Routledge Studies in Epistemology

RESPONSIBILITY FOR RATIONALITY

FOUNDATIONS OF AN ETHICS OF MIND

Sebastian Schmidt

Responsibility for Rationality

This book develops the foundations of an ethics of mind by investigating the responsibility that is presupposed by the requirements of rationality that govern our attitudes. It thereby connects the most recent research on responsibility and rationality in a unifying dialectic.

How can we be responsible for our attitudes if we cannot normally choose what we believe, desire, feel, and intend? This problem has received much attention during the last decades, both in epistemology and ethics. Yet, its connections to discussions about reasons and rationality have been largely overlooked. The book has five main goals. First, it reinterprets the problem of responsibility for attitudes as a problem about the normativity of rationality. Second, it connects substantive and structural rationality by drawing on debates about responsibility. Third, it supports recent accounts of the normativity of rationality by explicitly defending the view that epistemic reasons and other 'right-kind' reasons are genuine normative reasons, and it does so by drawing on recent discussions about epistemic blame. Fourth, it breaks the stalemate between rationalist and voluntarist accounts of mental responsibility by proposing a hybrid view. Finally, it argues that being irrational can warrant moral blame, thus revealing an unnoticed normative force of rational requirements.

Responsibility for Rationality is an original and essential resource for scholars and advanced students interested in connecting strands of normative theory within epistemology, metaethics, and moral psychology.

Sebastian Schmidt is a Postdoctoral Researcher in the Zurich Epistemology Group on Rationality (ZEGRa) at the University of Zurich, Switzerland. His work has appeared in journals such as *Grazer Philosophische Studien*, *Erkenntnis*, *Philosophical Studies*, *Australasian Journal of Philosophy*, and *Philosophical Issues*. He is the co-editor, with Gerhard Ernst, of *The Ethics of Belief and Beyond: Understanding Mental Normativity* (Routledge, 2020). He is also a Research Associate at the African Centre for Epistemology and Philosophy of Science (ACEPS), University of Johannesburg.

Routledge Studies in Epistemology
Edited by Kevin McCain, *University of Alabama at Birmingham, USA* and Scott Stapleford, *St. Thomas University, Canada*

Seemings and the Foundations of Justification
A Defense of Phenomenal Conservatism
Blake McAllister

Trust Responsibly
Non-Evidential Virtue Epistemology
Jakob Ohlhorst

Rationality in Context
Unstable Virtues in an Uncertain World
Steven Bland

Seemings
New Arguments, New Angles
Edited by Kevin McCain, Scott Stapleford, and Matthias Steup
The Epistemic Injustice of Genocide Denialism
Melanie Altanian

Misinformation, Content Moderation, and Epistemology
Protecting Knowledge
Keith Raymond Harris

How Should We Rationally Deal with Ignorance?
A Philosophical Study
Nadja El Kassar

Responsibility for Rationality
Foundations of an Ethics of Mind
Sebastian Schmidt

For more information about this series, please visit: https://www.routledge.com/Routledge-Studies-in-Epistemology/book-series/RSIE

Responsibility for Rationality

Foundations of an Ethics of Mind

Sebastian Schmidt

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

First published 2025
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

The Open Access version of this book was published with the support of the Swiss National Science Foundation.

ISBN: 978-1-032-46717-7 (hbk)
ISBN: 978-1-032-46718-4 (pbk)
ISBN: 978-1-003-38297-3 (ebk)

DOI: 10.4324/9781003382973

Typeset in Sabon LT Pro
by codeMantra

Contents

Acknowledgments

This work has been in development since 2015. On the journey, I received feedback from so many people that I would most likely forget some of them. I would like to thank all of them. However, some special thanks are appropriate. First, I thank my doctoral supervisors, Gerhard Ernst and Christian Seidel. Both were extremely encouraging during my PhD. While Gerhard flared up my interest in philosophical puzzles and helped me to figure out the role of philosophy in a good life, Christian encouraged me to think about issues in applied philosophy and the value of philosophy in society. Second, I would like to thank various supervisors and colleagues who helped me to develop many of my thoughts in this book. When I had opportunities to talk to David Owens (during a visit at the University of Reading, in 2015, and again later at a graduate conference in Pavia in 2018) and to Pamela Hieronymi (during a visit at UCLA in 2019), I was still determined to argue (against their views) that we're ultimately only responsible for our voluntary actions, but not for our attitudes. Both made me reconsider this view fundamentally. I now ended up arguing against my earlier view in book-length. This change in view is also due to discussions with Christian Kietzmann and Benjamin Kiesewetter between 2016 and 2019. Much is owed to Benjamin's clear treatment of the normativity of rationality, which allowed me to connect my main topic of responsibility to debates on rationality. Third, in more recent years, Anne Meylan and Veli Mitova have been extremely encouraging and supportive. I thank them for their unwavering belief in my philosophical abilities. Without Anne and Veli, I wouldn't be able to flourish in the competitive space of academic philosophy. Thanks also to Pablo Hubacher Haerle and Xian He for their helpful feedback on the whole manuscript. And thanks to Kathleen McCully for her very competent proofreading of the whole manuscript.

Finally, I thank series editors Scott Stapleford and Kevin McCain for their excitement about this book project and for taking it up in their excellent *Routledge Studies in Epistemology* series, Rosaleah Stammler and Andrew Weckenmann for their guidance throughout the very smooth

publication process, as well as Sashivadana from Codemantra for their competent typesetting of the manuscript.

This book developed out of my doctoral dissertation, titled *The Problem of Mental Responsibility. Outlines of an Ethics of Mind* (Friedrich-Alexander-Universität Erlangen-Nürnberg, 2020, available with online open access). Compared to the dissertation, the present book is much shorter. After finishing my PhD, I published parts of the dissertation as papers, some of which are now again partly reprinted throughout the chapters of the present book, sometimes in revised and appropriately modified versions. Some material from my (2020c) is included in Chapter 1. A few paragraphs taken from (2020a) that I modified are spread over Chapters 2 and 6, which are otherwise original chapters. Some parts from Schmidt (forthcoming b) can be found in Chapters 4 and 6. Chapter 3 is based on (2020b), but it develops the argument in much more detail. Chapters 4 and 5 include most of (2024a, forthcoming c), but the chapters connect this work in epistemology more explicitly to debates on responsibility and on rationality. Most of Chapter 7 is a slightly modified reprint of Schmidt (2024b).

This research was supported by the Swiss National Science Foundation Grant "Towards a Relationship-Based Epistemology" Number 210979. I thank the SNSF for the financial support of this postdoc mobility year that allowed me to spend a year (in total) at the Universities of Southern California and Glasgow during which I could work on this book.

The pre-print version of this book has been supported for open access publication by the Swiss National Science Foundation (Grant Number 230016).

Part 1

The Problem of Mental Responsibility

1 Introduction

> We ought not only to *try* to have good motives and other good states of mind rather than bad ones; we ought to *have* good ones and not bad ones.
>
> —Robert Marrihew Adams, "Involuntary Sins" (1985: 12)

1.1 The Project

We hold each other responsible for attitudes in a similar way as we hold each other responsible for actions. We think that we *ought* to believe in human-induced climate change, and we even consider it to be appropriate to criticize others if they fail to believe in it. A malicious desire, like the desire for another's suffering, can rightly provoke not only disapproval, but also resentment or indignation. An emotion like anger might turn out to be unjustified, and we might owe an apology to the person who was the target of our hostile emotion. And merely *intending* to become a better person can already be worthy of praise or credit.

These are commonplaces which indicate that we are responsible not only for what we do but also for the attitudinal mental states we are in. Understanding our *responsibility* for our attitudes (if there is such a responsibility), and the *norms* governing our attitudes (if there are such norms), are the tasks of an ethics of mind.

Our practice of holding each other responsible for our attitudes can become puzzling upon reflection. Attitudes are not themselves exercises of voluntary control – mental states are nothing we, strictly speaking, *do*. They are not actions, but states. It seems that we cannot *choose* what we believe or feel. Yet, according to a widespread assumption, we are only responsible for what is within the scope of our voluntary control. According to Descartes, for example,

> only one thing in us […] could give us good reason for esteeming ourselves, namely the exercise of our free will and the control we have over

DOI: 10.4324/9781003382973-2

> our volitions. For we can reasonably be praised or blamed only for actions that depend upon this free will.
>
> (1649, art. 152)

If Descartes is right, then how can we be responsible for our attitudes?

Here is a common response. Intuitively, we are not responsible for just *being* in a state – at most, we are responsible for causing a state or for failing to avoid it. For instance, we are not responsible for just *having* a headache. Nevertheless, we can control our headaches by taking painkillers, and this ability to indirectly control our headaches can explain why we are sometimes responsible for them. Similarly, we can often control our beliefs indirectly by engaging in inquiry, our emotions by going for a walk, and our desires and intentions by actively deliberating about what is good and right. Since this ability to indirectly control our mind (sometimes) makes us at least as responsible for our states of mind as we are responsible for our headaches, an answer to the question of how we can be responsible for our attitudes seems within reach:

> *Indirect Voluntarism*: We are responsible for our attitudes only in virtue of the fact that we are responsible for our actions by which we can sometimes foreseeably influence our attitudes.

According to this view, responsibility for attitudes can always be derived from responsibility for actions (or omissions) which had some foreseeable influence on our attitudes: since attitudes are states, we can only be responsible for them *indirectly* – that is, we can only be responsible for them *in virtue of* being responsible for actions and omissions. For only the latter are things over which we exercise direct voluntary control.[1]

Recently, this common response to the problem has fallen into disrepute. For it presents our attitudes as something *external* to us – that is, as something for which we are *only* responsible in the way we are responsible for the state of our apartment, for the behavior of our dog, or for suffering pain. We seem to be, however, *directly* subject to certain evaluations in virtue of our attitudes to which we could never be subject in virtue of the state of our apartment, our dog's behavior, or our pain. For instance, it is common ground in contemporary epistemology that we judge each other to be *epistemically justified* or *epistemically rational*[2] in holding a certain belief depending on whether the belief fits our evidence, which seems independent of whether we conducted a preceding investigation in a practically responsible manner. Since we *criticize* each other when our beliefs lack evidential support, such evaluations seem to imply that we are more directly responsible for our beliefs than Indirect Voluntarism can allow for.[3]

Indirect Voluntarism feeds on the intuition that we are responsible for states only insofar as we can control them by what we *do*. According to

this idea, a mental state can only be required of a person *indirectly* – that is, by requiring them to *do* something in order to bring about the required state. Importantly for the motivation of the present investigation, this idea is at odds with a widespread assumption in contemporary philosophy: that attitudes like belief and intention can be required *without* requiring us to bring about the attitude by means of an action. According to this idea, it can be required of a person to *just* believe or to *just* intend something. This idea is widespread in contemporary metaethics as well as in epistemology.

In metaethics (broadly construed), an extensive and specialized discussion about the requirements of rationality arose within the last decades. John Broome (2013, 2020) has argued that being rational consists in having coherent attitudes, rather than in responding correctly to (apparent) reasons.[4] According to Broome, we are rationally required to intend those means that we believe are necessary to achieve our ends.[5] This is a requirement to adjust our intentions and beliefs to each other. It is not presented as a requirement to perform certain *actions* in order to intentionally *ensure* by indirect means that we are coherent – as one can do, for example, by engaging in meditation. Rather, rational requirements are meant to govern our attitudes *directly*. This also holds for accounts that conceive of rationality as the capacity to respond correctly to one's reasons. For insofar as these accounts are concerned with responding to reasons for *attitudes*, they do not take themselves to be concerned with prudential or moral reasons for *managing* our attitudes. Rather, rationality often requires that we *believe* in accord with our epistemic reasons and *intend* in accord with our reasons for action.[6]

This point is essential both for the dialectical setup of the following inquiry as well as for the main argument of this book. So let me emphasize it. The current debate takes attitudinal rationality to require of us to *be* rational, rather than to *ensure* one's rationality, or to *manage* one's mind. Rational requirements are not reducible to requirements to perform certain actions – like inquiry, meditation, or controlling one's attention. Indeed, it might be a sign of irrationality if one must first perform such actions to make oneself rational.

For instance, if one is explicitly aware of the fact that one holds contradictory beliefs, the rational response is to drop at least one belief. It is *not* rational to maintain the contradiction while attempting to get rid of it. Suppose that you get convinced by your philosophy professor that facts aren't relative to the person who interprets them. However, you then remember that your history professor had convinced you earlier that historical facts are relative. The rational response is to drop at least one belief.[7] Here you would be irrational if you retain the contradiction, even *while* you attempt to get rid of it. Suppose that your beliefs are recalcitrant: they change only if you go for a walk after the seminar to give your brain a breather. In that case, you would *remain* irrational throughout your walk

as long as you hold two contradictory beliefs while being aware of their inconsistency. Rationality doesn't require you *to go for a walk*. Rather, it requires you to drop your inconsistent beliefs.

In the face of contradictory beliefs, it is rationally required to resolve the contradiction in a non-voluntary, non-effortful manner. If one does not revise one's irrational beliefs, but instead engages in an activity that later results in their revision, then the beliefs remain irrational *until* they are revised by one's action. One does not escape rational criticism by engaging in deliberation, or by going for a walk; rather, one escapes rational criticism only by revising one's attitudes in the way rationality requires. Thus, if we are directly responsible for complying with the requirements of rationality, then Indirect Voluntarism cannot be the correct account for all cases of responsibility for attitudes.

Some have argued that we must allow for *rational delay*: it must be rationally permissible to hold inconsistent attitudes for some time, given that normal agents cannot revise them without going through some reasoning process that takes time (Kiesewetter 2017: 62–70; Podgorski 2017). Note that the argument I make here is compatible with rational delay. My point is that one doesn't escape rational criticism by engaging in voluntary actions that aim at resolving one's irrationality. This point merely implies that we cannot argue for rational delay by appealing to the fact that sometimes agents need to perform actions in order to resolve contradictions, or to acknowledge their evidence. Nevertheless, it's still plausible that agents must go through some non-voluntary process of reasoning to resolve contradictions, which arguably takes time as well. So I can agree that they aren't criticizable as irrational *immediately* after they notice that they hold contradictory beliefs. This simple observation already implies rational delay.[8]

The idea of direct responsibility for attitudes is also central to contemporary debates in epistemology beyond the rationality debate (which I have located primarily in metaethics). William P. Alston (1988) famously attacked the 'deontological conceptions of epistemic justification', according to which deontic concepts like 'ought', 'reason', and 'justified' apply to beliefs in a similar way as they apply to actions. This has provoked elaborated defenses of the idea that we are directly responsible for our beliefs in a similar way as we are responsible for our actions, thus providing a basis for deontic language applying to belief.[9] Furthermore, discussions in the ethics of belief, and especially about epistemic reasons and practical reasons for belief, are concerned with what we ought to *believe*, rather than with how to *manage* our doxastic life.[10] If belief is directly subject to epistemic norms and reasons, and if we're responsible for complying with these norms and reasons, then Indirect Voluntarism cannot be the whole story.

