as bearing on the coherence of a belief instead of its truth. On coherence theories of truth, which understand truth as coherence with other propositions, these two might even coincide.[28] But this is a fairly radical position that should be independently motivated, especially if it is supposed to function as an objection to the pragmatic status of *(MD*)*'s reasons. Still, my claims are compatible with the commitments of a proponent of a coherence theory of truth or evidence. In a sense, they inhabit the area at the far end of the theoretical space that I have carved out. In this area, reasons that we traditionally consider pragmatic are subsumed under evidential reasons. This is in harmony with my intention to broaden the category of RKR for belief to reasons that are not strictly truth-related. Whether we call them pragmatic or evidential (on a coherence theory of evidence) or epistemic does not matter. What is more, this possibility only lends additional motivation to the second point I argued for – namely, the importance of specifying the conceptions of epistemic rationality at play. Without such specifications, it is possible that we actually agree on appropriate candidates for RKR (for example *(MD*)* and truth-related reasons), whilst being stuck in a superficial disagreement caused by the labels 'pragmatic' and 'evidential' or 'epistemic'. Theorising about and using distinctions like RKR/WKR then runs the risk of being confusing and pointless.

Now, with the pragmatic status of *(MD*)*'s reasons established, one might object that precisely this result casts doubt on the plausibility of the coherentist conception. Any conception of rationality which holds that pragmatic reasons could be RKR should immediately be dismissed, the objection goes.

In reply to this, I can only refer back to the methodological shortcoming of this debate that I have alluded to at various points in the chapter: the underspecification of the conception of rationality at issue. To illustrate, consider the structure of the debate so far. We started out with a list of characteristics of reasons of the right kind which we dubbed 'earmarks'. We then used this list to organise the domain of reasons into right-kind and wrong-kind reasons for belief. The objection departs from this structure. It wants to use the resulting categorisation of reasons as an argument for or against a conception of rationality.

But this argument only works if we can treat the set of RKR as a criterion for a plausible conception of rationality. A conception of rationality is only plausible if it results in this specific set of reasons as RKR, so the thought goes. There are two problems with this. One, prior to

[28] For an overview, see Young (2018).

investigation, we simply do not have the set of RKR. The whole point of the WKR/RKR debate and the use of earmarks is to determine which reasons belong to the set of RKR. So we cannot use something that we do not yet have to judge conceptions of rationality.

Two, even if we have agreed on a set of RKR, we cannot use this set as an arbiter between plausible and implausible conceptions of rationality. That is because the concept that we are trying to evaluate (rationality) is contained in the criterion we use to evaluate. We only arrive at a set of RKR through consideration of the earmarks, one of which contains rationality. This is unhelpful and potentially circular. Compare this to a situation in which the explanandum is already contained in the explanans. It would not be very enlightening to answer the question 'Why is it raining?' with 'Because there is rain'. It would be similarly pointless to try to shed light on what plausible conceptions of rationality there are by stipulating that they have to characterise the reasons that bear on rationality as RKR.

This points to a third difficulty with the objection's strategy: the second earmark is neutral with regards to the conception of rationality at play. The objection exploits this neutrality by assuming a more evidentialist conception of epistemic rationality for the second earmark. This then leads to a different set of RKR, which obviously does not contain pragmatic reasons. But this should not be surprising. Of course, the coherentist conception looks unattractive if we evaluate it with a set of RKR that has been provided by an evidentialist conception. Given that in the dialectic of the objection, we are in the business of finding plausible conceptions of rationality, we cannot assume that the best interpretation of the second earmark is an evidentialist conception. This would be deciding on an answer before we settled the question. Given these three problems, I contend that the objection does not pose a problem for my argument.

7.5 Concluding Remarks

Let me reconnect the preceding points to the aim of this chapter. I wanted to highlight a methodological shortcoming in the WKR debate by showing how exploiting this shortcoming opens up the possibility of pragmatic reasons bearing on belief's rationality. I conceived of this as an instance, or case study, of a more general methodological concern that has been underlying this entire thesis so far, namely the call for more complexity, specificity and explicitness in debates surrounding rationality. I then explored and motivated a coherentist conception of rationality, which paved the way for a decision rule *(MD*)* that provides pragmatic, non-evidential reason to alter, add or drop a certain belief because this would minimise overall change in one's belief system and hence is associated with less theoretic cost. So, *(MD*)* gives us a pragmatic reason that has an earmark of RKR. We have also seen how this finding can supplement the Weak Reasons Strategy in my ongoing defence of the Normativity of Rationality, which I will conclude in the next chapter.

This is not to say that all pragmatic reasons bear on belief's rationality, understood as being

appropriately connected to and supported by the other elements in the belief system and being the result of rational system-adaption. I am also not committing myself to saying that pragmatic reasons cannot be outweighed by other (evidential) reasons. But even if that were the case, they would still bear on belief's rationality. So, unlike Schroeder and the Evidentialists, I think there is at least one plausible conception of epistemic rationality on which pragmatic reasons bear on the rationality of belief.

Finally, let me finish with some considerations about why the predominant view takes pragmatic reasons to be reasons of the wrong kind. A possible explanation of why authors have been so quick to dismiss the possibility of pragmatic RKR might be found in the third earmark: correctness. Here, we deal with a different standard of evaluation. If what is at issue is the correctness of belief, surely evidential reasons are the right kind of reasons. Pragmatic reasons do not bear on a belief's correctness. But evaluating a belief for correctness is not the same as evaluating a belief for rationality. If the standard of evaluation is rationality, truth just is not all that matters, at least not on some plausible conceptions of rationality, and this opens up the domain of RKR for non-evidential, pragmatic ones. The classification of non-evidential reasons as WKR might be based on a conflation of standards of evaluation: we are evaluating rationality with the tools of correctness. Once we realise that the second and third earmarks correspond to different standards of evaluation, it is not very surprising that pragmatic reasons can satisfy one standard but not the other.

This casts doubt on characterising right and wrong kinds of reason with respect to different standards of evaluation. It might be more plausible to make the distinction specific to the relevant standard. But exploring this suggestion is beyond the scope of this chapter. I conclude that on a coherentist conception of rationality, pragmatic reasons as provided by *(MD*)* can bear on belief's rationality and be RKR.

Chapter Eight

Building New Bridges: Commitment and the Normativity of Rationality

8.1 Introduction – A Solution to the Normativity Problem

Recall, Chapter 6 highlighted an additional problem for the Normativity of Rationality – even if we accept wide-scoping as a solution to the initial detachment problem that bridge principles like *(BPR)* face, then we might still end up with reasons where there are none, if we also accept transmission principles. I have argued that this problem can be solved or weakened in a number of ways, which already did some work to establish the Normativity of Rationality. The previous chapter provided an additional source of support for one of these strategies that fell out of a case study of what can happen if we take seriously the methodological undercurrent of this thesis. I will now finally turn my attention to my positive contribution to the debate. The aim of this chapter is two-fold: I argue that the notion of *commitment* can solve the Normativity Problem for rationality, thereby establishing that rationality is normative; and I make a case for accepting *commitment* as a genuine member of the normativity family.

Accordingly, I propose to understand the Normativity of Rationality through the notion of *commitment*: whenever we are rationally required to do something, we have incurred a commitment to do so. This solution outperforms the existing proposals of understanding the Normativity of Rationality in terms of *reasons* or *ought* and also does justice to our practice of attributing rationality to agents. In addition, I show that commitment is indeed a plausible and genuinely normative notion that occupies a unique place in the normative landscape.

Before I turn to my arguments, briefly recall the version of the Normativity Problem I am concerned with. In short, we are looking for the correct way of spelling out this template of a bridge principle:

> If A, $B \models C$, then (normative claim about believing A, B, and C).

Given this, I can phrase my claim more precisely as follows: the correct normative claim, i.e. the correct normative notion that we can link rational requirements to, is commitment. Hence, I defend the following bridge principle:

> *(BPC)* If rationality requires you to x, you are committed to x.

8.1.1 *(BPC)* as the Solution

Let us see how commitment can provide not only a true bridge principle, but also one that sits well with my view of rationality attributions and the centrality of attitudes. *(BPC)* links rational requirements to the normative notion of commitment. Because commitment is not a paradigmatic normative notion, we need to spend some time understanding its features. The most comprehensive and detailed discussion of commitment and its features can be found in Shpall (2014), which I largely follow here.

In the moral realm, commitments are familiar. When I make a promise to pick up my friend from the airport, I have incurred a commitment to pick her up. If I then do not pick her up, I have done something wrong. I am criticisable in some way. But this is not to say that my commitment could not be overridden. There could have been mitigating circumstances, like me having to take my mother to the hospital instead. This could be an instance of me violating a commitment without having failed morally, but it would not change the fact that I have violated something – I did not honour my commitment. This simple case highlights a number of features of commitment that make it a normative notion that is distinct from the paradigmatic notions of *reason* or *ought*.

Most distinctively, commitment is a strict, yet pro tanto notion. Its strictness has it that commitments exercise significant normative force on agents – whenever a commitment is violated, the agent has fallen short in some way, e.g. by breaking a promise. But the pro tanto nature of commitment allows for the possibility of it being defeated. Unlike *ought*, pro tanto notions do not specify what the agent should do, all things considered. There can be competing and mitigating considerations, making it the case that what one should do overall is to violate a commitment (e.g. to take one's mother to the hospital). But they would have nevertheless gone wrong in some way by violating this commitment. Contrast this with strong pro tanto reasons: even if I have strong pro tanto reasons to pick up my friend from the airport, I need not have violated anything, or not done anything wrong, if I do not act on these reasons, since they might be outweighed by others. This unique combination of being both strict and pro tanto makes *(BPC)* a very attractive bridge principle, as we will see in more detail later.

What is more, commitments are agent-dependent and escapable. Whether a commitment exists depends on the agent. In the moral case, it depends on their actions, e.g. on whether they made a promise. Regarding rationality, commitments are incurred in virtue of having certain attitudes. For example, an agent incurs a commitment to believe q in virtue of believing p and believing

that if p, then q. Because the existence of a commitment depends so crucially on the agent, they can also be escaped or undone. In the moral case, this requires the other agent to release one from their commitment. The case is simpler for rationality: all it takes to escape a commitment is to change one's attitudes that gave rise to the commitment. This strong connection to agents and their propositional attitudes makes *(BPC)* not only a true bridge principle, but also one that can do justice to our practice of criticising agents on the grounds of their (ir)rationality.[1]

This is an additional advantage of this bridge principle and the turn to commitment. I have pointed out at numerous occasions that the crucial role rationality plays in our practices of criticising agents is one of the main motivations for thinking that rationality is normative. And given my account of rationality attributions, this also means that *attitudes* are crucial for these practices, since rationality attributions are ultimately based on sets of attitudes (and the corresponding applicable requirements). To further illustrate this familiar point, consider the following example.

We can imagine an agent, Adam, who believes that the world was created in six days (see ibid.). Is he rational in having this attitude? On the one one hand, we might think so, because Adam also believes that the Bible contains only true statements. On the other hand, we might think that he is not, because Adam also believes that the findings of geology and evolutionary theory are reliable. Relative to the set of attitudes containing his belief that the Bible contains only truths and that the world was created in six days, we can say that Adam is rational – his attitudes cohere with each other. But relative to a set of attitudes containing his belief about the reliability of geological findings and his belief that the world was created in six days, Adam seems irrational. He is violating some coherence requirement of rationality.

One might respond by denying that there is any puzzle here. Maybe the thought is that if Adam violates a rational requirement, he simply is not rational and that is all we should say about the matter. But I think this would misrepresent the situation. There is a positive rationality evaluation available for Adam. His belief that the world was created in six days does cohere with some set of his attitudes. By making our rationality attributions specific to sets of attitudes, we can account for this positive rationality evaluation without committing ourselves to an overall rationality attribution.[2]

Furthermore, there is independent reason to think that attitudes play a crucial role when determining the rationality of agents. As discussed in Section 2.1.1.1, one leading conception of rationality holds that rationality requires coherence amongst an agent's mental states. On this coherence conception, rational requirements require agents to display certain patterns amongst their attitudes, e.g. intentions to ends being accompanied by intentions to the means for these ends, and to avoid other patterns, e.g. a belief that p and a belief that not-p. If we accept

[1] This is reminiscent of Southwood's (2008) notion of a 'first-personal standpoint' on which he builds his defence of the Normativity of Rationality.

[2] Note that this is another application of what I called the 'Weak Rationality' Strategy in Chapter 6.

the coherence conception in addition to my claims about rationality attributions, attitudes are absolutely crucial in determining rationality: they determine which rational requirements apply to an agent and also feature in the content of these requirements.[3]

To sum up, acknowledging the centrality of attitudes for determining rationality enjoys two-fold support, from my account of rationality attributions and from the coherence view of rationality more generally. And deciding on rationality in turn determines whether criticism is warranted. Being able to issue this kind of criticism is a main motivation for thinking that rationality is normative. And this centrality of attitudes is reflected in understanding the Normativity of Rationality in terms of commitment, which is a key advantage of the bridge principle *(BPC)* that I defend here.

8.2 The Case for Commitment

Turning to commitment has a number of advantages: unlike its competitors, the bridge principle it provides holds up to scrutiny and can provide a solution to the Normativity Problem; the bridge principle also sits well with my independent claims about rationality attributions and criticism. So it provides an adequate understanding of the Normativity of Rationality and can be defended as a plausible normative notion in its own right. In what follows, I discuss each of the features of commitment in more detail, allowing me to develop these points along the way.

8.2.1 Commitment Sits on the Right Side of the Pro Tanto/Pro Toto Divide

Commitments, like reasons, are pro tanto – they only count in favour of something. This marks a notable difference to *ought*. Whenever someone ought to x, this means that x has emerged as the winner of a deliberation process in which all available options and all reasons for and against it have been considered and compared, with x being the favoured one. Being committed to x however only means that there is something that counts in favour of x, and that it is to be considered in the broader deliberation process – just like having a reason to x.[4] Having a commitment to x says nothing about whether x emerges as the favoured option or how x relates to other options. As we will see now, this feature allows *(BPC)* to outperform one of its competitors, *(BPO)*.

Being pro tanto makes commitment stand in stark contrast to *ought*, which is a pro toto notion. If it is the case that you ought to x, this means that x has in fact emerged as the winner of a deliberation process. It is what you should do, all things considered. There are no further competing considerations or reasons that could change this, since they have already been taken

[3] Note that the conclusions of this chapter do not require this theoretical commitment. Even if we accept a reasons view of rationality, attitudes still play an important role in determining which rational requirements apply to agents – the difference would lie in the content of these requirements. That aside, it should by now be obvious that my sympathies lie with the coherence view, and I will assume it throughout this chapter.

[4] This follows the now default characterisation of normative reasons put forward by Scanlon (1998).

into account. If we use *ought* to construct a bridge principle that could capture the Normativity of Rationality, we get something like the following:

> *(BPO)* If rationality requires you to x, you ought to x.

On a bridge principle like this, rationality would be connected to a normative notion that places the agent under an all-things-considered demand to comply with rational requirements. This puts forward a strong case for the Normativity of Rationality.

Unfortunately, this bridge principle gives rise to a number of counter-examples. The counter-examples all follow the same pattern. Since bridge principles are formulated as material conditionals, they allow for the now familiar detachment via modus ponens. From the fact that an agent is rationally required to x, we can detach the consequent, which features an *ought*-statement: she ought to x. This detached *ought*-statement is the target of the counter-examples – they aim to show that this statement is false, thereby falsifying *(BPO)*.

This should not be confused with the separate, albeit related detachment issue that arises in the debate about the correct formulation of rational requirements. As discussed in Section 2.1.3, there is substantial disagreement about whether rational requirements take narrow or wide scope (e.g. Broome, 2007a; Way, 2010b). To recall, the narrow scope version of requirements that are formulated as material conditionals takes the form of 'If you have attitude(s) x, then you are rationally required to have attitude(s) y'. Wide scope versions follow the form 'Rationality requires that, if you have attitude(s) x, you have attitude(s) y.' We have seen that this seemingly small difference in formulation has important consequences. For one, narrow scope requirements only leave open one way of satisfying the requirement – one is rationally required to adopt attitude(s) y. Wide scope requirements however leave open more than one way – one is rationally required to either adopt attitude(s) y or drop attitude(s) x. It is best to think of this debate as nested within the debate about bridge principles, which is our primary concern here. One's stance on the narrow scope/wide scope debate determines the antecedent in any given bridge principle. Despite this connection, the two debates are independent and I continue to remain neutral on the scope issue, but will point out later how one debate bears on the other.

The sort of cases that are invoked to falsify the detached *ought*-statement obtained from *(BPO)* often involve an appeal to competing values. It lies in the nature of an *ought*-statement that all different values should already have been taken into account. It is, after all, an all-things-considered notion. So pointing to values other than the value associated with rationality, such as instrumental value, should not make a difference to the truth of the statement. The counter-examples exploit our intuitions that often, these different values, external to rationality, do make a difference to the truth of the statement. It is specifically the pro toto nature of *ought* that is targeted here.