The main worry for Indirect Voluntarism, arising from these debates in contemporary metaethics and epistemology, is that we hold our attitudes *for reasons*. Compare the responsibility that we have for our attitudes with the responsibility that we have for our headaches, the behavior of our dog, or the state of our apartment. I might have reason to take a pain killer, to treat my dog well, or to clean up the place. Yet these are just reasons for actions that result in me not having a headache, my dog behaving well, or my apartment being clean. There aren't any reasons, strictly speaking, for not having a headache, for my dog behaving well, or for my apartment being clean. Plausibly, this is why we're not *directly* responsible for these things, but only *indirectly*, that is, in virtue of our actions. But there clearly *are* reasons for believing in climate change, desiring good food, feeling angry at a reckless driver, and intending to visit a friend. These reasons – say, scientific evidence for climate change, a pleasurable eating experience, the driver's disregard for others, and the friendship itself – aren't reasons for *bringing about* these attitudes by means of certain actions. First and foremost, these are reasons that *directly favor* these attitudes, that *justify* them, or that make it *rational* for us to hold them. We might be rightly criticized or blamed when we hold attitudes that aren't supported by sufficient reasons, or when we fail to hold attitudes that would be well-supported by our reasons.

Yet, if this is right, then how could our responsibility for our reasons-responsive attitudes be *merely indirect*? Instead, if there are genuine normative reasons for attitudes and corresponding requirements to believe, desire, feel and intend, then we must be evaluable in terms of whether we comply with these reasons and these requirements, and so we would expect our responsibility for attitudes to be just as direct as the responsibility that we have for our actions. This idea – that we are directly responsible for our attitudes – has been vividly expressed by Robert M. Adams:

> We ought not only to *try* to have good motives and other good states of mind rather than bad ones; we ought to *have* good ones and not bad ones. On my view the ethics of motives, and more generally of states of mind, has a certain independence, and is not merely a department of the ethics of actions. The subject of ethics is how we ought to live; and that is not reducible to what we ought to do or try to do, and what we ought to cause or produce. It includes just as fundamentally what we should be for and against in our hearts, what and how we ought to love and hate. It matters morally what we are for and what we are against, even if we do not have the power to do much for it or against it, and even if it was not by trying that we came to be for it or against it.
>
> (Adams 1985: 12)

In contrast to the present investigation, Adams didn't draw any tight connection between direct responsibility for attitudes and their rational evaluability. Rather, he appeals to the *moral* significance of our attitudes. As we will see throughout this book, and in particular in Part 3, rational mistakes can give rise to moral blameworthiness. In this sense, I agree with Adams that our attitudes matter for our *interpersonal* affairs. Nevertheless, I disagree with him insofar as I am skeptical about whether attitudes are themselves subject to robust moral requirements.

It seems that we face a classical philosophical puzzle. Intuitively, it is indeed puzzling how we can be held responsible for just *being* in a state, and how *being* in a state can be directly required of a person. However, if we are not directly responsible for our mental states, as Indirect Voluntarism claims, then it seems that we must treat our attitudes as on a par with brute sensations (like headaches) and other mere consequences of our actions (like the behavior of our dogs, or the state of our apartment). We might even have to deny that attitudes are held for reasons or can be evaluated as rational or irrational. For aren't we responsible for everything that is within the 'space of reasons', i.e., that is rationally evaluable? Our intuitive commitment to Indirect Voluntarism seems to lead us into a position that is unacceptable. But denying Indirect Voluntarism requires us to abandon this intuitive commitment. Which way should we go? It seems that we cannot have it both: either we are responsible for everything that can be held for reasons, or we are only responsible for what we control by our actions. For not everything that is held for reasons is controlled through our actions: our attitudes are not actions, and our attitudes often arise spontaneously without a chance for our actions to influence them.

To solve this philosophical puzzle, the present investigation pursues the topics of responsibility and blame as deeply intertwined with issues surrounding reasons and rationality. My strategy is to consider the normative status of reasons for attitudes: how, if at all, are we held responsible for (non-)compliance with our reasons for attitudes? In other words, how, if at all, are we held responsible for (ir)rationality, that is, for (not) complying with *rational requirements*? (I will defend a close connection between rationality and reason in Chapter 3.) I choose this strategy because what makes us pause when considering the plausibility of Indirect Voluntarism is the fact that our attitudes can be held for reasons. If we were not directly responsible for (not) complying with our reasons for attitudes, then the main worry with Indirect Voluntarism would be off the table. We are thus confronted with the challenge of making our direct responsibility for (non-)compliance with reasons for attitudes intelligible. Chapters 2–4 are mainly devoted to setting up this challenge. The remainder of the book then offers a reply. I summarize all chapters in Chapter 1.3 below.

To understand the responsibility that we have for our attitudes, I maintain, we must consider the normativity of reasons for attitudes. Properly understanding this normativity will reveal the responsibility that we have for our rationality. It will reveal that the normativity of rationality is, in a sense to be spelled out, an *evaluative* kind of normativity, which presupposes reasons-responsiveness but not control. I argue that our rationality matters for relationships between responsible beings – like friendships, romantic relationships, or any other relationship that requires attitudes like mutual trust, proper regard, and goodwill.[11] Someone's irrationality can impair these relationships in virtue of providing us with reasons to re-evaluate the irrational person's character, and to then modify our relationship toward them. It can give rise to forms of blame that *mark* relationship impairments (Hieronymi 2004). Such blame does not require that the person who is blamed for their attitude was in control of their attitude. Still, the resulting evaluative normativity implies *genuine* responsibility, despite being 'merely' evaluative. For, as I will argue against current orthodoxy, violating rational requirements can give rise to various forms of blame, and not the least among them being genuine *moral* blame. The dichotomy between 'mere evaluative criticism' and 'genuine blame' is a false one. My view thus implies a link between violating norms of rationality and genuinely moral blaming responses.[12]

I will also argue that the impression that we must decide between voluntarist and rationalist accounts of responsibility – that is, between accounts that ground our attitudinal responsibility in indirect control and those that ground it in reasons-responsiveness – is deceptive. We can instead acknowledge two different *faces* of responsibility for attitudes. Our philosophical puzzlement arises only because we assume that we deal with just one face of responsibility. But there are two. We are *directly answerable to rational requirements* that govern our attitudes, and we are *indirectly accountable to prudential norms and moral duties* that govern our actions of managing our attitudes. Both kinds of responsibility are central to an ethics of mind, and both can be understood from a broadly rationalist perspective, while still respecting the central intuitions that drive Indirect Voluntarism. However, both kinds of responsibility are not as distinct as they might seem – rather, rational answerability can partly ground moral blame. This is my *hybrid account* of responsibility for attitudes that this book defends.

1.2 The Ethics of Mind

This book is an exercise in the ethics of mind. Adams (1985: 12) already spoke of an "ethics of motives, and more generally of states of mind". Yet the field is only developing in the recent years from various debates

in (meta)ethics and epistemology. It will therefore be helpful to characterize this field in order to situate the account I just sketched within current debates. The ethics of mind is concerned, on its most abstract level, with two questions:

1 How can we be responsible for our attitudes?
2 What attitudes should we have?

The first question is concerned with explaining how there can be such a thing as an ethics of mind – how our practice of holding each other responsible for our beliefs, emotions, desires, intentions, and other attitudes can be appropriate. Proponents in this debate either argue that attitudes are, contrary to first impression, directly controlled in the same way as actions are directly controlled ('voluntarily')[13]; or they argue that we can exercise a form of direct, but non-voluntary control over our attitudes, that is, a form of direct control that is different from the control that we exercise when we act[14]; or they deny that attitudes are directly controlled in any way, and thus either deny some sense of 'Ought implies Can',[15] or they say that we are *never* directly responsible for our attitudes, but only indirectly responsible insofar as we control our attitudes through our actions (see references in endnote 2). By providing an explanation of our mental responsibility in this way, such accounts aim at providing a *foundation* for understanding the norms that underly our practice of holding one another responsible for our attitudes.[16]

The second question is central to debates about reasons for belief and about attitudinal rationality. It is concerned with the precise nature of the norms that govern our attitudes. If we can make sense of mental responsibility, then what are the requirements that govern our attitudes, and what are the kinds of reasons for which we can hold attitudes, and how do we distinguish them? Are all reasons for attitudes object-given, or are there also state-given reasons for attitudes?[17] How do these kinds of reasons determine what we ought to believe, desire, feel, intend, etc.?[18] And what is the normative force of these mental 'oughts', which include the requirements of epistemic rationality? Do rational requirements only govern our attitudes, or do they also govern our actions? If rational requirements have normative authority, where does it come from? What is the relationship between rationality and morality when it comes to the norms that govern our mind? How do reasons to manage our attitudes relate to reasons for *having* attitudes – for instance, how do reasons to *manage* belief relate to reasons for *believing*?

This – incomplete – list of questions illustrates how rich our investigation into an ethics of mind can be, but also how difficult it has become to navigate through this much-discussed dialectical terrain, in which contemporary discussions have achieved a high degree of sophistication (I have

barely referenced all the relevant literature in the endnotes). The ambition of the present inquiry into the ethics of mind is to provide an original answer to the first question on responsibility, and thereby to cast light on some issues arising from the second question that is concerned with mental norms, and thus with reasons and rationality (on the relationship between norms and reasons, see Chapter 2.3). In this way, the present investigation aims at motivating new research within the ethics of mind by providing the foundations for pursuing these further issues in normative theory (arising from question (2)) from a fresh perspective – a perspective informed by reflections about the responsibility that we have for being (ir)rational.

Importantly, the ethics of mind also addresses *applied questions*: How should we shape our own mind and the minds of others, and when should we refrain from such shaping? Questions about the relationship between autonomy and morality loom large here (see Meylan and Schmidt 2023). How should we relate to each other in light of the attitudes we hold, especially in concrete contexts that threaten to undermine responsibility, since they threaten one's ability to know and one's agency more generally, such as an epistemically polluted or epistemically hostile environment (Levy 2022; Nguyen 2023), epistemic oppression (Dotson 2014; Toole 2019), or ideological indoctrination (Tobi 2022)? Approaching such questions will benefit from the rich theoretical background that the present book provides. The presented account will thus also be of interest to scholars working in areas like vice epistemology, epistemic injustice, epistemic decolonization, the epistemology of the internet, and other issues in applied epistemology that raise the question of responsibility for our own mind. Yet the concrete applications in these areas remain to be explored.

Throughout the book, and especially in Part 3, I will clarify the relationship between the norms that *guide* our minds and norms that are relevant for *evaluating* each other in light of our attitudes. There, I will also argue that in potentially responsibility-undermining contexts, genuine moral blame can still be appropriate toward those who cause moral harm if they hold attitudes that are unjustified in light of their reasons. This shows how we can explain responsibility in contexts of oppression without going too externalist about blameworthiness: we can acknowledge that agents who lacked a significant amount of voluntary control over their mind are often blameworthy in an intuitive internalist sense of 'blameworthy': the agents were still in possession of the reasons that rationally required them to revise the attitudes. This part should thus be of interest to the discussions about vice epistemology's 'responsibility problem' (Battaly 2019).

However, the purpose of the book is to set out the foundations – as well as some future directions – for an ethics of mind, rather than answering all the specific yet intriguing and currently pressing questions that can be raised within this exciting field of philosophical research. Nevertheless,

the concluding Chapter 8 points to further directions for research in the ethics of mind, including issues in the recently thriving field of applied epistemology.

1.3 Overview

Chapter 2 argues in some more detail for the claim that in order to understand responsibility for attitudes, we must think about reasons and rationality. Traditional statements of the problem of mental responsibility see it as a conflict between responsibility for attitudes and a lack of voluntary control over attitudes: if we lack voluntary control over our attitudes, then we cannot be responsible for them. I argue that this doesn't get at the core of the philosophical puzzle. Instead, the puzzle arises from the fact that attitudes are held for reasons, or are rationally evaluable: even if most of our attitudes are acquired automatically, they're nevertheless often acquired and held *for reasons* in these cases, and so they're evaluable in terms of whether they're based on *sufficient* reasons, i.e., whether they're (substantively) rational. If there are normative requirements that we be rational, then we are directly responsible for complying with these requirements; but it seems that neither voluntary nor non-voluntary forms of control could explain our direct responsibility for rationality. The task of this book is to make sense of this responsibility for rationality, in particular by appealing to our accountability practices.

Chapter 3 then focuses on the relationship between reasons and rationality. I situate the project within the recent debate about the normativity of rationality. In particular, I show how the overall argument of this book allows us to defend a close connection between rationality and reasons. The chapter argues that, if we focus on cases in which we are *responsible* for being (ir)rational, then being irrationally incoherent *implies* that one didn't respond correctly to one's reasons. This is an important link between *structural rationality* and *substantive rationality*: being substantively rational implies that one is structurally rational (in an important sense of 'structurally rational'). The argument I develop presupposes that our capacity to respond to so-called 'right-kind' reasons grounds a kind of direct responsibility for attitudes. The remainder of the book defends this premise by showing that we can be blameworthy for not responding correctly to our right-kind reasons, even when our responses to these reasons aren't under our direct or indirect voluntary control. Thus, the main argument in Chapter 3, together with the main argument of the book, is a contribution to the recent debate on the relationship between rationality and reasons. Overall, the book thereby illustrates how theories of rationality can benefit from thinking more about responsibility.[19]

Part 2 of the book – containing Chapters 4 and 5 – engages in discussions about the normativity of epistemic reasons, that is, the kinds of reasons

that are central to evaluations of *epistemic* rationality. The part defends a view according to which the normativity of epistemic reasons is revealed in our practice of blaming each other for our distinctively epistemic failures. Within epistemology, this constitutes an argument against radical pragmatist and against a kind of instrumentalist account of epistemic norms and reasons. The focus in Part 2 is on responsibility for being *epistemically* rational, while Part 3 will generalize this conclusion to other kinds of rationality.

In contemporary epistemology, the normativity of epistemic rationality can seem puzzling. How can evidence provide us, all by itself, with normative reasons for or against belief? Chapter 4 is devoted to clarifying this challenge for the normativity of epistemic rationality, which comes in two shapes.

The first shape arises from the problem of clutter avoidance: it seems that, if epistemic rationality is normative, then we have reason to clutter our minds with uninteresting implications of our beliefs. Replying to this problem requires the normativist (about evidence and epistemic rationality) to specify the conditions under which evidence provides us with genuine normative reasons for belief. However, this leads them into a dilemma: either the condition fails to explain the normativity of epistemic rationality, or it eliminates epistemic rationality by rendering it practical.

The second shape of the challenge is revealed in cases in which our practical reasons require us to violate norms of epistemic rationality. Epistemologists have recently argued that such epistemic-practical conflicts show that there is no interesting sense in which we always ought to be epistemically rational, or that epistemic rationality is only part of an overall evaluation of what one ought to believe *simpliciter*, or *all reasons considered*. The chapter works out this challenge for the normativist from the recent literature on pragmatism and instrumentalism about reasons for belief, and it reconstructs the underlying argument.

Finally, the chapter ends with a diagnosis: both shapes of the challenge for normativism share a plausible assumption – namely, that there is a conceptual connection between genuine normative reasons and blameworthiness. This is why they are two shapes of the same challenge. Furthermore, the challenge *prima facie* presents strong arguments against the normativity of epistemic rationality. The challenge consists in making the normativity of right-kind reasons intelligible: it is not obvious that we always ought to respond correctly to our right-kind reasons. This is the more generalized version of the challenge, that is, a challenge that is applicable to *all* right-kind reasons for attitudes, instead of being restricted to epistemic reasons for belief. I return to this generalization of the challenge only in Chapter 6.

Chapter 5 defends the normativity of epistemic rationality against the challenge from the last chapter by appealing to our epistemic blaming practices. It first defends the connection between reasons and blameworthiness that the challenge in Chapter 4 rests on. Here I argue that the challenge

is right in assuming that there must be such a thing as epistemic blame if epistemic reasons are normative. Replying to the challenge thus requires me to defend the possibility of epistemic blame, especially in the context of the cases that give rise to the challenge, that is, cases of trivial belief (to reply to the clutter avoidance problem) and cases of epistemic-practical conflict. The chapter then develops such a reply by building on the recent literature on epistemic blame and on rationalist accounts of responsibility for attitudes.

In reply to the first shape of the challenge – the dilemma for the normativist arising from the problem of clutter avoidance – it is argued that a subject can be epistemically blameworthy for being epistemically irrational even in cases where no practical stakes are involved. That is, there are plausible conditions on the normativity of evidence that do not render the requirements of epistemic rationality practical, but that still preserve the normativity of purely evidential considerations. This allows us to meet the clutter avoidance problem by adding non-pragmatic conditions for evidence to be normative, such as the condition that one must *possess* sufficient evidence or *attend* to the question on which the evidence bears: as soon as these conditions are fulfilled, the subject is epistemically blameworthy (absent excuse) for not responding correctly to their epistemic reasons.

In reply to the second challenge – the one arising from epistemic-practical conflicts – it is argued that a subject can be epistemically blameworthy for violating a norm of epistemic rationality even when the subject *ought* to violate the norm of epistemic rationality on the basis of practical reasons. That is, a person can be epistemically blameworthy even when they were practically required to fail epistemically. This is revealed especially in cases where one had decisive practical reasons to cultivate an *epistemic vice*, which makes one blameless for cultivating the vice but doesn't excuse one from manifesting vice in irrational beliefs.

Taken together, both arguments that reply to our challenge from Chapter 4 reveal that epistemic rationality has normative significance independently of practical reasons to make oneself epistemically rational. Chapter 5 thereby shows that we are directly responsible for complying with the requirements of epistemic rationality.

Finally, by building on my previous arguments, Chapter 5 presents a conception of epistemic rationality as an evaluative kind of normativity that matters for how we ought to relate to one another within our epistemic community. The view implies that epistemic reasons cannot be weighed against practical reasons to derive what one *just plain ought to believe* or *ought simpliciter to believe* in cases of epistemic-practical conflicts; instead, there are genuine dilemmas between epistemic and practical reasons. Furthermore, epistemic normativity has pragmatic foundations but epistemic norms are not conditional on practical reasons to comply

with epistemic norms. The resulting view is incompatible with Indirect Voluntarism and certain versions of instrumentalism and pragmatism about reasons for belief, and so my defense provides a novel argument against these views.

At the end of this second part, some might still worry whether epistemic responsibility is 'genuine responsibility'. I leave room for these doubts and address them in the next part, especially in Chapter 7, where I argue that rational failure can even give rise to *moral* blameworthiness. This will give the idea of responsibility for rationality an even more solid basis.

Part 3 of the book clarifies the nature of the direct responsibility that we have for our attitudes. In Chapter 6, the argument from the second part is generalized: our evaluations of attitudes as (ir)rational imply that we are directly responsible for our attitudes, rather than merely for prior actions and omissions. It is argued that this gives rise to a plausible *hybrid account* of responsibility for attitudes that is rationalist in spirit, yet includes indirect voluntarist elements. In Chapter 7, the main worry with the overall approach that responsibility for rationality is not 'genuine' responsibility is addressed. It is argued that there is a continuity between rational criticism and moral blame: violating norms of rationality can sometimes cause moral harm, and if it does cause moral harm, our practice of apology and forgiveness can be fully intelligible, thereby revealing how moral blameworthiness can arise from epistemic irrationality.

Chapter 6 develops my hybrid account of mental responsibility from the previous discussion. I first generalize the challenge from Chapter 4 by applying it to right-kind reasons for desire, intention, and emotion: since there are plausible cases of trivial irrational attitudes and counterproductive rational attitudes, I argue that the doubts about the normativity of epistemic rationality can be generalized to the rationality of other attitudes. I then defend the normativity of right-kind reasons for attitudes by defending the view that we are directly blameworthy for irrationality more generally: holding irrational attitudes can impair our relationships in various ways and warrant distinctive yet genuine blaming responses. Building on this generalized version of my main argument, the chapter then argues that there is a twofold foundation of the ethics of mind. That is, it argues that we should allow for *two faces* of mental responsibility: direct responsibility for responding to right-kind reasons by directly forming or maintaining attitudes, and indirect responsibility for responding to wrong-kind reasons for attitudes by managing our attitudes. Both kinds of responsibility are essential to our blaming practices: any view that attempts to ground blameworthiness for attitudes either merely in reasons-responsiveness or merely in indirect control faces counterexamples. The chapter also defends a kind of internalism or perspectivism about blameworthiness by discussing the infamous case of the "rational racist" (Basu 2019).

The task of Chapter 7 is to reinforce the idea that being irrational can warrant genuine blaming responses – indeed, that it can sometimes legitimately give rise to *moral* blame. It starts out with the observation that many of our attitudes are non-culpable: there was nothing that we should have done to avoid holding them. It argues that we can still be blameworthy for non-culpable attitudes: they can impair our relationships in ways that make our full practice of apology and forgiveness intelligible. The argument has two steps. First, it is argued that we sometimes legitimately blame agents who behave in non-culpable ways, as when someone gets unreasonably angry at us but couldn't avoid their non-intentional, spontaneous behavior. I describe five different cases involving beliefs, desires, emotions, and intentions as causes of the relevant behavior and argue that our practice of apology and forgiveness is fully intelligible in each case, and that the full intelligibility of this practice implies that the agent is blameworthy. In a second step, I argue that the agent's blameworthiness cannot be located merely in the behavior itself, but must instead also be located in the non-culpable attitudes that cause the behavior. I do so by excluding alternative explanations of the agent's blameworthiness. This argument by appeal to our practice of apology and forgiveness poses a new distinctive challenge for Indirect Voluntarism which attempts to reduce all responsibility for attitudes to responsibility for prior actions and omissions. Rationalists (and hybrid views), who instead explain attitudinal responsibility (also) by appeal to reasons-responsiveness, can make sense of blameworthiness for non-culpable attitudes.

The purpose of this final chapter is to break the stalemate between rationalist and voluntarist views of attitudinal responsibility by watering down the contrast between 'mere rational criticism' and 'genuine moral blame', thus revealing a hitherto unnoticed normative force of rationality: evaluations as irrational can have a force just like moral blame. In the overall dialectic of the book, the chapter bolsters my case of blameworthiness for irrationality by showing how rational criticism or blame, including epistemic blame, need not be 'cool', but can plausibly be emotionally laden.

The concluding Chapter 8 summarizes the overall dialectic, points out the implications of the defended view, in particular also for the current debate on the normativity of rationality, and presents some further interesting avenues of research within the ethics of mind.

Notes

1 For defenses, see Meylan (2013, 2017) and Peels (2017) as well as Price (1954) on doxastic responsibility, Oakley (1992) on emotional responsibility, and Jacobs (2001) on responsibility for character. For more general views about

responsibility with this implication, see Rosen (2004), Levy (2007), and Fischer and Tognazzini (2009). See also Clarke (2023) for a recent article that locates direct accountability only in actions and claims that accountability must always be indirect when it comes to attitudes. I have defended a similar view in Schmidt (2020a) but revised it (see Part 3).

2 It is often common in epistemology to treat 'rational' and 'justified' as equivalent (see Wedgwood 2017: 8). As I explain in Chapter 3.1, I follow suit with this usage, at least for the present purposes. I agree, however, that there are theoretical purposes for the sake of which we must distinguish both concepts. In particular, many epistemologists regard epistemic justification as an externalist concept, while epistemic rationality is widely acknowledged to be an internalist concept. See my brief discussion of excuses in epistemology in Chapter 3.5.

3 See Owens (2000, 2017a: intro.), Smith (2005), Graham (2014), Hieronymi (2006, 2008, 2014, ms), McHugh (2013aa, 2014; 2017), McCormick (2015), Roberts (2015), Portmore (2019), White (2019), Schmidt (2020a, 2022), Ayars (2021), and Osborne (2021). Adams (1985) and Montmarquet (1993) are earlier opponents of Indirect Voluntarism. These accounts have their predecessors in accounts of responsibility which argue that our responsibility originates in our character – our 'quality of will' or our 'real self' (Wolf 1994). Arguably, our responsibility can originate in our character only if we were non-derivatively responsible for our attitudes, which (partly) constitute our character. See Luvisotto (2022) for a recent contribution that defends responsibility for belief as arising out of a person's character, or their 'evaluative orientation', but that explicitly disconnects this responsibility from reasons-responsiveness and from norms more generally (in contrast with the authors quoted above in this endnote and in contrast with the present investigation).

4 The account of rationality as responding correctly to one's apparent reasons – that is, the reasons one would have if one's beliefs were true – is associated with Parfit (2001, 2011), and it was further spelled out by Schroeder (2009) (for instrumental rationality) and Way (2009, 2010).

5 "Rationality requires of us that, if we intend an end and believe some means is necessary to that end, then we intend the means. This has to be a requirement of rationality. A person is necessarily irrational if she does not intend whatever she believes is a necessary means to an end she intends" (Broome 2005: 2).

6 See Kiesewetter (2017) and Lord (2018), who argue that no formulation of the coherentist requirements of rationality can make it intelligible why we ought to be rational or why we have reason to be rational. See also Kolodny (2005), who motivated these views. These authors instead argue that rationality consists in responding correctly to the reasons 'within our ken', that is, the reason-providing facts that are epistemically available to us or possessed by us. Note that these accounts are different from the accounts of rationality as responding correctly to apparent reasons mentioned in endnote 5. See Worsnip (2021) for a recent reply to the challenges for coherentism about rationality that were raised by these authors.

7 I take this case from Eva Schmidt (2023), who argues that the inconsistency here is itself a reason to suspend judgment about both propositions. I object in Schmidt (2023) that merely the balance of evidential reasons determines the rationality of suspension: the fact that your evidence is tied provides you with a decisive reason to suspend. See Eva Schmidt (2024) for a reply.

8 Rational delay must be distinguished from the view that rational requirements govern mental processes of reasoning rather than mental states (see Kiesewetter

2017: 70–79). Following Kiesewetter, I take it that states rather than processes are directly governed by norms of rationality. Indirect Voluntarists would have trouble also with responsibility for complying with such process requirements, since the relevant processes (forming, maintaining, and revising attitudes) aren't always voluntary either. So my argument against Indirect Voluntarism doesn't depend on whether rationality governs states or processes. What matters is that it governs non-voluntary responses.

9 See Boyle (2011), McCormick (2015), and Hieronymi (2008) for some defenses.

10 At least this is how many proponents of the debate wish to understand their own views, e.g., Adler (2002), Kelly (2002), Owens (2000), Wedgwood (2002). The discussion partly originates in the papers of William K. Clifford (1877) and William James (1896), who did not explicitly distinguish between reasons for belief and reasons for belief-management (see Lindner 2020 for a helpful discussion). However, especially in recent years, pragmatists like Leary (2017), McCormick (2015), Reisner (2009), and Rinard (2015, 2017) have questioned the strict distinction between norms of belief and norms of belief-management. However, see Chapter 3.2, Chapter 5, as well as Schmidt (2022) and Schmidt (forthcoming c) where I provide some motivation to maintain this distinction.

11 These are the attitudes Strawson (1962) viewed as essential for the kinds of relationships we can enter with other responsible beings.

12 As far as I'm aware, the only explicit defense of the view that not responding correctly to epistemic reasons or other right-kind reasons for attitudes can give rise to genuine moral blameworthiness in the recent literature is Ayars (2021) – of course next to Schmidt (2024b) and chapter 7 in this book. However, the issue creeps up regularly in debates on tracing and culpable ignorance. The view is normally dismissed – too quickly, as I claim.

13 See Ginet (2001) for the classical defense of doxastic voluntarism. However, the position I appeal to here need not rely on an ability to explicitly decide to believe. McCormick (2015) argues that both belief and action are subject to 'guidance control' (see Fischer and Ravizza 1998), which she distinguishes from control by decision, and she argues that guidance control can ground doxastic responsibility.

14 See Boyle (2011), Hieronymi (2006, 2008, 2009a; 2009b, 2014, ms), Montmarquet (1993), Raz (2011), Smith (2005).

15 See especially Adams (1985) and Owens (2000, 2017a: intro.). Owens only denies "Ought implies Can" for beliefs and emotions, but not for intentions. According to Owens, we are responsible for our intentions in virtue of having reflective control over them – a control we also exercise over actions, but which is not voluntary control.

16 This is the idea of Meylan (2013), titled *Foundations of an Ethics of Belief*. There are positions that do not neatly fit into one of the categories I present here. First, Graham (2014) thinks that we are never directly responsible for our actions, but only for our attitudes. He does not explicitly ground this direct responsibility in control, and thus most likely belongs to the same category as Adams and Owens (see endnote 16). Second, Gaultier (2020) argues that exercising what he calls 'doxastic strength' – resisting one's conative attitudes that pull one to believe against the evidence – is a form of indirect control that is not voluntary control. However, doxastic strength seems effortful, and so I count his view among the indirect voluntarist views mentioned in endnote 2.

17 Traditionally, the distinction between object-given and state-given reasons is seen as matching the distinction between reasons 'of the right kind' and reasons

'of the wrong kind' (see Gertken and Kiesewetter 2017 for more on this distinction). It has been argued that some state-given reasons are right-kind reasons because they bear on the rationality of the attitude at issue (Schroeder 2012, 2021; Lord 2020). I return to this issue in Section 3.2.

18 For proposals of how to weigh or compare practical reasons for belief (state-given or wrong-kind) with epistemic reasons for belief (object-given or right-kind), see Reisner (2008, forthcoming), Steglich-Petersen and Skipper (2019), Howard (2020), and Meylan (2020). McCormick (2017) proposes how to weigh different kinds of reasons for hope with one another. For doubts about such weighing, see Berker (2018), Kauppinen (2023), Schmidt (forthcoming c) as well as Chapter 5 of this book.