Take Kolodny's (2005) toy version of a closure requirement, requiring you to believe the entailments of your existing beliefs. If *(BPO)* is true, it would be true to say that you ought to, all things considered, believe the entailments of your beliefs. Whilst this may sound plausible at first sight, it has repeatedly been pointed out that this has absurd results. Harman (1986) famously coined the term 'Clutter Avoidance', which forms the basis of the clutter objection. Our beliefs entail a large number of things, many of which are trivial and completely useless for the agent. To illustrate, take the extreme case of an agent who only holds a single belief p. One implication of p is $(p \vee q)$. This means that now the agent ought to believe $(p \vee q)$, given a closure requirement of rationality and *(BPO)*. But $(p \vee q)$ entails $(p \vee q) \vee r$; and so she ought to also believe $(p \vee q) \vee r$, and so on. This strikes many as highly implausible: it is clearly not the case that what an agent should do, all things considered, is to clutter her mind with an infinite amount of trivial entailments that are of no use to her and might even prevent her from performing good reasoning elsewhere, given that our computational capacities are limited. Competing considerations such as pragmatism and utility would plausibly trump the demands of rational requirements here.[5]

Now, one might want to reply that the problem here lies with the requirement and not *(BPO)*. The toy version of a closure requirement that I have been considering might just not be restrictive enough. But the problem for *(BPO)* remains even on much more restricted requirements, like a requirement to believe the relevant and known consequences of our beliefs. To illustrate, take the leader of a party who believes that the party platform should include the best policy. Assume that they also believe that some radical proposition, like nationalising banks, is the best policy. Presumably, a known and relevant consequence of these beliefs is that the party platform should include a policy to implement the nationalisation of banks. So the party leader would be required to have this belief. On *(BPO)*, it would furthermore be the case that they ought to believe that the party platform should include a policy for the nationalisation of banks. But a convincing case can be made against this. Having such radical policies on the platform makes the party less likely to do well in elections and might complicate finding potential coalition partners; all of which would be to the detriment of the party's influence, without which it cannot effect any change whatsoever. So presumably, the party leader should not believe that the party platform should include a policy to effect the nationalisation of banks, all things considered. There are competing considerations, such as the pragmatics of politics, that cast doubt on whether holding this belief is the best course of action overall. This is another instance where it is not obvious that the requirements of rationality always emerge as the winner of the competition between competing considerations – contrary to what the all-things-considered nature of *ought* in *(BPO)* suggests. In many cases, what agents ought to do might be what is most convenient, pleasurable, pragmatic, or broadly speaking what leads to most utility – rather than what is rationally required of them.

At this point, let me point out how the related but separate scope debate bears on these issues.

[5] Note that you do not even need to be convinced that the agent ought *not* to believe $(p \vee q) \vee r$. To deny *(BPO)*, it is enough to acknowledge that it has not been settled what she ought to do.

Detachment worries along this line become a lot more salient on narrow scope views, because these leave the agent with only one option for compliance. On a narrow-scope reading of the requirements in our examples, the agent would be required to believe $(p \vee q) \vee r$ and the party leader to believe that the nationalisation of banks should be on the platform. What is more, forming these beliefs would be the only way in which they could satisfy the rational requirement. Combining this with *(BPO)* leads to an even stronger result: the agents now ought to form these beliefs, all things considered. But this seemed false, given considerations of clutter avoidance and political pragmatism.

Things look a bit different if we construe the requirements in the examples as having wide scope. Our first agent then would be required to either form the belief that $(p \vee q) \vee r$ or to give up their belief that p. Our party leader would be required to either believe that nationalisation of banks should be included in the platform or to give up the belief that this is the best policy, or that the best policy should be on the platform. This set of options is preserved when we apply *(BPO)*, which yields that the agent ought to go for one of these options. This leaves us with a way out of the clutter avoidance objection – there is another option of satisfying the requirement and the *ought*-statement provided by *(BPO)* that does not involve unnecessarily cluttering our minds. And it gives us a way of accommodating political pragmatism in the second example. For example, the party leader could satisfy the requirement, and thereby do what they all-things-considered ought to do, by e.g. dropping the belief that the best policy should feature on the platform.

Unfortunately, these appeals to wide-scoping cannot save *(BPO)* entirely. For one, it still seems objectionable that doing what one ought to do, all things considered, could correspond to simply dropping beliefs, absent any reasons to doubt their reliability. Moreover, the discussions in Greenspan (1975) and Setiya (2007) have taught us that wide-scope versions might not always be available. There are cases where wide-scope views collapse into narrow-scope views, e.g. when some of the options left open by a wide-scope requirement are ruled out by other factors, e.g. cognitive limitations or practical concerns, leaving the agent with only one way of satisfying a requirement, which brings back the full force of objections to *(BPO)*'s detached *ought*-statements.

Independently of the scope debate, *(BPC)* is not prone to counter-examples of this kind, since it features commitment, which is a pro tanto notion. It manages to avoid the counter-examples not because it avoids detachment – detachment is obviously still possible, since the structure of *(BPC)* as a material conditional has not changed. The important difference is that the resulting detached consequent is not problematic. Since commitments are pro tanto, they do not claim that its content is the best course of action, or that it corresponds to what the agent should do, all things considered. Once we realise this, it is no longer problematic to say that on *(BPC)*, in the previous examples, agents are committed to believing the entailments of their beliefs, even if this means believing long disjunctions, or believing propositions that are in tension with political pragmatism. We can allow for this and still hold that all things considered, they should not hold

these beliefs. And so we can preliminarily conclude that *(BPC)* is to be preferred to *(BPO)*, since its pro tanto nature allows it to avoid these counter-examples.

8.2.2 Commitment Sits on the Right Side of the Slack/Strict Divide

The success of *(BPC)* might prompt one to ask why we should not prefer a bridge principle that links rationality to reasons. After all, reasons are pro tanto notions too and have the additional advantage of being generally recognised as paradigmatic normative notions. But as we will see, reasons have the additional feature of being slack, which leads to problems of its own. Since commitment is a strict notion – a feature it shares with *ought* – it can outperform reasons as well. To see this, consider the following reasons-version of a bridge principle:

> *(BPR)* If rationality requires you to x, you have a reason to x.

It should be noted first that *(BPR)* can deal with the detachment problem that *(BPO)* faced in much the same way as *(BPC)*. Since *(BPR)* also connects rationality to a pro tanto notion, detachment, whilst still being possible, is no longer problematic. All we can detach is that the agent has a reason to satisfy the rational requirement, e.g. to believe a long and complex disjunction to satisfy a closure requirement.[6] This is compatible with it being the case that what they should do, all things considered, is to not believe the disjunction for reasons of clutter avoidance.[7]

It is not surprising that *(BPC)* and *(BPR)* share an advantage over *(BPO)* that is rooted in a shared feature, namely their being pro tanto. The distinct advantage of *(BPC)* comes out when we focus on their differences, namely commitment's strictness. The fact that reasons are slack can be seen as a problem for the right account of the Normativity of Rationality: *(BPR)* might be true, but it does not appropriately capture the way in which rationality is normative.

This is in part brought out by Broome's (1999) paper, where he is careful to keep rationality and reasons apart.[8] He argues that the Normativity of Rationality cannot be captured in terms of reasons: one, satisfying rational requirements need not coincide with doing what one has reason to do. For example, we are satisfying a rational requirement by believing the consequences of

[6] Note that some authors are sceptical even of detached reasons, as opposed to detached *ought*-statements (see e.g. Raz, 2005). Moreover, the nested detachment problem arising from the scope debate that existed for *(BPO)* also exists for *(BPR)*.

[7] Schroeder (2004) makes the same observation in his treatment of the scope debate, which also shows traces of the Weak Reasons Strategy as discussed in Section 6.3.1. He notes that counter-examples to wide-scope principles (in his case, a version of the hypothetical imperative) can be avoided by "replacing the "ought" itself with a weaker normative concept, such as that of a reason. [...] [W]e can replace talk about what you ought to do with talk about what there is a reason for you to do. [...] After all, the reason might simply not be very good [...]. So even if you do have some reason to do these things, we need have no worries about whether it will turn out that you ever ought to do them" (ibid.: 338, 344, 345).

[8] This also assumes the coherence framework of rationality, on which rationality is about coherent attitudes, rather than responding to reasons.

our beliefs.[9] But these beliefs might not be what we have reason to believe. If I believe that everything the Bible says is true, then I satisfy a rational requirement by believing that the world was created in six days – even though I do not have reason to believe that the world was created in six days, given the vast amount of evidence to the contrary.

Two, it need not be indicative of something being wrong if reasons conflict. Quite the opposite: since reasons are pro tanto notions, they are well-suited to deal with conflicts. When we have reasons for two incompatible courses of action, we simply weigh them against each other. In contrast, it is indicative of something being wrong if rational requirements conflict. If you violate a rational requirement by adopting a belief that satisfies another rational requirement, then something needs to be sorted out.