19 I have urged theorists of rationality to think more about responsibility in Schmidt (2020b). There I could only do so on very limited space. I thus take Chapter 3 as an opportunity to spell out these ideas.

2 Mental Responsibility

> Attributions of responsibility and other deontological judgments in the doxastic realm are puzzling. For much of what we believe is beyond our control; we cannot decide to believe the way we can decide to act. It seems that such lack of control should excuse us from responsibility and judgment.
>
> Miriam Schleifer McCormick, *Believing Against the Evidence* (2015: 77)

According to McCormick's characterization, the problem of mental responsibility is not restricted to responsibility for beliefs. It also seems that *deontic judgments* about beliefs are puzzling in the face of our lack of doxastic control. After all, in some sense, 'Ought' implies 'Can'. But how can sentences like 'You should not believe this nonsense' or 'There is no reason at all to believe this' be true if there is no doxastic control, that is, no 'Can'? The very possibility of an ethics of belief, as well as the project of normative epistemology, seems to become questionable (see Alston 1988).

Are we dealing with one problem here, or with two? Is there a problem about responsibility *and* a problem about deontic attribution? I argue that the problem of mental responsibility is best understood as the problem about reasons and rationality. This will dissolve the impression that there might be two different problems: there's only one. It will also allow us to spell out the problem with Indirect Voluntarism: attitudes can be held for reasons, which implies *direct* responsibility. Furthermore, candidate solutions will become visible. Overall, this chapter will give us a clearer grasp of the central theme of this book: the problem of mental responsibility as a problem about rationality and reasons. The next chapter then dives deeper into the relationship between reasons and rationality. Together, both chapters provide the framework for our inquiry.

Chapters 2.1 and 2.2 develop a natural line of thought that leads us to the core of our philosophical problem. Chapters 2.3 and 2.4 then provide some clarifications around the concepts of norms, requirements, responsibility,

DOI: 10.4324/9781003382973-3

and blame. Chapters 2.5 and 2.6 round up the dialectic by applying this conceptual framework to our main problem.

2.1 Control and Direct Responsibility

Let us consider, first, why it seems that we cannot *control* our beliefs. Most of our beliefs seem to be passively caused by our environment. We experience the acquisition of a perceptual belief – like the belief that you are reading this text right now – as something that *happens* to us, rather than something we decide for. You do not stop and reflect upon whether you should acquire this belief after the text is in front of your eyes. Rather, belief comes immediately with perception. The same holds in cases where our perception provides us with ambiguous evidence. In these cases, we *refrain* from judgment quite automatically. To take a classical example (cf. Sextus Empiricus, *Outlines of Pyrrhonism*, I.32): If a tower seems to be round from a distance and I know that I am not close enough to judge the tower's shape, I cannot just decide to believe that the tower is round. Rather, I refrain from judging quite automatically. If, however, I am close enough to the tower, I will come to believe what shape it really is without having to perform any special action (except for walking toward it).

Even when we actively reflect about what to believe, there seems to be no place for genuinely *doxastic* freedom. Imagine a scientist who wants to find out whether a certain substance is water. She conducts some experiments, reflects about them, and comes to believe that it is water. The only actions the scientist performed were the experiments and her intentional deliberating or reasoning about what is true. She did not perform a *further* action when she was done with her active thinking: she did not actively *form a belief*. Rather, her thinking concluded in the formation of a belief, but this conclusion is not an additional action that she performed.

To illustrate this intuitive lack of doxastic freedom or control, imagine a manual telling the reader what to do in order to acquire justified, true beliefs about a certain topic. Let us assume that just *reading* this manual is not enough to justify certain beliefs. For instance, suppose that the author of the manual does not provide the information necessary to justify these beliefs because it is important for her that her readers come to hold these beliefs due to their own efforts, rather than just because she tells them that this is so and so. These manuals would mention various activities the readers should perform, like reading certain books, investigating certain topics, conducting experiments, or thinking about certain issues. The manuals would not tell the reader what to do *after* they are done with all those activities. They would not tell them *how to form a belief* after they are finished with their inquiry, consisting of careful reasoning, experimenting,

observing, etc. This is because there is nothing left for the reader to *do* after they tried to find out the truth by inquiry. Any further advice beyond how to conduct the relevant inquiry will not make any sense. The reader would just not know what to do with this advice. Suppose the manual says: 'First, read a book about birds of the kind xy. Then watch bird x and think about whether this bird has property E. Finally, form a belief about whether bird x has property E.' If the last piece of advice is not meant as urging us to inquire even more into the issue, it just would not make any sense as an advice.

Such observations about the automaticity of belief formation are the flipside of the *reasons for which* we form beliefs. Return to our scientist from above, who could engage in the experiments she conducts and in her activity of thinking for *practical reasons*: she could decide to engage in them or refrain from doing them depending on whether it was *important* or *of interest* to know the nature of the substance. By contrast, she could not decide to conclude her investigation for such reasons: considerations about whether it would be good, valuable, important, interesting, or useful to believe that the substance in question is water could not guide the formation of her belief after she is done with her previous activities. Rather, she will form the belief based on her evidence that was uncovered by the preceding investigation. This confirms that forming beliefs as immediate responses to evidence or epistemic reasons isn't under our control in the way performing actions as immediate responses to practical reasons is under our control.[1]

Thus, there is a dilemma for the proponent of doxastic control: Either we form our beliefs spontaneously, or we form them reflectively. Spontaneous belief formation seems to happen quite automatically on the basis of evidence, and thus, there seems to be no room for genuine freedom. However, even when we form our beliefs reflectively, only our reflection, that is, our intentional thinking, is active and done for practical reasons, but not the formation of our belief that results from it. As Richard Moran (2001: 119) points out, "there is no further thing the person *does* in order to acquire the relevant belief once his reasoning has led him to it". Our rational capacities just work the way they do – we do not exercise direct control over the results of their proper functioning, even if we can influence their proper functioning through various activities (such as getting enough sleep). It seems that it is not *us*, but rather our evidential situation, together with the functioning of our rational capacities, that determines what we believe in each situation – in which we might happen to be either with or without our contribution (Strawson 2003).

This is a dilemma not only for the proponent of direct *doxastic* control but for the proponent of direct control *over attitudes* in general. Most of our attitude-forming takes place without us being reflectively aware of it.

We spontaneously form not only numerous *beliefs* about our environment but also other attitudes. We form desires and intentions to make it in time to the meeting, hopes that we will still make it, fears that we won't, regrets that we did not get up earlier to make it in time, and feelings of anger and resentment directed at the slow driver in front of us. Even if we *could* decide on our attitudes, we would not have time do so. And even if we *have* time to form our desires, intentions, and emotions reflectively, then we can at best inquire into what is good, right, fearsome, regrettable, or deserving of our anger and blame. Yet the attitudes that arise from such inquiry are nothing we choose. Even if we actively think about how to decide between actions, it seems that our resulting decision or intention is itself nothing that we could ever directly decide on (see Kavka 1983).[2]

The proponent of direct control might object that, until now, I have only considered cases of *rational* belief formation. Sometimes, our desires, passions, emotions, and moods influence what we believe in a situation. In the words of William James (1896: 9), our 'willing nature', that is, "all such factors of belief as fear and hope, prejudice and passion, imitation and partisanship, the circumpressure of our caste and set", influence, and often determine, what we in fact believe. Will this not make room for genuinely *doxastic* freedom?

It does not seem so. For, as James himself notes, even in these cases "we find ourselves believing, we hardly know how or why" (1896: 9). Here, James means to strengthen his point about the influence of our 'willing nature' on belief. Interestingly, his discussion of how the will influences our beliefs here isn't meant to support the idea that we have voluntary control over our beliefs but rather to further *undermine* this idea. Again, we might take certain measures in order to ensure that our passions have this or that influence on what we believe. But this is indirect control over belief, not *direct* control. Manuals which would tell us how to deceive ourselves into believing that p by instrumentalizing our 'willing nature' might, for example, advise us to deceive ourselves by attending to this and that, and by avoiding attending to other things. Remember Pascal's advice to surround yourself with religious people and engage in the religious live in order to cultivate belief in God. But there cannot be any meaningful advice that we should, after we are finished with these activities, *form the belief that p*. Intuitively, this response is just settled automatically, whether by evidence or by non-evidential reasons and other factors that influence our states of belief.

The direct control we normally have over our actions is commonly called 'voluntary control', that is, control by the will, or by our intentions or decisions. To be more precise, we should say that our actions are themselves *exercises of voluntary control*. Some authors prefer to say that our actions are *under* our voluntary control, or that we exercise voluntary

control *over* our actions. This is misleading, however. It suggests that there are separate 'exercises of voluntary control' which are different from our actions (maybe some mental activity), and that we control our actions by means of these further exercises of voluntary control. This misleading picture suggests that we are homunculi who control our own body from within our own head, and so gives us a weirdly estranged picture of our agency. While I will grant that there might be a sense in which we exercise control when we intend or decide to act, intention or decision are not themselves exercises of voluntary control. The whole point of the concept 'voluntary control' is to highlight a kind of freedom which we enjoy when we act for reasons.[3]

The *prima facie* lack of direct voluntary control over attitudes can be further illustrated by what Matthew Chrisman (2008: 346) has called the 'no rewards principle': In a situation where you are offered a reward for believing something for which you have no, or contrary, evidence, you cannot believe it 'just like that'. Suppose someone offers you a reward for believing that the number of stars in the galaxy is even. Arguably, nothing hinges on whether you believe something about this, except that you will get a lot of money if you acquire the belief. So why not just believe it to get the money? It seems to be in some sense *impossible* for you to acquire the belief *just like that*. You might deceive yourself into believing that you have evidence for the desired belief, or you might swallow a futuristic pill that induces beliefs, or you might try to convince the people offering you the money that you have the relevant belief, and act as if you believe it (e.g., you might reply with 'yes' when you are asked whether the number of stars is even, even though you do not believe it). However, in the latter case, you do not really believe it, and in the former cases, you bring yourself to believe it by more indirect strategies. It is difficult to see how we could just directly form the belief that the number of stars is even. So it seems that beliefs are not exercises of direct voluntary control.

We might conclude that we never have direct voluntary control over our beliefs, that we can never 'believe at will'. For having such control over belief would require that we can believe *for* reasons of the kind provided by the reward in situations like these.[4] However, if we do not have such voluntary control over belief, how can beliefs be subject to norms? After all, we (often) can perform an action just because we see that the action would be good to perform. When someone offers me a reward for lifting my arm, it is quite easy for me to do so and collect the reward. Arguably, the fact that we can control actions in this way ('voluntarily') is what explains (at least partly) why we sometimes *ought* to do one thing rather than another, why there are *reasons* for actions, and why we can be *blamed* and *praised* for what we do.

Forming an emotion or an intention just because it would be good to have the emotion or intention seems at least as problematic as forming a belief for such reasons.[5] This *prima facie* problem can be experienced by each of us if we imagine situations analogous to the one described above: if someone offers you a reward for desiring something completely undesirable (either something completely neutral or something very bad), or for merely *intending* an action which you have no reason to perform,[6] or for being angry about someone who did not do anything bad to anyone, or for fearing something which you do not consider dangerous or fearsome, then you cannot desire, intend, or feel 'just like that'.

I will not go into detailed descriptions of examples for every attitude, for the examples are easy to imagine, and they have been described in detail by others (see endnote 5 above). Note only that, *prima facie*, it does not seem to be the case that we have direct voluntary control over mental states. Thus, given that this kind of control seems to explain why we can be normatively required to act, the question arises of how mental states can *ever* be subject to normative requirements.

Why, one may wonder here, is this all about *direct* voluntary control? Why is the fact that we have indirect control over our beliefs and other mental states through actions not sufficient to explain mental responsibility? We must answer this question in order to get to the core of the problem at hand, and to see which *kind* of responsibility we are seeking to understand by invoking control. For the problem is not that we do not control our attitudes at all: we are sometimes responsible for them in the way we are responsible for other consequences of our actions. After all, nobody denies that we can sometimes control our mind indirectly.[7]

The main reason why indirect control cannot fully explain why we are responsible for our attitudes is that attitudes are *held for reasons*. This distinguishes them from non-intentional mental phenomena, like sensations. The question 'Why do you feel pain?' can only be answered by giving a causal explanation ('I fell from the roof'). It cannot be answered by giving *justifying reasons* why it would be *appropriate* or *rational* or *make sense* to feel pain now. For pain, understood as a mere sensation which can occur in various bodily parts, cannot be rational or irrational. By contrast, 'Why are you angry?' cannot only be answered by a mere causal explanation ('I talked to my neighbor, and this talk brought about certain brain states of mine which caused me to be angry'), but also by giving *justifying* reasons ('My neighbor said a lot of offensive stuff'). It can be *rational* or *irrational* to be angry in a situation, and one's anger can be more or less intelligible. By contrast, 'I understand your pain' will only make sense if 'pain' refers to a complex emotion, rather than to a mere sensation: besides physiological processes, there is nothing more to understand about a hurt foot.

In contrast to beliefs and other attitudes, sensations are *brute*: they do not belong to the space of reasons.

The crucial point is that we have *indirect control* over our sensations and yet our sensations are not rationally evaluable. We can influence whether we have pain or how much pain we will have by seeking or avoiding dangerous situations, or by going or not going to the doctor. We can cause ourselves to feel pain by tweaking ourselves. This kind of control makes us indirectly responsible for our sensations. But this is obviously *not sufficient* for pain to be subject to rational evaluation. The rational evaluability of attitudes indicates that our agentive relation to our own attitudes is much more direct than our relation to our sensations: we can be irrational merely by believing, feeling, desiring, and intending something (see Chapter 1.1). As Angela Smith (2005: 251) puts it: "we are not merely producers of our attitudes, or even guardians over them; we are, first and foremost, inhabiters of them". That is, our agentive relation toward our intentional mental states, and consequently our *responsibility* for them, seems to be much more direct than mere indirect control could explain.

This might lead us to think that we must appeal to some kind of *direct* control to explain our *direct* responsibility for attitudes. However, it is not obvious what direct control over attitudes is supposed to be. *Prima facie*, attitudes are not actions, and actions seem to be our only paradigm of what we control directly. If the voluntary control that we exercise in acting is the only control that can legitimately be called *direct*, then there seems to be no direct control over attitudes. But then it seems that neither appealing to indirect voluntary control nor appealing to direct voluntary control could explain why we are directly responsible for (ir)rational attitudes.

2.2 Reinterpreting Our Problem

How could a lack of control over attitudes pose a problem for our practice of holding each other responsible for our attitudes? It is helpful to state the problem as a conflict of three claims which, when all true, would result in a contradiction. Thus, to resolve the conflict, we have to reject at least one of the three claims.[8] A first attempt would be to state the problem as follows:

1* We are responsible for our attitudes.
2* We are responsible for our attitudes only if we can control them.
3* We cannot control our attitudes.

The problem with this way of formulating the puzzle is that (3*) would be obviously false. As mentioned earlier, we have at least *indirect* control over our mental states: we can meditate, investigate, and actively engage in thought and reasoning. We can also control our mind indirectly by

engaging in projects with the aim of acquiring, for example, true beliefs about, say, the city of Bucharest by going there and walking through the city. So, this attempt at formulating the problem does not get at its core.