8.2.2.1 Commitment and Criticism

This observation leads to a third point that is not explicitly mentioned by Broome. When agents violate rational requirements, they open themselves up to criticism, regardless of whether this violation was backed up by reasons. Whenever an agent violates a requirement, they are definitely failing in some respect. For example, someone who does not believe the consequences of their beliefs is criticisable for that. Adam is criticisable for not believing that the world was created in six days, given that he believes that everything the Bible says is true. This is the case even if his initial beliefs are not backed up by reasons. There is still a dimension of criticism that is available.

What does all of this mean for *(BPR)* and its ability to account for the Normativity of Rationality? Even if *(BPR)* could successfully be defended against the objections, we might still want to reject it because it does not capture the dimension of criticism an agent is subject to when they violate a rational requirement. To violate a rational requirement on *(BPR)* would only amount to not acting on one, potentially very weak reason, which in itself is not criticisable.[10] This does not sit well with the fact that we take criticisms of (ir)rationality to be a serious charge.

(BPC) can outperform *(BPR)* here: because commitments are strict, albeit pro tanto notions, they allow for this dimension of criticism. If someone violates a rational requirement on *(BPC)*, this amounts to violating the strict normative notion of commitment, which is criticisable. To further illustrate this, consider again the parallel to a moral case. Say I have promised to pick up my friend from the airport but I have also promised to pick up my mother from the airport at the other end of town. If I go pick up my mother from the airport, I am breaking the promise

[9] Broome thinks that rational requirements should be understood as instances of the more general category of 'normative requirements'.

[10] In a sense, this is the counterpart to the 'Weak Rationality' Strategy I explored in Section 6.3.2. There, the positive appraisal available to an agent who conformed to a weak reason was only limited. Here, the negative evaluation, or criticism, that they can be subjected to if they do not conform to a (potentially weak) reason is also limited. Whereas this feature provided a solution to the problem transmission principles posed for the Normativity of Rationality, it now turns out to create an additional, different problem. This lends additional motivation to the search for a bridge principle other than *(BPR)*.

I made to my friend. I violate a commitment. But it might be the case that I have done what I should have done, all things considered; or that my breaking a commitment was backed up by reasons; given that my technophobe mother would have a much harder time making her way home, whereas my friend can easily look up directions on her phone. Nevertheless, I have still done something wrong by breaking a promise and not honouring a commitment, which makes me criticisable.

One might want to push back on the idea that agents are criticisable if they violate rational requirements (and, on *(BPC)*, commitments), even if this violation is in line with the reasons. Surely it is a good thing if e.g. someone does not believe that the world was created in six days, even if they believe that everything the Bible says is true? But this reaction relies on a confusion between strictness and pro toto-ness. These two properties must not be conflated. Whether a notion is strict or slack is just a question of how authoritative its voice is. Strict notions speak with absolute authority. But authorities can be wrong. So even though commitment speaks with an authoritative voice, this does not mean that an agent who fails to honour a commitment is not doing what they ought to do, all things considered. In this regard, strict notions can be likened to the rules of chess. The rules of chess are strict. Whenever someone breaks a rule of chess, they are definitely doing something wrong: they have broken a strict rule. But there could be many situations in which breaking a rule of chess is what one ought to do, all things considered – e.g. when playing chess with a child whom we do not want to discourage.

The analogy between the rules of chess also provides us with a better understanding of the Normativity Problem: rational requirements are like the rules of chess in that non-compliance opens agents up to one sort of criticism. But they might be unlike the rules of chess with regard to their normativity. Whilst there need not be genuine normative pressure to follow the rules of chess, there might be genuine normative pressure to comply with the requirements of rationality. And indeed, as this chapter argues, there is such normative pressure, which should be understood in terms of commitment.

This does not exclude the possibility of an agent doing what they ought to, all things considered, if they violate a requirement and hence do not honour a commitment. Only *(BPC)* allows us to capture this two-fold judgment – criticising agents for violating a requirement on the one hand, whilst leaving open the option of letting them off the hook with regard to what they should do all-things-considered on the other hand – because it features the unique normative notion of commitment, which is both strict and pro tanto.

8.2.2.2 Defending the Strictness of Commitment

At this point, one might worry that commitment is not really strict. After all, commitments are grounded in the attitudes of agents and can in principle come in very strange shapes and forms. An agent's commitments might be irrational or morally reprehensible, like commitments to believe that vaccines cause autism or to intend to lie to your friend. And they could also seem

entirely banal, like a commitment to intend to wear a green shirt on Tuesdays. Should we really think of these commitments as strict?

In reply to this, I want to point out again that being strict and being pro toto should not be confused. Granted, this seems like a natural combination and also one that we are familiar with, because we encounter it in *ought*. But the two properties do not necessarily have to co-exist in the same notion – they do not for commitment. The kind of thinking that motivates scepticism about the strictness of commitment follows the line of reasoning that leads to counter-examples against *(BPO)*. Recall that these counter-examples take the strict and pro toto nature of *ought* as a starting point and derive conclusions that seem to be incompatible with its pro toto nature. In doing so, they appeal to what the agent should do, all things considered. This makes the counter-examples forceful – of course, one should not believe that vaccines cause autism, all things considered, and there is nothing seriously wrong if one does not hold this belief. But this appeal to what one should do, all things considered, is not available in cases of commitment.

In contrast, commitments are grounded in particular subsets of attitudes that agents hold. Whether an *ought*-statement holds of an agent – whilst also taking into account attitudes that the agent has – is primarily the result of a complex weighing process, concluding in what the agent should do, all things considered. Given this, the strictness of *ought* is difficult to separate from its pro toto nature. The latter seems like a natural explanation of its strictness: you are criticisable because you are not doing what you should do, all things considered. But this is not the only available explanation of strictness, and the pro toto-interpretation of strictness also is not the only available interpretation.

When we talk about commitment, it is important not to adopt the same pattern. The strictness of commitment cannot be explained by a failure to do what you all-things-considered should do. What we need is an explanation of strictness within the framework of a pro tanto notion. The chess-analogy is once again illuminating. The rules of chess are strict in that they speak with authoritative voice. But their authority is also clearly restricted to a particular domain, namely the game of chess. If someone breaks the rules of chess, they are definitely doing something wrong (within the domain of chess). But this does not preclude the possibility that they are doing something right in another domain by breaking the rules of chess. Once we leave the domain of chess behind, other authorities in other domains become salient, e.g. morality or prudence. We enter a weighing process which might conclude in a pro toto demand to violate a rule of chess, whilst still maintaining that the agent is doing something wrong within the domain of chess.

We can now understand the strictness of commitment as featured in *(BPC)* as follows: its strictness is specific to the domain of rationality. When you fail to honour your commitment to believe that vaccines cause autism, or to wear a green shirt on Tuesdays, you are definitely doing something wrong in the domain of rationality (assuming that these attitudes are rationally required of you), but plausibly not in the domains of friendship or convenience. And it is still possible that what you should do, all things considered, is precisely to violate these commitments.

The norms of other domains can outweigh the domain of rationality.

To clarify, this section is concerned with defending the strictness of commitment. Asking whether a notion is strict is separate from asking whether it is normative. My defence of commitment's strictness has relied on the notion of domain-specificity, and I used the analogy to rules of chess to illustrate. By itself, this does not yet amount to a defence of the Normativity of Rationality. This would correspond to yet another different question, namely whether the domain of rationality is normative. It is possible for a domain to be governed by (domain-specific) strict rules without itself being a normative domain – and indeed, some might think that the domain of chess with its domain-specific strict rules of chess is an example for this. But of course, the normativity of a domain and the notions that govern it can interact. If a domain is governed by notions that are themselves normative (e.g. reasons, or ought), regardless of their strictness, then we can take this as establishing the normativity of the domain. This is why I argue in Section 8.2.4 that commitment is also a normative notion in its own right. In that way, we can establish the Normativity of Rationality, i.e. the normativity of the domain in which commitments are strict.

But for now, I have only focussed on establishing the strictness of commitment. And with this defence in place, we can conclude that *(BPC)* outperforms *(BPR)* because it can appropriately capture the role of rationality in criticism, thanks to its strictness,

8.2.3 Commitment's Agent-Dependency Captures the Normativity of Rationality Adequately

I have pointed out in Section 8.1.1 that any account of the Normativity of Rationality should give a special role to the attitudes of agents, because these attitudes play a crucial role in rationality attributions, which in turn are central for our practice of using rationality as a criterion for criticising agents. This criticism relies on whether we can attribute rationality to agents and this in turn depends on the specific set of the agent's attitudes that we are looking at. Neither *(BPO)* nor *(BPR)* are directly concerned with the agent's attitudes and so cannot give a special role to attitudes in the accounts of the Normativity of Rationality that they provide. Commitment, and hence *(BPC)*, however, is explicitly related to an agent's attitudes. This is due to commitment being agent-dependent.