Given that it is not any control that seems lacking, but the kind of *direct* control we have over our actions, we could put (3*) not in terms of control, but rather in terms of *direct voluntary* control. Still, this would not get to the core of the problem. There are two reasons for this.

First, it would ignore the possibility that we could be criticizable or even blameworthy for failures of our rationality. To see this, note first that if we had direct voluntary control over beliefs, we could choose our beliefs based on whether they are beneficial to ourselves or to others. We would be criticizable or blameworthy based on whether we violate the norms of prudence or morality in choosing our beliefs. It might then happen that a particular belief – say, an overoptimistic belief that you are the best pianist of all times – is all things considered best to hold. Suppose, for instance, that this belief would boost your confidence for the next pianist competition. This belief would be epistemically irrational, given your evidence about your pianist abilities. Assume that you have voluntarily chosen this irrational belief due to its practical benefits. In this case, we could not regard you as criticizable for your epistemic irrationality *in virtue of* your exercising direct voluntary control in choosing this belief. For you have exercised direct voluntary control in the way you *ought* to have exercised it: you brought about the belief that is all things considered best to hold. Yet you might still be blameworthy in a distinctively *epistemic* way by virtue of your epistemic irrationality.

So, even if we had direct voluntary control over attitudes, it seems that blameworthiness for rational failures cannot be explained by it. We should thus not formulate the problem in terms of voluntary control, if we do not want to just assume from the start that we aren't *epistemically* responsible for our beliefs. One might deny that we are responsible for our beliefs in a distinctively epistemic way. But this is a position that has to be argued for. Furthermore, it is a position that fails, as I will argue in Chapters 5–7.

Second, even if there were some cases in which we *could* exercise direct voluntary control over an attitude, we surely cannot *always* exercise direct voluntary control over our attitudes. Spontaneously arising attitudes seem to be outside the reach of any kind of voluntary control whatsoever. Yet such attitudes are often rationally evaluable. If this rational evaluability implies that we are in some sense responsible for holding attitudes, then the responsibility for such attitudes cannot be explained by reference to voluntary control. So, even if we had direct voluntary control over attitudes in some cases, other cases in which we lack such control, but in which we still seem to be responsible (due to the possibility of rational evaluation), remain puzzling. The problem can thus not be a problem about voluntary control.

Again, one reply is to just deny that there is such a thing as blameworthiness for purely rational failures: we never blame people merely for being epistemically irrational; rather, we blame them for irrationality only if their irrationality was under their direct or indirect voluntary control. I will discuss this view at length throughout the book. However, we should not assume this position right away when formulating the problem. Rather, we should formulate the problem in such a way that *all* candidate solutions – including those that *allow* for direct responsibility for rationality – can be put on the table. Formulating the problem in a way that presupposes that all responsibility must be explained by voluntary control just frames the discussion in a way that makes us oblivious to other interesting candidate solutions, and it makes us overlook important connections between responsibility and rationality.

Since the problem of mental responsibility also calls for an explanation of the kind of responsibility we have for rational failures, and since it seems that this responsibility can be made intelligible neither by direct nor by indirect voluntary control, we should now try the following formulation:

1** We are responsible for being (ir)rational.
2** We can only be responsible for being (ir)rational if we have *direct non-voluntary* control over our attitudes.
3** We do not have such direct non-voluntary control.

I think this is a better way to state the problem. Yet we might wonder why (2**) is true as stated. For what is *non-voluntary control* supposed to be? Intuitively, the control we exercise when we act – direct voluntary control – is our only paradigm of direct control, and this is the kind of control that explains why we are responsible for our actions and their consequences. So how could *non-voluntary* control explain our responsibility for being (ir)rational? Since we do not know much about the nature of this control, we also do not see, *prima facie*, how it could do this trick. It thus seems that, intuitively, there is *no* form of control – neither direct nor indirect, neither voluntary nor non-voluntary – that could explain how we are responsible for being (ir)rational. I therefore propose the following formulation of the problem of mental responsibility:

The Problem of Mental Responsibility, First Formulation

1 We are responsible for being (ir)rational.
2 We can only be responsible for being (ir)rational if there is *some* form of control that could explain why we are responsible for being (ir)rational.
3 There is no form of control that could explain why we are responsible for being (ir)rational.

The plausibility of (1) derives from the intuition that being irrational is blameworthy or criticizable in some sense. This assumption is widespread in the literature on rationality. It is often a starting point for motivating the idea that *rationality is normative* – that is, the idea that rationality provides us with normative reasons or places genuinely normative requirements on us to be rational. Derek Parfit, for instance, points out that he uses 'irrational' "in its ordinary sense, to mean, roughly, 'deserves strong criticism of the kind that we also express with words like 'foolish', 'stupid', and 'crazy'" (Parfit 2011: 123), and Kiesewetter (2017: chapter 2) points out that we use 'irrational' as *personal* criticism (see also Way 2009: 1; Lord 2018: 4). That is, we use this word (sometimes) in order to criticize another person for a response. Such *personal* criticism contrasts with merely evaluating a response as bad and with merely criticizing the person's rational subsystem for malfunctioning (say, in cases of severe pathology).[9]

Thus, proponents in the current debate on rationality take irrationality to be criticizable. They thereby commit to the view that we are responsible for being (ir)rational. For otherwise being criticizable for violating norms of rationality wouldn't be intelligible. This also supports the view that rational requirements have a certain authority that other norms – say, those of etiquette – lack. I will return to the idea that this authority can be captured in terms of responsibility and blame or criticism – a central theme of this book – in Chapters 2.4 and 2.5 below.

The plausibility of (2) derives from the intuition that responsibility requires control, or 'Ought implies Can' – a claim that is also, in one version or another, accepted by philosophers concerned with the requirements of rationality (cf. Kiesewetter 2017: 28; Wedgwood 2017: chapter 3). Here the relevant 'can' is not understood as a kind of voluntary control, but as exercising a specific capacity to directly form attitudes in response to reasons for those attitudes.

I have also motivated the idea that there seems to be no such thing as direct control over our attitudes (in Chapter 2.1), and that our familiar paradigm of control – voluntary control (direct or indirect) – cannot do the job of explaining why there is such a thing as responsibility for complying with rational requirements (in Chapter 2.2). This, in turn, motivates (3).

Thus, all the premises are intuitively plausible. In order to resolve our philosophical perturbation caused by the three claims, we need to find reasons for rejecting at least one of them. There are two further advantages of understanding the problem of mental responsibility in the way I have proposed here.

First, we now know better *what it is* that requires an explanation. Merely saying that we need to explain how we can be responsible for our attitudes leaves the explanandum unclear. For it is obvious that we can be said to be sometimes responsible for our attitudes in the same way as we can be said to be sometimes responsible for the consequences of

our actions (Meylan 2017). We are indirectly responsible for our attitudes insofar as we have *indirect control* over our attitudes: we can actively reason to beliefs or inquire, manage our emotions by meditation, and determine our intentions and decisions by actively thinking about what to do. If we understand the problem as requiring an explanation of our *responsibility* for attitudes without specifying what *kind* of responsibility we have in mind, then it is unclear why indirect voluntary control does not provide a satisfying explanation. Yet, as I have argued above, our capacity to indirectly control our mind doesn't seem suited to explain why we are responsible for being (ir)rational. Indirect responsibility for attitudes thus seems not to *exhaust* our responsibility for them. Since we do not see what kind of control could explain the direct responsibility for complying with the requirements of rationality, we are faced with a philosophical problem.

Second, stating the problem in this way has a *metaphilosophical* advantage. It shows us how different debates are relevant for solving the problem. For one, we need to think about the nature of responsibility and control. However, we also need to think about *the nature and the status of requirements to be rational*. Thinking about their status means thinking about how we are responsible, and how we are sometimes blameworthy, for forming, maintaining, and revising our attitudes in response to reasons for and against these attitudes. Getting clear about responsibility for rationality might also give us a clue as to what kind of control we are searching for, or show us that this search is futile, because the relevant responsibility might not require control at all. Thus, theorists of responsibility need to think about rationality and reasons.

On the other side, theorists of rationality (and reasons) also need to think about responsibility and blame (Schmidt 2020b). For as long as we do not understand in what sense we are responsible for complying with rational requirements and thus in what sense we can be legitimately blamed or criticized for irrationality, we do not fully understand the normative status of these requirements. To better understand this status, we need to think more about norms, rationality, and their connections to responsibility and blame.

I will begin to discuss some basic connections between these concepts throughout the rest of this chapter as well as in Chapter 3. I start with norms and reasons (Chapter 2.3) and continue with their relationship to responsibility and blame (Chapter 2.4). This will allow us to set the stage for the remaining inquiry, and to relate our discussion to other debates within the field of the ethics of mind (see Chapter 1.2). I will again return to our main philosophical problem that I have just outlined after these conceptual clarifications (that is, in Chapter 2.5).

2.3 Norms and Reasons

There are descriptive uses of the term 'norm' that are not of primary interest for our investigation. Sometimes we say things like 'It is the norm to be married by the age of 30', thereby referring to some regularity within a society. Any such purely descriptive use of the term does not imply anything about what one normatively *ought* to do. Such a sentence might also express the beliefs of people within the society about when one ought to be married. But the beliefs of these people will not have any straightforward implications for what individuals normatively ought to do. Such norms are just conventional, similar to the A4-norm for paper sizes, as well as for sizes of other artifacts, like cars, bottles, chairs, windows, etc. Yet there are also non-conventional norms. John Broome, while distinguishing types of 'ought's or 'should's,[10] mentions the natural norm that 'an oak should have deep, sturdy roots', and writes:

> An oak's having deep, sturdy roots could fairly be called 'a norm', and that is enough to justify the term 'normativity'. However, in the context of an oak, natural normativity is not what I call true normativity. To say an oak should, in the sense of natural normativity, have deep, sturdy roots *is only to say it needs deep, sturdy roots to complete its life cycle*. This is not a truly normative statement.
>
> (Broome 2013: 12)

This norm is a standard of *normal development* of an oak. It is how we expect oaks to develop if they grow in the right environment. The norm is non-conventional because it is set by the natural development of an organism, rather than human conventions. But it's not normative.

We get closer to 'genuinely normative' norms – i.e., *requirements* – if we consider the relationship between norms and reasons. Norms can be expressed with 'ought's or 'should's as in 'one ought not to steal' (which is equivalent to 'one is not allowed to steal') or 'an oak should have deep, sturdy roots'. As we just saw, not all norms provide us with reasons (see also Kiesewetter 2017: 3–4).[11] Just because it is true, in some sense, that oaks should have deep, sturdy roots, the truth of this statement doesn't imply that anyone has a reason to, say, plant well-developing oaks or maintain the roots of oaks. Similarly, just because the rules of a game require one to make a certain move does not imply that one has a reason to make the move (maybe there is no reason to play the game). Or, to give one last example, just because the rules of etiquette require one to use a certain fork for a specific meal does not imply that one always has a reason to use that fork for that meal (sometimes we do not care about etiquette,

and sometimes nobody will be offended by our non-compliance with etiquette).[12] Those norms that provide us with reasons – requirements of morality and prudence, and (as I will argue) of rationality – are often called *normative* requirements. We can, in line with the metaethical literature on reasons (see, for instance, Kiesewetter 2017: 8–9), define the following conceptual relations between normative requirements and reasons:

Decisive reasons. S has decisive reason to φ *iff* S (normatively) ought to φ.
Sufficient-allowed-link. If S has sufficient reason to φ, then S is (normatively) allowed to φ.
Sufficient reasons. S has sufficient reason to φ *iff* S (normatively) ought to [φ or give any other response that is allowed].

Decisive reasons are not only to be contrasted with sufficient reasons but also with *pro tanto* reasons. A *pro tanto* reason to φ neither needs to be sufficient nor decisive. Rather, it can sometimes be easily outweighed by other reasons even though it favors φing.[13] For example, the fact that jumping on the neighbor's flowers would be lots of fun is a reason to jump on the neighbor's flowers. Yet this reason is neither decisive for jumping on the flowers – I am not required to jump on them. Nor do I have sufficient reason to do so: it would not be allowed. Other reasons, like the fact that it would upset our neighbor, that the flowers are beautiful, and that I might get sued, count against jumping on the flowers. And yet the fact that jumping on the flowers is fun favors jumping on the flowers and is thus a *pro tanto* reason to jump on them.

The distinction between decisive or sufficient reasons and *pro tanto* reasons holds for reasons for attitudes as well. That there is a car driving toward my house can be – given my overall background information – a reason *to believe* that my expected guest is arriving. However, if it is possible that another person is driving toward my house, then this reason does not make it the case that I *ought* to believe that my guest is arriving. Nor would it provide me with *sufficient* epistemic reasons for believing that my guest is arriving. I might be epistemically at fault if I gave in to wishful thinking and believed that my guest was arriving just because there was a car driving toward my house. The car might just turn around and drive back, so that it turns out that it was not my expected guest who was driving it, but just someone who made a wrong turn. The fact that the car is driving toward my house is, under these circumstances, merely a *pro tanto* reason to believe that my guest is arriving: it doesn't meet the threshold of permissible belief. By contrast, if the car does not turn around and I instead see someone waving to me from the car, then I would have, *ceteris paribus*, sufficient or even decisive reasons to believe that my guest is arriving. Similar examples can be construed with respect to reasons for other attitudes.[14]

Having decisive reasons to φ means that one ought to φ *all-things-considered*. That is, given all the reasons that are relevant for whether to φ, φing is the thing to do (or the attitude to have). Some authors call this the 'deliberative ought' (Williams 1965: 184; Kiesewetter 2017: 9) or the 'central ought' (Broome 2013: 22–25).[15] This 'ought' of decisive reasons is to be distinguished not only from 'oughts' that do not provide us with reasons. For instance, standards of normal development ('an oak should have deep, sturdy roots'), mere epistemic predictions ('the train ought to arrive in a minute'), or merely evaluative 'oughts' ('the world ought to be a better place'). Evaluative 'oughts' provide us with reasons only if we can, in some relevant sense of 'can', make true what they say ought to be the case. The central or deliberative ought must also be distinguished from the domain-relative 'oughts' of prudence or morality: If one *prudentially* ought to take a bath, it is not thereby something one has decisive reasons to do (there might be strong moral reasons against taking a bath due to the water and energy wasted); and it is a question open to philosophical dispute whether one always ought to do what one *morally* ought to do. However, the 'ought's of prudence and morality differ from other domain-relative 'ought's that don't provide us with normative reasons at all, such as etiquette or game rules: at least when it comes to practical normativity, they figure in the deliberative 'ought'.

A question that will be of high relevance for our discussion is whether what we *epistemically* ought to believe expresses a domain-relative ought, or whether we always have decisive reasons to believe what we epistemically ought to believe. I will argue in Chapter 5 that epistemic rationality should *not* be understood as one domain among others that are relevant to the overall deontic status of belief. Rather, even if there *are* distinctively practical reasons for belief that do not bear on belief's epistemic rationality (say, reasons provided by the *pleasure* of believing something), which I won't rule out in the present investigation (but see Chapter 3.2 for some discussion), these reasons do not interact with epistemic reasons in a way that would allow us to say what we should believe 'all epistemic and practical reasons considered'. Rather, I argue in this book that there is a point in retaining a distinctively *epistemic* normativity of belief that cannot be weighed or compared with practical normativity.

There are extensive discussions about what kinds of things reasons are,[16] how they are related to an agent's responses to reasons,[17] and what types of reasons there are.[18] Some of these questions will become relevant for our discussion, and I will note them when appropriate, but many can be put aside. Here I just mention one of my commitments: reasons – at least *normative* reasons – are *facts*. Here is a quick and dirty argument. When we φ for a reason, then our φing can be made intelligible in light of that reason: we can explain why we φ by reference to the reason for which we φed.

Yet in order to explain something one needs to cite a fact: something that is not the case cannot explain anything. I thus take reasons neither, *pace* Mitova (2017), to be beliefs or other mental states (though some reasons are facts about one's mind), nor, *pace* Schroeder (2021), to be sometimes false states of affairs or considerations. If a state of affairs or a consideration turns out to be false, then it thereby turns out not to be a reason. Yet false states of affairs or considerations can be *apparent reasons*. This is important because there are cases in which the same (or at least, a similar) relation holds between a false consideration and a subject's response as it holds between a reason and a response given for this reason. In such cases, the subject responds merely to an apparent reason (Alvarez 2010).

2.4 Responsibility and Blame

As I will put it, there are norms *to which* we are responsible, and norms *to which* we are not responsible. To be responsible to a norm means that one can, at least in the absence of excuse or exemption, be legitimately blamed or personally criticized for violating the norm. Again, the requirements of morality and prudence are norms to which we are responsible. Stealing someone's purse is something you can be resented for by the victim, and others can legitimately feel indignant about your thievery. You can regret that you did not start out to do earlier what you really want to do in life, and in this sense blame yourself for prudential failure.

By contrast, we are not responsible to the rules of etiquette, or to the rules of spelling. Although it seems that we sometimes *can* be legitimately blamed for failing to use the correct fork for the meal, or for misspelling another person's name, such blame is ultimately *moral* – or maybe, in some cases, prudential. What we are blamed for is not showing proper regard to people attending the dinner, or to the person whose name we misspelled. There is nothing wrong *per se* with using the other fork or writing down letters in another order than one is supposed to do by the rules of spelling. The blame at issue is not 'etiquette blame' or 'spelling blame'. Rather, it is moral blame. Because of this, we can say that we are not responsible to etiquette or spelling rules.

The distinction between norms to which we are responsible and norms to which we are not responsible displays a striking parallel to the distinction between norms that provide us with reasons and those that do not provide us with reasons, which I discussed above. I think this is not a coincidence. Rather, a norm that provides us with reasons – like a rational or moral requirement – is a norm "that someone, the subject of the norm, is accountable for conforming to (in suitable conditions)" (Kauppinen 2018: 3). I will elaborate on this idea throughout Parts 2 and 3 of this book. A close connection between responsibility and reasons is central to

motivating my investigation, and it will also be central to my solution of our problem.

The idea that responsibility is a matter of potentially being an appropriate object of blame (as well as positive attitudes, such as praise or gratitude) for one's actions, attitudes, or character can be traced back to Peter Strawson's (1962) influential idea that we can understand our responsibility practices by understanding the reactive attitudes involved in the kinds of relationships that only responsible beings can engage in with one another. Paradigm examples of blame are resentment and indignation as other-directed blame and guilt as self-directed blame. These emotions are only appropriate reactions toward actions and attitudes for which we are responsible. The other-directed versions of blame often have a certain significance for our relationships with other people in that they communicate that we take the other's action or attitude to impair our relationship with them. Considering the nature of blame – and the metaphilosophical question of what this dispute amounts to – will become essential for understanding our responsibility for attitudes throughout Parts 2 and 3 of the present book.[19]

For starters, we can work with a provisional understanding of blame, following Thomas Scanlon (2008: 122), who notes that blame is neither mere evaluation of a person nor a kind of punishment. I take this to be a consensus about the nature of blame. If I note that someone is lacking good eyesight, I evaluate the person, but I do not blame them. And if I blame them, I do not normally punish them: blame is an involuntary reaction to perceived wrong or norm-violation, and is thus not normally a kind of punishment, which is voluntarily imposing a burden on another person (cf. Hieronymi 2004). For sure, blame might be *instrumentalized* as punishment: I might cultivate my blaming emotions toward another person in such a way that I see them as punishments and ways of controlling another person. However, such instrumentalizing of one's own emotional life to manipulate others is toxic.

Understanding responsibility in terms of the openness to reactive attitudes, and openness to blame in particular, is non-committal about other conditions that must be satisfied for someone to be blameworthy besides that person being responsible. We are responsible for many of our actions and attitudes without being blameworthy for them. I am responsible for typing these words, but I am not blameworthy for doing this (I hope). One condition that needs to be fulfilled for me to be blameworthy is that by performing an action, I violate a norm to which I am responsible, like an (all-things-considered) requirement of morality. However, I am not always blameworthy for violating a norm to which I am responsible. For there are *excuses*. If you forget your best friend's birthday, but only because some horrible event distracted you from remembering it, then your friend should

excuse you from this interpersonal expectation to remember each other's birthdays. A promising definition of blameworthiness thus seems to be:

S is blameworthy for φing iff
- a S is responsible for φing, and
- b by φing, S violates a norm with which S (normatively) ought to comply, and
- c S is not excused for violating the norm by φing.

The definition implies the absence of exemptions because I take it that, if one is not responsible for φing, one is exempted. According to this definition of blameworthiness, responsibility is the most essential precondition for blameworthiness in the sense that the person must be both responsible for what they are blameworthy for (the action or attitude) and *responsible to the norm they violate* (the moral or prudential requirement, or the rational requirement). Responsibility, one might say, is the basis for all our practices involving the reactive attitudes – it is what makes these practices legitimate or appropriate (in some sense).[20]

It is common to distinguish *moral* responsibility from *causal*, *legal*, and *role* responsibility. A stone can be causally responsible for smashing a window, but the stone is neither legally nor morally accountable for the broken window because the stone is not an agent. In certain contexts, we are legally responsible without being morally responsible. We might be required by law to pay the bill for some damage we caused, although there is nothing that we could have done to avoid the damage we have to pay for.[21] We might also be responsible for something in virtue of our social role without being morally responsible. If we are the head of a company, we will have certain 'responsibilities' in the sense of *obligations* in virtue of that role. If we turn out to be incapable of fulfilling these responsibilities, and we are thus not *morally* responsible for fulfilling them (due to our incapability), then we might still have to suffer the consequences (like losing our position).

Yet it would be misleading to call the property I am focusing on *moral* responsibility. As I have mentioned above, we might also be *prudentially* responsible, and even prudentially blame ourselves, as when we feel regret about not having done something earlier. Furthermore, a central question I will pose is whether and how we are *responsible to requirements of rationality*. If we are responsible to these requirements, then the responsibility in question might not be the same as moral responsibility. In fact, even though I argue that sometimes rational failure can warrant moral blame, I will also argue that many responses, such as forms of *epistemic* blame, are quite distinct from moral blame, and thus do not presuppose *moral* responsibility. I thus use the term 'responsibility' as encompassing what philosophers usually understand as moral responsibility but also including prudential responsibility, and potentially responsibility to other norms

that provide us with reasons, like the requirements of epistemic rationality. Distinguishing our responsibility to different kinds of norms is a task for Part 3.

2.5 Back to Our Problem

We are now in a position to formulate our philosophical problem more precisely. In Chapter 2.2, I stated claim (1) of the trilemma as saying that we are directly responsible for being (ir)rational. This claim is puzzling. How can you be directly responsible just for *being* in a state, rather than for *causing* the state or *maintaining* it in some way? Intuitively, we might just solve the problem by rejecting this claim. However, I have already noted that it gains its support from the idea that irrationality is *criticizable* or *blameworthy* – an idea which presupposes responsibility for being (ir)rational. Moreover, I have pointed out that this responsibility seems to be more *direct* than indirect control over attitudes could explain: we are responsible for directly revising or even just for holding our attitudes in response to our reasons for those attitudes, rather than only for managing our mind with indirect methods, such as going for a walk or meditating. According to the terminology introduced above, this is just another way of saying that rationality places genuine *requirements* on us to revise or hold attitudes by directly responding to reasons. We can now grasp our problem better by specifying it as follows.

The Problem of Mental Responsibility, Final Formulation

1. We are directly responsible to the requirements of rationality.
2. We can only be directly responsible to the requirements of rationality if there is *some* form of control that could explain this direct responsibility.
3. There is no form of control that could explain why we are directly responsible to the requirements of rationality.

There are still questions about the content and normativity of rational requirements, which I will address in Chapter 3. In Chapter 4, we will see how responsibility to the requirements of *epistemic* rationality can become questionable. If epistemic reasons were not genuinely normative reasons, either because they are non-normative reasons (*epistemic anti-normativism*) or because they are not reasons at all (*epistemic nihilism*)[22] – then the intuition that we are directly responsible for complying with the requirements of epistemic rationality fades. For remember that this intuition was driven by the idea that beliefs are, in contrast to mere sensations or brute physical states, directly reasons-responsive. If beliefs are not responsive to genuinely normative reasons but are instead just caused by our evidence

(and by non-evidential factors), then maybe we are after all not directly responsible for complying with the 'requirements' of epistemic rationality: we would never be blameworthy *merely* in virtue of being irrational. Indeed, we might prefer not to talk about rational 'requirements' any longer because epistemic rationality wouldn't substantively require anything of us. It would just be a standard for evaluating how well our beliefs fit our evidence, but facts about whether our beliefs fit our evidence would, by themselves, have no further implications for our praise- or blameworthiness.

Thus, these solutions – *epistemic anti-normativism* or *epistemic nihilism* – would deny claim (1) of the problem of mental responsibility, at least when it comes to responsibility *for belief*: we aren't responsible to the requirements of epistemic rationality. This, in turn, would prepare the stage for Indirect Voluntarism (see Chapter 1). For remember that the main problem for Indirect Voluntarism is that beliefs and other attitudes are responsive to reasons, or evaluable as rational and irrational, which seems to imply direct responsibility for holding (ir)rational attitudes. This direct responsibility does not seem to be explainable by reference to indirect voluntary control. However, if reasons-responsiveness or rational evaluability does not imply responsibility for complying with the norms of rationality, then the path is free for Indirect Voluntarism: we wouldn't be *directly* responsible for beliefs since we wouldn't be responsible to rational requirements, and so any remaining *indirect* responsibility could be explained by reference to indirect control. Interestingly, Indirect Voluntarism about belief fits well with positions that deny the normativity of epistemic reasons, and it gets into trouble as soon as we allow that epistemic reasons are genuinely normative reasons.[23]

Part 2 will instead defend *normativism about epistemic reasons*: epistemic reasons exist, and they are genuinely normative reasons (just like moral and prudential reasons).[24] This will imply, I argue, that we are directly responsible to the requirements of epistemic rationality. The argument will also demonstrate, as I show in Chapter 6 by generalizing this conclusion, that we are directly responsible to other rational requirements – like rational requirements to desire, intend, and feel certain things. This refutes Indirect Voluntarism.

Importantly, the kind of responsibility I spell out won't require us to specify what kind of control is implied by it. For the normativity of reasons for attitudes must indeed be understood, in a sense, as an *evaluative* kind of normativity: it serves us for evaluating each other's character, which matters for how we should relate to each other. So, the solution to the problem that I will defend denies (2), i.e., the link between responsibility for rationality and control, while still maintaining a strong link between rationality and responsibility.

However, I won't argue in this book that responsibility to rational requirements does not presuppose *any* kind of control. Some authors claim

that responsibility for attitudes can only be explained by reference to a form of direct, non-voluntary control that we exercise in believing, desiring, feeling, or intending – that is, they argue that we exercise our agency just by *having* these attitudes or by just *being* in a mental state.[25] I will not argue against this. I grant that attitudes are *responsive to reasons*. According to these authors, this is often taken to imply that we exercise our non-voluntary agency in believing, desiring, feeling, or intending. Here, I will remain neutral about whether attitudes are exercises of control in some substantial sense. Rather, my claim is merely *that we need not assume* that they are exercises of control in order to make it *intelligible* how we can be directly responsible for our attitudes. That is, in order to understand how there can be such a thing as direct responsibility to rational requirements, we merely need to understand the kind of normativity attached to rational requirements. We might have to consider concepts of non-voluntary control in order to solve other problems. But we need not do so in the context of understanding mental responsibility.[26]

2.6 Summary

What gets our problem going is the assumption that we are directly responsible for our seemingly automatic responses to reasons for attitudes. This assumption gains its plausibility from the intuition that we are *criticizable or blameworthy as irrational* for failing to respond correctly to these reasons. If we were to ignore this intuition, then all attitudinal responsibility might well derive from responsibility for prior actions and omissions, which are clearly subject to normative requirements (namely, the requirements of prudence and morality). However, given this intuition, it is an open question whether we should maintain the claim that we are directly responsible to the requirements of rationality, or whether we should instead reduce all responsibility for attitudes to responsibility for managing our attitudes through actions. I will argue throughout this book that we should acknowledge genuine responsibility for rationality.

I have not yet talked much about the connection between rationality and reasons, but rather assumed that there is a tight connection between both concepts. This chapter showed us that thinking about how we can be responsible for responding to our reasons for attitudes, which often goes hand in hand with being criticizable or blameworthy as irrational, is central to solving our philosophical puzzle. So, to fully understand how the problem of mental responsibility comes down to a problem about the normativity of rationality, I need to explain how reasons for attitudes connect to rationality. This is one task of the next chapter. Its other task is to show how the following investigation contributes to the debate about the conceptual relationship between rationality and reasons.

Notes

1 In line with this, Owens (2003) argues that our aims in forming beliefs cannot interact with our other aims we have as agents. Importantly, the influential critique by Steglich-Petersen (2009) of Owens' argument is not in conflict with what I say here, for Steglich-Petersen claims that certain activities, which conceptually aim at forming a true belief (like inquiry or reasoning), can interact with our wider aims. A belief-formation in the non-intentional sense (see Schmidt 2016 and Arpaly 2023 on the ambiguity in 'belief formation') cannot interact with our wider aims as practical agents. To accept that a belief-formation (in the non-intentional sense of the term) can interact with our wider aims is to commit to pragmatism about reasons for belief (see Chapter 3.2).