Its existence and content depends on the agent: "in order to come into existence, a commitment must be grounded in an activity of the agent who is to become committed. [...] [They] depend, in a broad sense, on the activities of the agent who comes to stand in the commitment relation, it makes sense that different agents are committed to very different things" (Shpall, 2014: 154). This is an intuitive thought. Adam is only committed to believing that the world was created in six days because he holds other beliefs, like that everything the Bible says is true, and not because some feature of the world that is independent of him and his attitude(s). Things look similar for the anti-vaxxer: if they did not have the belief that vaccines cause autism, or had beliefs with different content, they would not be committed to believing that their child will

develop autism if vaccinated. This commitment might not exist at all, or have different content, as soon as their beliefs change.

To further motivate this feature of commitment, take the moral case. Moral commitments are also agent-dependent in that they depend on the agent: if I had not promised to pick up my friend from the airport, I would not have incurred a commitment to doing so. The difference lies only in what gives rise to the commitments. In the moral case, these are often practical actions of the agent. In the theoretical case, commitments come into existence in virtue of the agent holding certain attitudes.

So attitudes are at the core of commitments, which makes *(BPC)* a more than apt bridge principle to account for the Normativity of Rationality, which allows for criticisms that are also crucially focussed on attitudes. This sets *(BPC)* further apart from *(BPO)* and *(BPR)*. On the coherence picture of rationality that we have been assuming here, what rationality requires of agents crucially depends on their attitudes, and whether they fit together in a coherent way. On the competing picture, which holds that rationality consists in responding to reasons, what is rationally required of agents is determined by the reasons there are, which can but need not coincide with coherent attitudes on the part of the agent. Since *(BPC)* attributes a crucial role to attitudes, it is a much more fitting bridge principle for the coherence picture of rationality.[11]

The agent-dependency of commitment has another feature as its consequence: commitments are escapable. Because commitments depend on agents in that they are grounded in the attitudes of agents, one can exit a commitment by merely changing one's attitudes. If the initial attitude has changed or is no longer there, the commitment has lost its ground and no longer exists. Once I stop believing that vaccines cause autism, I also stop being committed to believing that vaccinating my child will cause her to have autism.[12] Shpall takes this to mean that commitments are always escapable – all it takes is a change of attitudes.

He claims that "[w]e can always rationally fulfill, or escape all of our rational commitments" (ibid.: 163). We can escape a commitment "by relinquishing the attitude(s) that constitutes the ground of the commitment" (ibid.: 153). However, there are counter-examples to this, as Setiya's (2007) cigarette case shows. Setiya's agent, by stipulation, cannot change some of their attitudes. This makes it the case that in order to satisfy *(Means-End Coherence)*, the only thing they can then do is to intend to buy cigarettes. The case shows that Shpall's claim is too strong: commitments are not always escapable. Agents in cigarette-like scenarios have acquired a commitment to intend to buy cigarettes, in virtue of their attitudes, but cannot escape this

[11] Let me note that assuming a coherence view of rationality adds additional support to *(BPC)* but is by no means necessary. If we adopt, say, a substantive view of rationality, the other arguments in favour of *(BPC)* that do not focus on the centrality of attitudes still have purchase.

[12] There is of course the issue of doxastic (in)voluntarism. If it is impossible to change beliefs at will, it is not so easy after all to escape a commitment. But Shpall's claim is not specific to beliefs and also does not rely on it being *easy* to escape such a commitment. What matters is that it is possible and that it entirely depends on the agent, unlike moral commitment which cannot be unilaterally escaped.

commitment because they are simply unable to "relinquish its ground".

So we cannot claim escapability as a general characteristic of commitment. But I also do not think that this is very problematic. The role it plays in Shpall's argument is to highlight one of many similarities between moral and rational commitment. Rejecting general escapability as a similarity does not amount to rejecting the more general claim about sameness in kind between the two commitments. It just amounts to there being one less piece of evidence. Giving up the general escapability claim for commitment therefore does not cause much damage. Because commitment so heavily depends on our attitudes, it seems right that it should be bound in the same way and face the same restrictions as our attitudes. If some of our attitudes cannot be given up in certain situations, neither can commitment.

8.2.4 Commitment Is Genuinely Normative

By now, the superiority of *(BPC)* to *(BPO)* and *(BPR)* should have been established. However, I want to address one more worry. One might doubt that a bridge principle like *(BPC)*, whilst being plausible and true, could vindicate the Normativity of Rationality because it does not involve a paradigmatic normative notion. And indeed, commitments seem to be suspiciously well-suited to solve the Normativity Problem. Have we just invented a notion that conveniently combines the properties that we need? After all, we have not said much about the normativity of commitment so far. If commitment turned out to be a merely stipulated notion that is not normative, then *(BPC)* could not establish the Normativity of Rationality, even if the principle were true. Luckily, commitment is a genuinely normative notion, and whilst it might not be considered a paradigmatic one, it is definitely one that we are familiar with, at least in the moral domain. So let us have a brief look at the kind of work that normative notions usually do for us and how commitment fares in this regard.

8.2.4.1 The Functions of Normative Notions

Normative notions allow us to make judgments, e.g. when we we condemn an act ('Stealing is wrong') or praise an agent ('She is a good student'). Such judgments can be about states of affairs or outcomes, and also about someone's intentions or their character.[13] For example, we might think that doctors who recommended thalidomide, an over-the-counter drug that was thought to alleviate nausea and sickness, to pregnant women in the 1950s did the right thing, even though the drug turned out to cause severe deformities in foetuses. This judgment is clearly not based on the outcome of the doctors' actions – after all, they resulted in thousands of new-borns with deformed limbs or organs. But there is a sense in which we cannot fault the doctors, and may even hold them in esteem. After all, they did not act recklessly and only had their patients' best interest in mind. This sense is captured if we use normative notions to make judgments

[13] There are interesting relations between this distinction and the familiar distinctions of internal vs external and objective vs subjective judgments and blame and praise. Unfortunately, exploring these will have to be left for another time.

about character or intentions, rather than states of affairs or outcomes. We can also imagine corresponding cases, where judging an outcome is at the forefront, e.g. a scrupulous business manager investing in a company whose product unexpectedly turns out to provide a cure for cancer.

The preceding points obviously gloss over a number of details. The important point is that they let us identify an important function of normative notions: we use them to make judgments, where we can further differentiate these judgments into judgments about outcomes or states of affairs on one side and judgments about intention or character on the other.

What both have in common is that they are backward-looking. When we make judgments, we use normative notions to evaluate something that has already happened. But normative notions also play an important forward-looking role: they are prescriptive in that they tell us what to do. A statement like 'You should call your grandmother' puts the agent under serious pressure to comply or to carry out the respective action. Importantly, this is the case independently of the goodness or the value of the action. Even in a case like 'You should follow my orders', where these orders facilitate some atrocious crime, the agent is still under pressure to conform. Note further that prescriptivity bears a close link to criticisability. When a prescriptive normative norm applies to an agent, this agent is prima facie criticisable if they do not comply.

This discussion draws out an important lesson. Normative notions are used in a backward-looking way, where we use them to judge states of affairs or outcomes; or intentions or character. And they are also used in a forward-looking way to prescribe courses of actions, which goes hand in hand with criticisability.

With this in place, let us briefly consider if and how commitment, alongside with our familiar notions of *ought* and *reason*, perform these functions.

8.2.4.2 Commitment and the Functions

Let us start with the backward-looking functions of normative notions. For example, *ought* as a paradigmatic normative notion is easily compatible with the practice of judging outcomes or states of affairs. If an agent, as a result of their actions, failed to do what they ought to do, this outcome would be judged negatively, since they failed to do what they should have done, all things considered.

Despite not being an all-things-considered notion, things look similar for commitment. If an agent is committed to x, and honours this commitment, we can make a positive judgment about this state of affairs (and a negative judgment in the opposite case, if they do not honour their commitment). Importantly, this is objectively the case, since the force of commitment "obtains independently of how [the agent] thinks about it" (Shpall, 2014: 149) – for example, the commitment to pick up my friend from the airport that I have incurred by promising to do so persists, even if I no longer think that this is what I should be doing.

But equally importantly, this value judgment is merely pro tanto because commitment is only a pro tanto notion. Whilst the outcomes of someone believing $(p \vee q) \vee r$ because they are committed to believing the consequences of their beliefs, or me picking up my friend instead of my mother because I am committed to keeping my promise, are judged positively to some extent, this does not mean that these outcomes are the best or most valuable overall. It is entirely possible that another outcome where the agent does not hold this belief, or chooses a different course of action, is better overall.