2 See Soteriou (2020) on a sense in which we exercise control over our decisions by realizing them in future action.

3 Hyman (2015) argues that even inanimate things, like the sun, can act (it is *shining*). By contrast, I reserve 'action' or 'to act' for full-blooded agency that is intentional under a description. This is not to say that intentions and decisions are on a par with things that merely happen to us. It is merely to say that they are not actions: we do not exercise our agency when we act in the same way as we exercise our agency when we believe, desire, feel, intend, or even decide something. Attitudes or mental states could at most be exercises of what Hieronymi (2006, 2014, ms) calls 'evaluative control', but it's not plausible that they're exercises of voluntary control (Schmidt 2016).

4 I will refer to the relevant reasons as practical reasons, which are a subtype of state-given reasons (those provided by practical value). For this terminology, see Chapter 3.2. Note that, as Schroeder (2021) argues, some practical reasons might be reasons of the right kind. But they are not practical reasons of the reward-type of practical reasons.

5 For the problem with respect to emotions, see Oakley (1992: chapter 4). For the problem with respect to intentions, cf. Hieronymi (2006), Kavka (1983), Owens (2000: 81–82), and Pink (2009: section 6).

6 Assume that you know that you will not get the reward if you will actually perform the action: you receive the reward only if you merely intend the action without actually doing it. Without some roundabout routes (like ensuring that you will not perform the action in the future and making yourself forget about this fact), you will not be able to intend the action insofar as you are rational. Cf. Kavka (1983).

7 Levy (2007) is sometimes mentioned as an exception (cf. McCormick 2015: 84–86). However, even Levy grants that we can sometimes influence our mind indirectly in a controlled manner, which presupposes merely that we can reasonably foresee some of the consequences (under some description) that our actions have on our mind.

8 On stating philosophical problems as trilemmas, see Ernst (2008: 65–71).

9 For some recent opposition to the idea that irrationality implies personal critcizability, see Worsnip (2021: 27–32). I'll discuss this issue in more detail in Chapter 3.

10 Throughout this book, I will not make a difference between 'ought' and 'should'.

11 For a related understanding of normativity in terms of reasons, see Raz (2011: 85). I will follow Kiesewetter and Raz in this understanding for now. A main theme of this book will be a close connection between normativity and blameworthiness. I later propose that normative requirements – those that provide

us with reasons – are those that make us blameworthy if we violate them in the absence of excuse or exemption.

12 Brössel et al. (2013) argue that every domain provides us with domain-specific reasons. For example, there are 'financial reasons' that we have whenever a decision is financially beneficial (cf. Brössel et al. 2013: 285). Thus, even if we do not need any money at all (say, because we live self-sustainably somewhere far out in the forests), we would still have a financial reason to enrich our finances. Similarly, we would always have an 'etiquette reason' even though there is no independent reason to comply with the norms of etiquette in a given situation. Kiesewetter (2021) convincingly criticizes this view as merely stipulating a notion of domain relative reasons. It is much more natural to say that, if we do not need any money at all, then we also have no reason to enrich our finances. Thus, the norms of financial maximization or the norms of etiquette are paradigms of norms that do not provide us with reasons. It is this genuinely normative sense of 'reason' in which I will focus on as well.

13 However, it is not essential for a pro tanto reason that it can easily be outweighed. Some pro tanto reasons are very weighty reasons for an action or attitude. In cases where an action is strongly supported by a pro tanto reason, and yet we are not allowed to do the action because there is a better course of action available (i.e., one supported by weightier reasons), we might feel some kind of regret for not being able to both do what we ought to do and comply with all our pro tanto reasons.

14 I will throughout this book assume that we sometimes ought to believe certain things, that is, I assume that there are positive epistemic obligations (for a defense, see Simion 2024). The same applies to other attitudes: sometimes, there are decisive reasons for desiring, intending, and feeling certain things. This is evident because we would sometimes be criticizable or blameworthy if we didn't hold a certain attitude. The notion of blameworthiness that I develop in Part 3 of this book allows us to make intuitive sense of obligations to hold attitudes.

15 Broome (2013: 22–25) characterizes this 'ought' as the 'ought' that figures in the idea that it is irrational not to intend what one believes one ought to do. This gives rise to the enkratic requirement of practical rationality.

16 For a recent helpful discussion, see Alvarez (2018), who defends factualism about reasons, and Mitova (2017) for the view that epistemic reasons are true beliefs, as well as the contributions in Mitova (2019) for the contemporary dispute across epistemology and ethics. For the classical version of psychologism about reasons (reasons are mental states), see Davidson (1963). As I explain below, I follow Alvarez in her factualism.

17 Here debates about the 'basing-relation' between a reason and a response for this reason are relevant. On acting for a reason, again see Davidson (1963) as the locus classicus of the debate. Davidson defends the idea that the reason for which we act is the primary cause of that action. Anscombe (1957) provoked positions that deny that explaining actions by citing the agent's reasons are causal explanations. For some overview of the debate, see the volume of D'Oro and Sandis (2013). On the basing-relation in epistemology, see Korcz (2019).

18 The probably most salient distinctions are between normative, motivating, and explanatory reasons. These distinctions are not necessarily ontological, but functional: what justifies my action or attitude (normative) might be what motivates me to do it (motivating), and thus explains why I do it (explanatory). Yet these reasons can come apart in various ways. For a good overview, see Alvarez (2017).

19 There is a significant discussion about what other reactive attitudes are instances of blame (see Coates and Tognazzini 2013). Does blame necessarily have an emotional or even passionate component (like the reactive attitudes just mentioned) (Wallace 1994, 2011)? Or can we also blame someone by merely distrusting the person, or even by merely judging (dispassionately) that our relationship towards this person is impaired due to their action or attitude (Hieronymi 2004, ms; Scanlon 2008; Smith 2013)? Can blame be analyzed in terms of behavioral dispositions that are organized around a characteristic belief-desire pair (Sher 2006, 2009)?

20 Recently, Lasonen-Aarnio (forthcoming: chapter 6) has argued that we should radically detach blameworthiness from the violation of norms, and rather link it to a failure of manifesting success-conducive dispositions. However, maintaining (b) in my definition of blameworthiness is compatible with her argument. This is because I focus on norms that require one to respond correctly to reasons, which involves manifesting success-conducive dispositions insofar as it requires one to properly base one's beliefs on one's possessed or available reasons (see also Chapter 3.1 and endnote 3 in that chapter).

21 Following Frankfurt (1969), some authors deny that alternative possibilities are necessary for moral responsibility. This does not matter for my argument here. Whatever the conditions are on moral responsibility that one accepts, they might not be fulfilled, and one would still have to pay the bill because one is legally responsible.

22 I take the labels from Kiesewetter (2021).

23 The reader might wonder why I assume such a close connection between normative reasons and responsibility. I motivate my view about this connection in more detail in Chapters 4–5. On my view about the connection between epistemic reasons as well as other 'right-kind' reasons and rationality, see Chapter 3.

24 To my knowledge, the only explicit defenses of this view in the present literature are Kiesewetter (2021), Kauppinen (2023), Paakkunainen (2018), and Schmidt (2024a, forthcoming c).

25 See esp. Boyle (2011), Hieronymi (2006, 2008, 2014, ms), and Smith (2005). For some objections against these views, see Chrisman (2020).

26 One such context in which we need to consider concepts of direct, non-voluntary control might be the problem of understanding how there can be agency in a world that is dominated by natural law – that is, the traditional problem of free will (see Hieronymi ms; Wagner 2015).

3 Rationality and Reasons

> The notion of irrationality we are interested in when asking for the normativity of rationality – the one that is associated with legitimate criticism – does, I think, require the capacity to modify one's attitudes in the light of reflection, and thus the absence of compulsion.
>
> Benjamin Kiesewetter, *The Normativity of Rationality* (2017: 100)

To introduce the topic of rationality and why I use the term 'rational' as intimately connected with reasons, I will first rehearse the recent debate insofar as it is relevant for the present investigation (Chapter 3.1). I will then explain the distinctions between the right kind of reasons and the wrong kind of reasons, as well as between object-given and state-given reasons (Chapter 3.2). These distinctions are relevant to our overall investigation, and they will help us to specify a central use of 'rational'. To be rational, according to this use, implies that one *responds correctly to one's reasons of the right kind*. In the remainder of the chapter (Chapters 3.3–3.7), I argue why we should make room for such a concept of rationality, at least for philosophical purposes of theorizing about genuinely normative questions in the ethics of mind, and how this concept implies a kind of normative structural rationality. The impatient reader who is familiar with debates on reasons and rationality might skip Chapters 3.1 and 3.2 and proceed to the main argument of this chapter that I begin developing in Chapter 3.3.

My ambition in the remainder of this book is not to put forward yet another theory of rationality. Instead, I defend the view that we are *directly responsible* for (ir)rational attitudes. The present chapter further sets up my defense. While the last chapter has focused on connecting mental responsibility with attitudinal norms and reasons, this chapter connects norms and reasons with rationality. It will finish our dialectical setup by showing in more detail how the problem of mental responsibility is a problem about the normativity of rationality. Moreover, the chapter shows

DOI: 10.4324/9781003382973-4

how thinking about responsibility can help us to clarify the relationship between rationality and reasons.

3.1 The Requirements of Rationality

As explained in Chapter 2.3, not all norms provide us with reasons. The fact that certain sheets of paper 'ought' to be A4-sized or that oaks 'should' have deep sturdy roots does not imply that one always has a reason to ensure that sheets of paper are A4-sized or to ensure that oaks have deep sturdy roots. Sometimes, it might be true that we should mess with the size of a sheet of paper or that we should dig out the roots of an oak – and in these cases, we might rationally do so without regret or feelings of guilt, and there's no need to excuse our actions. It is different when it comes to the requirements of morality. When something is morally required, we at least have a strong *pro tanto* reason to do it, or maybe even a *pro toto* or decisive reason to do it, and we are blameworthy for violating the moral requirement absent excuse or exemption.

What about the requirements of rationality? Does the fact that we *rationally* ought to have a belief, desire, emotion, or intention imply that we have *a reason* or that we *all-things-considered ought* to be rational? Answering this question with 'yes' is to defend the *normativity of rationality* (see Kiesewetter 2017, 2020; Lord 2018; Worsnip 2021). Many worries with the idea that rationality is normative arose from a conception of rationality as mental coherence. To be rational, according to this conception, is that one's attitudes are coherent with one another in specific ways. Requirements of so-called *structural rationality* might thus take forms like: 'If you believe that you have sufficient evidence for p, then you ought to believe that p' or 'If you believe that you ought to φ, then you ought to intend to φ' or 'If you believe that you ought to φ, and you believe that ψing is a necessary means to φing, then you ought to intend to ψ'.

A main worry with the normativity of these requirements is that they would, if they were normative, give rise to unacceptable *bootstrapping* (Kolodny 2005: 514–542; Kiesewetter 2017: chapter 4): we could make it the case that we ought to believe or intend something just by adopting the antecedent attitudes without any reason for them. According to the second standard of rationality, for example, it would be true that I ought to intend to scream at you if I now just arbitrarily adopt the belief that I ought to scream at you. This seems implausible. In reply, it has been suggested that the standards of rationality take wide-scope rather than narrow-scope form.[1] However, it has been argued convincingly that the wide-scope versions of the standards also give rise to unacceptable bootstrapping (Kiesewetter 2017: chapters 4.4–4.7). Furthermore, wide-scope standards

of rationality seem to implausibly imply that each way of satisfying the standard is rationally on a par (Kiesewetter 2017: chapters 6.4–6.5).

Such problems[2] for a normative conception of structural rationality have provoked defenses of the idea that rationality is not just a kind of mental coherence. Rather, to be rational is *to respond correctly to one's possessed or available reasons* (Kiesewetter 2017; Lord 2018). These accounts avoid the bootstrapping problem because we don't respond correctly to our reasons when we adopt an antecedent attitude of a coherence requirement without any reason for this attitude. For example, if I now believe that I ought to scream at you without any reason, then it does not implausibly follow, according to this conception of rationality, that I ought to adopt the intention to scream at you. This is because neither my belief nor my intention is a correct response to my reasons: the fact *that I believe that I ought to scream at you* is not a reason to intend to scream at you, nor is my belief or my intention self-justifying.

Indeed, rationality as reasons-responsiveness seems to imply immediately that we always ought to be rational. For it seems trivially true that we always ought to respond correctly to our reasons. Proponents of rationality as reasons-responsiveness took this assumption for granted. However, as we will see in Chapter 4, this claim is not as trivial as it seems. There I will spell out a *neglected challenge* for the normativity of rationality arising from recent doubts about whether reasons for attitudes are 'genuinely normative reasons'. The remainder of the book will offer a reply to this challenge in Chapters 5–7.

The notion of 'responding correctly' to reasons means both to give the response that is favored by the reasons (for example, if one has decisive epistemic reasons to believe that p, then the correct response to these reasons is to believe that p) *and* to give the response *for* those reasons (that is, to believe p *on the basis* of the relevant evidence). That is, I might believe what my reasons support but my belief need not be related to my reasons in the right way so as to count as a *correct* response. For instance, if you tell me that you are in town right now, this might give me decisive reasons to believe that you are in town. But if I then believe that you are in town because I flipped a coin and it showed heads, then I did not respond correctly to my reasons and I am not rational. I will not be concerned in this book with spelling out what it is to believe something *for* a reason (for an account, see Lord 2018: chapters 5 and 6). If I believe that you're in town, and this belief is supported by my evidence, the belief is *propositionally* rational, yet it's not thereby *doxastically* rational for me to hold this belief: after all, I might believe that you're in town on the basis of my coin-flipping rather than on the basis of my evidence. My focus is on norms that require responding *correctly* to reasons, and so on *doxastic* rationality when it comes to the rationality of belief.[3]

Accounts of rationality as responding correctly to one's reasons are not just concerned with the reasons there are, but with reasons which are *available to* or *possessed by* the subject. This is because the accounts must capture the intuition that rationality *supervenes on the mental*: mental duplicates that live in radically different environments will both be rational, even if one duplicate's attitudes are inaccurate about their environment. One duplicate is misled (say, by an evil demon), but it is still rational. This is the plausible *internalist* conception of rationality (Wedgwood 2017, 2023): whether one is rational cannot be a matter of facts that are inaccessible to us. Even if there *is* a reason to leave my office now because it will all blow up in five minutes, this reason does not make it rational for me to leave my office if I have no clue at all about the fact that it will all blow up (Parfit 2001). Proponents of reasons-based accounts of rationality thus spell out theories according to which it is the agent's *available* or *possessed* reasons that determine what responses the agent ought to give.[4]

Reasons-responsivist accounts of rationality thereby position themselves against *objectivism* about 'ought' and reasons. Objectivism states that 'S ought to φ' means that φing is the best option, no matter whether S is in a position to know or has some kind of cognitive access to whether φing is the best option. For example, the objectivist would claim that I *ought* to leave my office in the example above, even if I have no clue that it will blow up soon. By contrast, the *subjectivist* would deny this and say that I only *ought* to leave the office if I – in some way or other – have cognitive access to the fact that it will blow up, i.e., if the reasons for leaving the office are *possessed* by me. If I possess the reasons because I know that the office will blow up, but I do not leave the office, then I am blameworthy. According to the objectivist, I might not be blameworthy for staying in my office, and yet I *ought* to leave the office – namely, when am not in a position to know that it will blow up, but it will blow up.