With regard to judging intention or character, commitment can easily enable such judgments because of its agent-dependence. We can say something positive about an agent's character if they honour their commitments, and something negative if they do not. Importantly, the reason for this is not because what someone is committed to is the best option all things considered, as would be the case if we were dealing with a pro toto notion, but because this commitment is grounded in their attitudes. Endorsing the grounds of a commitment (by endorsing the attitudes in the first place) without endorsing what arises from these grounds (the commitment) is an inconsistency that can be seen as warranting a negative value judgment. Here is how Shpall puts the point:

> "I have a reason to eat dim sum tomorrow morning – namely, the fact that I find it delicious. I have a reason to learn how to play the trumpet – namely, the fact that I've always wanted to. Yet it is perfectly okay for me to fail to act on these reasons, and countless others. But it is not perfectly okay for me to fail to satisfy my commitments" (Shpall, 2014: 158-159).

And so there is an available positive value judgment about the doctors who prescribed thalidomide – they honoured their commitment to further their patients' well-being, despite the terrible outcome.

As for the forward-looking function of normative notions, strict notions like commitment and *ought* are particularly well-suited here. Whenever they apply, they place the agent under a precise demand. Commitments "put genuine pressure on the committed agent to form the attitude to which he's committed" (ibid.: 149), and so they are clearly prescriptive; they tell the agent what to do. This also means that there is a clear form of criticism available when agents do not comply. But again, it is important to keep in mind that commitment's prescriptiveness and the ensuing criticism is different from the one we find in *ought*, since commitment is only a pro tanto notion. Even though it places agents under strict demands to comply, complying might not be what the agent should do overall. They might only be criticisable with regard to the commitment they violate, within the domain of rationality, not with regard to what they should do overall.

This concludes my discussion of the normativity of commitment. I have sketched some functions of normative notions and have explored how commitment sits well with all of them. Hence, its normativity should no longer be doubtful.

8.3 Objection: Commitments Are Normative Epiphenomena

I want to consider one final objection, which does not question the plausibility of *(BPC)* or the normativity of the notion of commitment itself, but instead argues that commitments "do not make any contribution to what an agent ought to do. Instead, this is determined only by reasons and/or obligations. Call this the objection that commitments are normative epiphenomena" (Liberman and Schroeder, 2016: 115). In order to better grasp this objection, it is instructive to consider the parallel to debates about the causal powers of mental events. According to Epiphenomenalism as it appears in the philosophy of mind, mental events have no effect on or do not cause physical events (Robinson, 2019). So if we want to draw a parallel to our debate, we are concerned with commitments instead of mental events; the weighing process corresponds to the causal relation in question; and what we ultimately care about is the effect on what we ought to do, instead of on physical events. The objection's claim then becomes that commitments do not feature in the weighing process that determines what an agent ought to do. This is how we have to understand the charge of commitments being normative epiphenomena. If successful, Liberman and Schroeder take the objection to establish that "commitments do not play a unique or ineliminable role in the normative landscape after all" (Liberman and Schroeder, 2016: 116).

Their reply to this objection is short: "It is not surprising that it should be possible to describe what agents ought to do and to believe without mentioning 'commitments' [...]. Yet we believe that the fact that commitments can be ignored does not show that they should be" (ibid.: 116). They can still play an important explanatory role in the weighing or deliberation process. Instead of directly determining what an agent should do, they can sometimes explain the result of the process more efficiently than a direct appeal to *reasons* and *ought*. But I think we can say more to disperse the worry.

For one, I hope to have shown that commitment is a normative notion in its own right, rather than just a hybrid of *reasons* and *ought*. It differs from both notions in important ways. For example, commitments are crucially connected to the attitudes of agents – their existence depends on them – whereas *reasons* and *ought* are not. But even if we wanted to conceive of commitment as a hybrid notion, this would not necessarily have to mean that commitment has no impact on the weighing process that determines what one ought to do. In fact, having commitment in the picture can lead to different results of the weighing process.

Hybrid notions are made up of two or more already existing notions. But this does not mean that they are dispensable. There are situations where hybrid notions perform differently than the original notions taken together. As an example outside of philosophy, take amphibious vehicles. Such vehicles are able to drive both on land and water, due to being a hybrid between e.g. a truck and a boat. This hybrid nature allows them to cover areas that a combination of truck and boat could not cover, e.g. shallow waters that are not deep enough for boats to float but too deep for trucks to drive in. There are of course many cases where an amphibious vehicle is

dispensable and could be replaced by alternating between using a truck and using a boat. But we cannot infer from this that amphibious vehicles are always dispensable, e.g. in shallow waters. Amphibious vehicles present solutions to situations in which their constituents fall short.

We can observe something similar for commitment. In a framework that only contains *ought* and *reasons*, the weighing process should result in one of two situations: either the existing reasons give rise to an *ought*-statement, thereby placing the agent under a strict, pro toto demand; or no decisive result can be reached, leaving the agent with potentially multiple slack, pro tanto reason-statements. With commitment in the framework, more situations become possible: the result of the weighing process could also place a strict demand on the agent without issuing an *ought*-statement; or it could result in more than inconclusive reasons, whilst still only placing the agent under a pro tanto demand.

The case of the agent who believes that vaccines cause autism can illustrate this. Presumably, there are some reasons that count in favour of there being a causal relation between vaccines and autism, like cases where children were diagnosed with autism shortly after being vaccinated. But there are also many reasons that speak against it, such as facts concerning the the unreliability of this evidence or about the difference between correlation and causation. So the result of the weighing process about what to believe certainly does not issue an *ought*-statement of the form 'You ought to believe that vaccines cause autism'. Yet, it is plausible that the agent is still under a strict demand to form this belief, given their prior beliefs about this relation and about the reliability of available evidence. This situation can be captured by an appeal to commitment: the result of the weighing process would not be a reason- or an *ought*-statement, but a strict commitment to form this belief.

This result is plausible even from the point of view of the anti-vaxxer: on their perspective, there are more reasons in favour of believing that vaccines cause autism. Yet, we would not want to go so far as to claim that the agent is under a pro toto demand to form this belief; after all, there are still many reasons against this belief. Again, commitment allows us to capture this: the demand the agent is under to believe that vaccines cause autism would only be pro tanto. Given these deliberations, it is not clear that the objection is successful. If commitments can change the outcomes of the weighing process, then maybe they are not epiphenomena after all.

But even if these points do not move the reader to accept that commitment does indeed bear on the weighing process, one can still reply to the objection that nevertheless commitment does play a "unique and ineliminable" role in the normative landscape: commitment still features a unique combination of pro tanto and strictness and hence allows for an attractive kind of two-fold criticism, as we have seen earlier. For example, it allows us to criticise someone for not picking up their friend in one sense, whilst acknowledging that in another sense, they did what they ought to do by picking up their mother instead. *Reasons* and *ought* alone could not capture this.

8.4 Conclusion

With all of this in place, we can see that turning to commitment allows us to solve the Normativity Problem of rationality and comes with a number of additional benefits. Unlike its competitors that feature the normative notions of *ought* and *reasons*, the bridge principle *(BPC)* that features commitment is much more plausible. It either avoids or can deal with the problems that *(BPO)* and *(BPR)* face. This is largely due to its unique combination of features, and in particular its being both pro tanto and strict. Not only does this set commitment apart because it occupies a new space in the normative landscape, it also allows it to accommodate our practice of criticising agents for their (ir)rationality in a way that does justice to the crucial role that attitudes play here. Moreover, it can supply a nuanced understanding of the Normativity of Rationality and is a genuinely normative notion in its own right. In this sense, *(BPC)* represents the achievement of the second goal of this thesis – to provide a positive argument in favour of the Normativity of Rationality. What is more, we have seen at various points how the two goals of this thesis interact with each other and provide reciprocal support, through their common theme of accounting for and explaining the role of rationality and attitudes in criticism.

On a more general note, this discussion has also provided interesting insights into the relations between the properties of being strict and slack and being pro toto and pro tanto and into the meaning of claiming that something is a normative notion. These positive results, along with the fruitful discussion it led to, should be enough to firmly place commitment on the map of normative notions and should motivate further research into the benefits that a turn to commitment could yield with regard to the normativity of notions other than rationality.

Conclusion

Let us take a step back and review the ground covered. I have started this investigation with the observation that attributions of rationality matter greatly when it comes to criticising or praising agents. This observation motivated the two main aims of this thesis. One, if rationality attributions play such an important role, we need a firm understanding of their structure and of how we are to make such attributions. The first half of this thesis was dedicated to the aim of providing a more accurate, flexible and nuanced account of rationality attributions. Two, the fact that they play such a crucial role for a number of evaluative purposes motivates the Normativity of Rationality, which in turn would explain why rationality attributions play this crucial role. This constituted the aim of the second half of this thesis: assessing what stands in the way of rationality being normative, and providing a positive argument in defence of the Normativity of Rationality.

These two main aims were accompanied by a methodological observation. At various points in my discussion, I pointed out that adopting a more fine-grained approach that makes implicit commitments explicit can lead to progress in our debates about rationality in general.

Here is a quick summary of my argument. In Chapter 1, I provided some examples that showed how we rely on rationality attributions when we evaluate agents, but also how making these attributions is no trivial task. Given the connections between this practice and the claim that rationality is normative, I briefly introduced the Normativity Problem for rationality.

In Chapter 2, I explored one strategy that promised progress with regard to our first aim. The hope was that if we could find the definitive list of rational requirements, we could simply read off rationality attributions, depending on whether agents satisfied these requirements. But once the plentitude of choice points about nature and form of rational requirements – and especially the absence of consensus on any of them – became obvious, I concluded that this strategy is ultimately hopeless.

Chapter 3 then approached the problem from a different perspective and focussed directly on rationality attributions. The question of where to look for compliance with requirements led to some important insights about the desiderata that an account of rationality attributions has to meet.

Taking these on board, Chapter 4 developed my answer to the question of how we should think about attributing rationality to agents. I argued that a more fine-grained approach that makes explicit the subset of attitudes to which a given attribution is relevant can meet the desiderata. What is more, it allowed us to make progress with the problem cases that have accompanied us since the beginning. Chapter 4 marks the achievement of the first aim of this thesis – we now have a firm understanding of the structure and practice of making rationality attributions.

In Chapter 5, I took some time to distinguish my account from an alternative contextual account that might be seen to naturally suggest itself. Despite some similarities, I argued that my claims do not require endorsing contextualism about rationality.

Chapter 6 marks the beginning of the second half of the thesis, which is concerned with defending the Normativity of Rationality. In this chapter, I took a closer look at common problems for the claim that rationality is normative. These arise once we introduce transmission principles into the picture. In my conclusion, I again relied on the same methodological observation and proposed that if we adopt either a more nuanced understanding of transmission, or of rationality attributions as previously developed, we can at least weaken the raised problems for rationality's normativity, if not avoid them altogether.

Chapter 7 could be seen as a case study of taking seriously my methodological observation. I showed that on a coherentist conception of epistemic rationality, pragmatic reasons can be considered reasons of the right kind for belief. This supported the conclusions of the previous chapter, by supplementing an existing strategy of defending the Normativity of Rationality.

Finally, Chapter 8 provided my preferred argument for rationality's normativity. By surveying the problems of existing solutions, I settled on an understanding of the Normativity of Rationality in terms of commitment, which I argued is a genuine but understudied normative notion. The turn to commitment avoids the problems of its competitors and also sits well with my rationality attributions account – they both share an appreciation for the importance of an agent's attitudes and hence give this thesis its title: Attitudes First.

Bibliography

Alchourrón, Carlos E., Peter Gärdenfors, and David Makinson (1985). "On the logic of theory change: Partial meet contraction and revision functions". In: *The journal of symbolic logic* 50.2, pp. 510–530.

Alchourrón, Carlos E. and David Makinson (1982). "On the logic of theory change: Contraction functions and their associated revision functions". In: *Theoria* 48.1, pp. 14–37.

Bedke, Matthew S. (2009). "The Iffiest Oughts: A Guise of Reasons Account of End-Given Conditionals". In: *Ethics* 119.4, pp. 672–698.

BonJour, Laurence (1985). *The Structure of Empirical Knowledge*. Harvard University Press.

Bratman, Michael (1987). *Intention, Plans, and Practical Reason*. Center for the Study of Language and Information.

Bratman, Michael E. (2009). "Intention, Practical Rationality, and Self-Governance". In: *Ethics* 119.3, pp. 411–443.

Brentano, Franz (1969 [1889]). *The Origin of Our Knowledge of Right and Wrong*. Routledge.

Broome, John (1999). "Normative Requirements". In: *Ratio* 12.4, pp. 398–419.

— (2005). "Have We Reason to Do as Rationality Requires? - a Comment on Raz". In: *Journal of Ethics and Social Philosophy* 1.Symposium, pp. 1–10.

— (2007a). "Is Rationality Normative?" In: *Disputatio* 2.23, pp. 161–178.

— (2007b). "Requirements". In: *Hommage À Wlodek; Philosophical Papers Dedicated to Wlodek Rabinowicz*. Ed. by Toni Rønnow-Rasmussen et al.

— (2007c). "Wide or Narrow Scope?" In: *Mind* 116.462, pp. 359–370.

— (2013). *Rationality Through Reasoning*. Wiley-Blackwell.

Brunero, John (2010). "The Scope of Rational Requirements". In: *Philosophical Quarterly* 60.238, pp. 28–49.

— (2012). "Instrumental Rationality, Symmetry and Scope". In: *Philosophical Studies* 157.1, pp. 125–140.

Cappelen, Herman and John Hawthorne (2009). *Relativism and Monadic Truth*. Oxford University Press UK.

Cappelen, Herman and Ernest Lepore (2005). *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Wiley-Blackwell.

Cherniak, Christopher (1986). *Minimal Rationality*. MIT Press.

Crisp, Roger (2000). "Value ... And What Follows by Joel Kupperman New York: Oxford University Press, £25.00". In: *Philosophy* 75.3, pp. 452–462.

D'Arms, Justin and Daniel Jacobson (2014). "Wrong Kinds of Reason and the Opacity of Normative Force". In: *Oxford Studies in Metaethics* 9.

DeRose, Keith (2009). *The Case for Contextualism: Knowledge, Skepticism, and Context, Vol. 1*. Oxford University Press.

Dinges, Alexander (2016). "Epistemic Invariantism and Contextualist Intuitions". In: *Episteme* 13.2, pp. 219–232.

Fink, Julian (2012). "The Function of Normative Process-Requirements". In: *Dialectica* 66.1, pp. 115–136.

— (2018). "The Property of Rationality: A Guide to What Rationality Requires?" In: *Philosophical Studies* 175.1, pp. 117–140.

Fogal, Daniel (2018). "On the Scope, Jurisdiction, and Application of Rationality and the Law". In: *Problema* 12, pp. 21–57.

— (forthcoming). "Rational Requirements and the Primacy of Pressure". In: *Mind*.

Gärdenfors, Peter (1988). *Knowledge in Flux: Modeling the Dynamics of the Epistemic States*. MIT Press, Cambridge, MA.

Gärdenfors, Peter and David Makinson (1988). "Revisions of knowledge systems using epistemic entrenchment". In: *Proceedings of the 2nd conference on Theoretical aspects of reasoning about knowledge*. Morgan Kaufmann Publishers Inc., pp. 83–95.

Gensler, Harry J. (1985). "Ethical Consistency Principles". In: *Philosophical Quarterly* 35.139, pp. 156–170.

Gerken, Mikkel (2013). "Epistemic Focal Bias". In: *Australasian Journal of Philosophy* 91.1, pp. 41–61.

Gertken, Jan and Benjamin Kiesewetter (2017). "The Right and the Wrong Kind of Reasons". In: *Philosophy Compass* 12.5, pp. 1–14.

Gibbard, Allan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press.

Gibbons, John (2009). "You Gotta Do What You Gotta Do". In: *Noûs* 43.1, pp. 157–177.

— (2010). "Things That Make Things Reasonable". In: *Philosophy and Phenomenological Research* 81.2, pp. 335–361.

— (2013). *The Norm of Belief*. Oxford University Press.

Ginet, Carl (2001). "Deciding to Believe". In: *Knowledge, Truth and Duty*. Ed. by Matthias Steup. Oxford University Press, pp. 63–76.

Ginnivan, Leah. *The Dirty History of Doctors' Hands*. URL: `http://www.methodquarterly.com/2014/11/handwashing`.

Greenspan, Patricia S. (1975). "Conditional Oughts and Hypothetical Imperatives". In: *Journal of Philosophy* 72.10, pp. 259–276.

Groenendijk, Jeroen and Martin Stokhof (1984). "Studies on the Semantics of Questions and the Pragmatics of Answers". PhD thesis.

Hamblin, C. L. (1958). "Questions". In: *Australasian Journal of Philosophy* 36.3, pp. 159–168.

— (1973). "Questions in Montague English". In: *Foundations of Language* 10.1, pp. 41–53.

Harman, Gilbert (1984). "Logic and Reasoning". In: *Synthese* 60.1, pp. 107–127.

— (1986). *Change in View*. MIT Press.

Hawthorne, John (2006). "Testing for Context-Dependence". In: *Philosophy and Phenomenological Research* 73.2, pp. 443–450.

Hedden, Brian (2015). *Reasons Without Persons: Rationality, Identity, and Time*. Oxford University Press UK.

Hieronymi, Pamela (2005). "The Wrong Kind of Reason". In: *Journal of Philosophy* 102.9, pp. 437–457.

— (2013). "The Use of Reasons in Thought (and the Use of Earmarks in Arguments)". In: *Ethics* 124.1, pp. 114–127.

Higginbotham, James (1996). "The Semantics of Questions". In: *The Handbook of Contemporary Semantic Theory*. Ed. by Shalom Lappin. Oxford: Oxford University Press.

Hinchman, Edward S. (2013). "Rational Requirements and 'Rational' Akrasia". In: *Philosophical Studies* 166.3, pp. 529–552.

Horty, John F. (2012). *Reasons as Defaults*. Oxford University Press USA.

Kaplan, David (1989). "Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and Other Indexicals". In: *Themes From Kaplan*. Ed. by Joseph Almog, John Perry, and Howard Wettstein. Oxford University Press, pp. 481–563.

Kelly, Thomas (2002). "The Rationality of Belief and Other Propositional Attitudes". In: *Philosophical Studies* 110.2, pp. 163–96.

Kiesewetter, Benjamin (2015). "Instrumental Normativity: In Defense of the Transmission Principle". In: *Ethics* 125.4, pp. 921–946.

— (2017). *The Normativity of Rationality*. Oxford: Oxford University Press.

Kolodny, Niko (2005). "Why Be Rational?" In: *Mind* 114.455, pp. 509–563.

— (2007). "How Does Coherence Matter?" In: *Proceedings of the Aristotelian Society* 107.1pt3, pp. 229–263.

— (2008). "The Myth of Practical Consistency". In: *European Journal of Philosophy* 16.3, pp. 366–402.

— (forthcoming). "Instrumental Reasons". In: *The Oxford Handbook of Reasons and Normativity*. Ed. by Daniel Star. Oxford University Press.

Kyburg, Henry Ely and Choh Man Teng (2001). *Uncertain Inference*. Cambridge University Press.

Liberman, Alida and Mark Schroeder (2016). "Commitment: Worth the weight". In: *Weighing Reasons*. Ed. by Errol Lord and Barry Maguire. Oxford University Press New York, pp. 104–120.

Lord, Errol (2014). "The Coherent and the Rational". In: *Analytic Philosophy* 55.2, pp. 151–175.

— (2018). *The Importance of Being Rational*. Oxford, UK: Oxford University Press.

MacFarlane, John (2014). *Assessment Sensitivity: Relative Truth and its Applications*. Oxford University Press.

— (ms). "In What Sense (If Any) Is Logic Normative for Thought?"

McCormick, Miriam Schleifer (2014). *Believing Against the Evidence: Agency and the Ethics of Belief*. Routledge.

— (2018). "No kind of reason is the wrong kind of reason". In: *Believing in accordance with the evidence*. Springer, pp. 261–276.

Murphy, Peter. *Coherentism in Epistemology*. URL: `https://www.iep.utm.edu/coherent/`.

Nair, Shyam (2016). "Conflicting Reasons, Unconflicting 'Ought's". In: *Philosophical Studies* 173.3, pp. 629–663.

Nozick, Robert (1993). *The Nature of Rationality*. Princeton University Press.

Parfit, Derek (1971). "Personal Identity". In: *Philosophical Review* 80.January, pp. 3–27.

— (2011). *On What Matters*. Oxford University Press.

Podgorski, Abelard (2016). "A Reply to the Synchronist". In: *Mind* 125.499, pp. 859–871.

— (2017). "Rational Delay". In: *Philosophers' Imprint* 17.

Quine, W. V. O. (1951). *Two Dogmas of Empiricism*. Longmans, Green.

Quine, W. V. O. and J. S. Ullian (1970). *The Web of Belief*. New York: Random House.

Rabinowicz, Wlodek and Toni Rønnow-Rasmussen. *Ethics Discussions at PEA Soup: Rabinowicz and Ronnow-Rasmusson on Schroeder*. URL: `http://peasoup.typepad.com/peasoup/2012/06/i.html`.

— (2004). "The Strike of the Demon: On Fitting Pro-Attitudes and Value". In: *Ethics* 114.3, pp. 391–423.

Raz, Joseph (1999). "Explaining Normativity: On Rationality and the Justification of Reason". In: *Ratio* 12.4, pp. 354–379.

— (2005). "The Myth of Instrumental Rationality". In: *Journal of Ethics and Social Philosophy* 1.1, p. 28.

Reisner, Andrew (2009a). "The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem". In: *Philosophical Studies* 145.2, pp. 257–272.

— (2009b). "Unifying the Requirements of Rationality". In: *Philosophical Explorations* 12.3, pp. 243–260.

— (2011). "Is There Reason to Be Theoretically Rational?" In: *Reasons for Belief*. Ed. by Andrew Reisner and Asbjørn Steglich-Petersen. Cambridge University Press.

— (2013). "Is the Enkratic Principle a Requirement of Rationality?" In: *Organon F: Medzinárodný asopis Pre Analytickú Filozofiu* 20.4, pp. 436–462.

Rippon, Simon (2011). "In Defense of the Wide-Scope Instrumental Principle". In: *Journal of Ethics and Social Philosophy* 5.2, pp. 1–21.

Roberts, Craige (2012). "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics". In: *Semantics and Pragmatics* 5, pp. 1–69.

Robinson, William (2019). "Epiphenomenalism". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Summer 2019. Metaphysics Research Lab, Stanford University.

Ross, W. D. (1930). *The Right and the Good*. Clarendon Press.

Scanlon, Thomas (1998). *What We Owe to Each Other*. Belknap Press of Harvard University Press.

Schroeder, Mark (2004). "The Scope of Instrumental Reason". In: *Philosophical Perspectives* 18.1, pp. 337–364.

— (2005). "Instrumental Mythology". In: *Journal of Ethics and Social Philosophy* 1.2.

— (2007). *Slaves of the Passions*. Oxford University Press.

— (2009). "Means-End Coherence, Stringency, and Subjective Reasons". In: *Philosophical Studies* 143.2, pp. 223–248.

— (2012). "The Ubiquity of State-Given Reasons". In: *Ethics* 122.3, pp. 457–488.

Setiya, Kieran (2007). "Cognitivism About Instrumental Reason". In: *Ethics* 117.4, pp. 649–673.

Shah, Nishi (2006). "A New Argument for Evidentialism". In: *Philosophical Quarterly* 56.225, pp. 481–498.

Sharadin, Nathaniel (2015). "Problems for Pure Probabilism About Promotion (and a Disjunctive Alternative)". In: *Philosophical Studies* 172.5, pp. 1371–1386.

Shpall, Sam (2014). "Moral and Rational Commitment". In: *Philosophy and Phenomenological Research* 88.1, pp. 146–172.

Snedegar, Justin (2014). "Contrastive Reasons and Promotion". In: *Ethics* 125.1, pp. 39–63.

Southwood, Nicholas (2008). "Vindicating the Normativity of Rationality". In: *Ethics* 119.1, pp. 9–30.

Staffel, Julia (2019). *Unsettled Thoughts: A Theory of Degrees of Rationality*. Oxford University Press.

Stalnaker, Robert (2014). *Context*. Oxford University Press.

Stanley, Jason (2004). "On the Linguistic Basis for Contextualism". In: *Philosophical Studies* 119.1-2, pp. 119–146.

Steinberger, Florian (2019). "Three Ways in Which Logic Might Be Normative". In: *Journal of Philosophy* 116.1, pp. 5–31.

Sylvan, Kurt (2016). "Epistemic Reasons I: Normativity". In: *Philosophy Compass* 11.7, pp. 364–376.

Turri, John (2017). "Epistemic Contextualism: An Idle Hypothesis". In: *Australasian Journal of Philosophy* 95.1, pp. 141–156.

Way, Jonathan (2010a). "Defending the Wide-Scope Approach to Instrumental Reason". In: *Philosophical Studies* 147.2, pp. 213–233.

— (2010b). "The Normativity of Rationality". In: *Philosophy Compass* 5.12, pp. 1057–1068.

Wedgwood, Ralph (2009). "Intrinsic Values and Reasons for Action". In: *Metaethics*. Ed. by Ernest Sosa and Enrique Villanueva. Wiley Periodicals, pp. 342–363.

Williams, Bernard (1970). "Deciding to Believe". In: *Problems of the Self*. Ed. by Bernard Williams. Cambridge University Press, pp. 136–51.

— (1981). *Moral Luck: Philosophical Papers 1973–1980*. Cambridge University Press.

Worsnip, Alex (2018). "The Conflict of Evidence and Coherence". In: *Philosophy and Phenomenological Research* 96.1, pp. 3–44.

Young, James O. (2018). "The Coherence Theory of Truth". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2018. Metaphysics Research Lab, Stanford University.