In the present investigation, I will mostly set aside the debate between objectivism and subjectivism by restricting my discussions to cases in which the relevant reasons are available to or possessed by the subject. If it turns out that Broome (2020) was right that normativity is about objective reasons, while rationality is about the reasons that you have or that are available to you from your perspective, then my view is that rationality is only normative when the objective reasons are possessed by you or are available to you. However, if they are possessed or available, then rationality is normative at least *in these cases*, according to my account. I won't bother to discuss the relevant notions of availability or possession.

When I use the term 'rational' in the following investigation, I will use it in line with the reasons-based conception of rationality à la Kiesewetter (2017) and Lord (2018), rather than a conception of rationality as mental

coherence à la Broome (2013). However, this is not a mere stipulation. For, as I will argue throughout Chapters 3.3–3.7, *if* accounts of rationality as mental coherence are interested in a property that is closely linked to personal criticizability (and so to responsibility), then being irrational in the coherence-sense *implies* that one fails to respond correctly to one's reasons, at least when the relevant reasons are possessed the subject. In other words: if one responds correctly to one's (possessed) reasons, then one is also rational in the coherence sense, insofar as "coherence" is meant to pick out a property linked to criticizability. This argument is a novel contribution to the debate on the normativity of rationality.[5]

I will nevertheless remain neutral about whether all kinds of rationality are *substantive* in the sense that they consist in responding correctly to reasons. As Worsnip (2021) has recently argued, the view that there is a *substantive* kind of rationality could be compatible with the view that there is also a *structural* kind of rationality that consists in a kind of coherence. To illustrate the difference between both kinds of rationality, Worsnip (2021: 5–6) contrasts a case in which a person, Tom, believes against his evidence that he is superman, that superman can fly, but that he (Tom) cannot fly, with another case in which another person, Tim, believes against his evidence that he is superman, that superman can fly, and that he (Tim) *can* fly. According to Worsnip, Tim is structurally more rational yet substantively more irrational than Tom. For Tim seems consistent in his beliefs while harboring two beliefs against his evidence, whereas Tom is inconsistent in his beliefs albeit merely harboring one belief against his evidence.

As we will see later (Chapter 3.7), the distinction between structural and substantive rationality might collapse if we think that only evidence *that is believed* can justify one's beliefs. Tom's failure to respond correctly to his epistemic reasons would then just consists in an inconsistency of his beliefs about evidence with his belief that he is superman. However, I do not commit to substantial claims about how best to distinguish substantive from structural rationality here, or about whether they can be distinguished at all. For, as I explain in Chapter 3.7, discussing this would bring up deeper philosophical disputes that I cannot settle or even address in detail. I am thus open to Worsnip's account that there might be a *dualism* of rationality.

What I will argue for now is that structural irrationality, by itself, cannot amount to a very serious failure if one doesn't always have decisive reasons to be structurally rational. The present investigation therefore focuses on substantive (ir)rationality. This is the rationality that seems to place genuine normative requirements on us – requirements we are responsible for complying with, and which give rise to blameworthiness when we fail to comply without excuse. In cases where structural rationality does not imply decisive reasons to be structurally rational, we won't be blameworthy

or criticizable for failing to comply with its demands. That is, the charge of irrationality would be inappropriate in such cases.

3.2 Rationality and Kinds of Reasons

To be rational implies, according to my use, to respond correctly to one's *reasons of the right kind* for the attitude (or *right-kind* reasons).[6] They can be distinguished from reasons of the wrong kind (or *wrong-kind* reasons) by two features.

First, right-kind reasons are those reasons that bear on the distinctive rationality of an attitude. Beliefs, for instance, are subject to *epistemic* rationality. Epistemic reasons are those reasons that bear on this kind of rationality. Beliefs might also be subject to *practical* rationality: it might be practically, but not epistemically, rational to believe that your friend will get the job, if this belief allows you to boost your friend's self-confidence by being more supportive, while you lack sufficient evidence that they will get it (say, you don't know who else has applied). However, practical rationality is not distinctive for belief, because actions and maybe other attitudes are also subject to practical rationality. Reasons bearing *only* on the practical rationality of belief, but not on its epistemic rationality, are thus 'of the wrong kind' in this specific sense: they don't bear on belief's distinctive rationality, which is epistemic.[7]

Second, epistemic reasons and other reasons of the right kind (for other attitudes) are characterized by the fact that they are reasons *for which* you can clearly adopt an attitude: they are those normative reasons that can also clearly be your motivating reasons. For instance, it is clearly possible to believe that your friend will get the job *for the reason* that your friend is the best candidate (and the application system is fair); but it is not clearly possible to believe that your friend will get the job *for the reason* that this belief would make you more supportive. It seems somehow *difficult* to adopt your belief for such practical reasons. This is why they seem to be, intuitively, 'of the wrong kind' and not bearing on belief's *epistemic* rationality.[8]

Traditionally, right-kind reasons are therefore conceived of as being co-extensive with *object-given reasons* for an attitude (see Parfit 2001: 21–22). Object-given reasons for an attitude are reasons that indicate (or constitute) facts about the attitude's object rather than about the attitude itself: object-given reasons are facts that support the attitude, or make it rational to have the attitude, by indicating that or making it likely that the attitude fulfills its constitutive aim. For example, object-given reasons for *beliefs* are (or are provided by) evidence, because evidence indicates the truth of the object of the belief, i.e., the truth of the belief's propositional content. Scientific reports on climate change are thus object-given reasons

for belief: they indicate that human-induced climate change takes place. By contrast, that I feel less existential angst if I do not believe in climate change is a *state-given* reason not to believe in climate change. It is not an object-given reason against climate change, because this fact does not indicate that the belief fulfills its constitutive aim of truth.[9]

Analogously, an object-given reason for a *desire* shows the object of the desire to be desirable in some respect; object-given reasons for *fear* indicate the danger of what you fear and thus make it rational to experience fear; object-given reasons for *intention* are reasons for the intended action: that I will get poisoned if I drink a toxin is an object-given reason not to intend to drink it in virtue of being a reason not to drink it; and that I get a lot of money for intending to drink a toxin is a state-given reason to intend to drink the toxin.[10]

In contrast to object-given reasons, *state-given reasons* indicate (or constitute) facts about the state of the attitude itself. The most important category of state-given reasons under discussion is *practical reasons for attitudes*, i.e., reasons that support the attitude by indicating the attitude's value. According to those who believe that there are such practical reasons – i.e., *pragmatists* – the fact that it would be beneficial to believe in a guardian angel (e.g., because one would sleep better at night) could be a practical reason to believe in a guardian angel (even in the absence of sufficient evidence).[11] Analogously, facts about the value of any other attitude are considered to be reasons for this attitude by pragmatists.

It is often argued that practical reasons are not even *reasons* for attitudes, but at best *considerations* indicating the attitude's value.[12] It is sometimes argued that they are not reasons because if they were, then we could sometimes form attitudes *for* practical reasons: we could believe, desire, feel, or intend at will. We could just decide not to fear a dangerous tiger or just decide to intend to drink a poisonous substance just because we regard fearing or intending to be beneficial. It thus seems that, intuitively, since practical reasons for belief cannot properly motivate belief, practical reasons for belief are not really reasons for belief, but merely reasons for *bringing beliefs about* or for *causing beliefs*.

However, the issue is more complicated. For we might be able to believe for practical reasons although we cannot believe at will. Pragmatists now employ various strategies to support this view. According to the *first strategy*, which I have elsewhere dubbed the *new pragmatists' strategy* (Schmidt 2022), our ability to control our beliefs indirectly is sufficient for believing for practical reasons: that we can *cause* or *bring about* our beliefs for practical reasons is sufficient for the possibility of basing our beliefs on practical reasons.[13] Opponents of this form of pragmatism (myself included) argue that indirect doxastic control does not amount to *believing* for practical reasons but merely to *acting* for practical reasons. Indeed,

these forms of pragmatism face the serious challenge of finding a theoretical purpose for assuming that there are practical reasons for belief *in addition* to practical reasons for causing belief (Schmidt 2022: 1815–1816): what is the point of saying that if believing in God is pleasurable for you, this fact does not merely provide you a reason for *causing* the belief (by going to church, engaging only with believers, etc.), but also a reason for *believing* in God? The practical reason to believe seems *superfluous* to the practical reason to cause the belief. Pragmatists have not yet answered this challenge, which was already noted by Parfit (2011: 432).

This gives rise to the current dialectic: while pragmatists argue that their opponents are overly restrictive in allowing for practical reasons to *cause* beliefs but not for practical reasons to *believe*, their opponents ask why one should allow for the latter. If we can say all that we might wish to say about the practical side of belief normativity by talking about reasons for managing our beliefs indirectly (that is, by talking about causings or bringings about), on the one hand, and about epistemic reasons for belief, on the other, then it is unclear why we should allow for practical normativity attaching to belief itself.[14]

However, there are other strategies forward for pragmatism that I won't consider within the present investigation. The *second strategy* is to appeal to a form of direct control over beliefs that does not amount to believing at will, but that still enables us to believe for practical reasons (McCormick 2015: chapter 6). The *third strategy* is to argue that sometimes there are normative reasons for an action or attitude even though we cannot believe *for* such reasons (Schroeder 2007). The *fourth strategy* is to restrict the relevant kinds of practical reasons to some state-given reasons that can plausibly motivate *not* having a belief and that bear on beliefs *epistemic* rationality rather than on a separate kind of practical rationality: the practical stakes of holding a false belief or the fact that further evidence will become available (Schroeder 2021). Such reasons might also be interpreted as epistemic reasons for suspending judgment (Lord 2020). If the fourth strategy is successful, then some state-given reasons for belief would bear on belief's epistemic rationality and so be reasons 'of the right kind'.[15]

The present investigation will largely remain non-committal on these issues. What I defend is the view that there is a distinctively *epistemic* rationality of belief that has normative force, and that other kinds of attitudes display their own distinctive kinds of rationality that have their distinctive normative force. Reasons of the right kind for an attitude are those reasons that bear on these kinds of rationality that are distinctive for each mental state. These right-kind reasons might include state-given reasons. Furthermore, reasons 'of the wrong kind' might still be relevant to the distinctively *practical* rationality of holding an attitude, if there is such a thing. At the very least, wrong-kind reasons are relevant for practical norms of

managing our attitudes. What I will defend is that object-given reasons, which are clear-cut cases of reasons of the right kind, have normative force on their own: for instance, evidence provides us with genuinely normative epistemic reasons for belief, and it does so *all by itself* (Chapter 5).

With this overview over the relevant dialectics in recent debates, we can now turn to my argument that we should focus on rationality as reasons-responsiveness, insofar as we are interested in a serious failure of irrationality that implies criticizability absent excuse.

3.3 The Object of Theories of Rationality

John Broome claims that "[m]ost philosophers who write about rationality intend to write about it as it is commonly understood" (Broome 2020: 300). That is, he claims that not only the object of his own theory of rationality is determined by the ordinary use of 'rational' but also the object of the theories he disagrees with is determined by ordinary use. Broome's claim is:

(I) The object of theories of rationality is determined by our ordinary concept of rationality.

As Broome specifies, rationality is commonly understood as a mental property of a person; furthermore, rationality can be 'reified' in the same way as morality can be reified: it can be conceived of as an entity that places requirements on us. If we fulfill all those requirements, then we possess the property of rationality to the highest possible degree.

Another claim that is suggested by Broome's brief remarks on the metaphilosophy of rationality should be distinguished from (I):

(II) The method of theories of rationality is to analyze 'rational'.

I am not sure whether Broome analyses 'rational'. Conceptual analysis – spelling out necessary and jointly sufficient conditions for the correct application of a word – contrasts with conceptual *explication*. The latter seems closer to what other theorists of rationality do. Lord (2018: 6) says that he provides a "real definition" of the property of rationality, which he contrasts with conceptual analysis. The idea must be, roughly, that our ordinary concept might fail to track the property that philosophers are, or should be, interested in. Wedgwood (2017: 23) points out that he engages in "constructive theory-building" – a kind of theorizing that contrasts with philosophy that is "closer to everyday thought". Both remarks suggest that what these philosophers do is to explicate a concept of rationality, maybe in order to carve normativity at its joints (Lord), or maybe to show that we can conceive of a property that supervenes on the mind, has

normative authority, and is a good means to achieve an external aim, like truth (Wedgwood). Theorists of rationality might thus start out by making claims about our ordinary use of 'rational', but they then deviate from this use for specific theoretical purposes.

It can lead to significant misunderstandings in a debate if there are disagreements about its object or its method: we might end up talking about different things or engage in different kinds of projects. In what follows, I focus on the *object* of current theories of rationality. I will argue that theorists of rationality are not just concerned with the ordinary use of 'rational'. Rather, due to their philosophical interests they pick out a specific use that is intimately connected to blame and praise. I call the property this use refers to 'rationality$_{\text{RESP}}$', because it presupposes the subject's direct responsibility for their attitudes. I then present an argument why rationality$_{\text{RESP}}$ plausibly implies responding correctly to reasons.

Philosophers often pick out specific uses of a term in order to determine the object of their theory. For instance, Broome thinks that

> [f]or the sake of philosophical analysis, we must expect to have to give 'ought' a more precise meaning than it has in common English. At the very least, we may exclude some ordinary uses of 'ought' in order to avoid ambiguity.
>
> (Broome 2016: 6)

The uses of a word on which a philosophical theory focuses are ideally determined by the philosopher's interests. One of the central philosophical interests in thinking about rationality is to understand the normative significance of rational requirements – as a central interest in thinking about morality is to understand the normative significance of moral requirements. I argue in this subchapter that this specific theoretical interest justifies a restriction to uses of 'rational' that refer to instances of rationality *for which we are held responsible*.

To see what I mean by rationality$_{\text{RESP}}$, remember how irrationality is used as implying criticizability in the debate (see Chapter 2.2). When Parfit uses 'irrational', he uses the term "in its ordinary sense, to mean, roughly, 'deserves strong criticism of the kind that we also express with words like 'foolish', 'stupid', and 'crazy'" (Parfit 2011: 123). Kiesewetter (2017: chapter 2) points out that we use 'irrational' in order to criticize another person for a response (see also Kauppinen 2023: 540–542), rather than merely for evaluating a response as bad or the person's rational subsystem as malfunctioning. Proponents of the current debate take the criticizability of irrationality to support the thought that rational requirements have a certain authority that other norms – say, those of etiquette – lack (Way 2009: 1; Lord 2018: 4).

Insofar as theorists of rationality are interested in a property closely connected to criticism and praise, they restrict their argument to the uses of '(ir)rational' that interest them, and exclude uses of 'irrational' that, for example, refer to pathology. In this vein, Kiesewetter writes that

> [t]he notion of irrationality we are interested in when asking for the normativity of rationality – the one that is associated with legitimate criticism – does, I think, require the capacity to modify one's attitudes in the light of reflection, and thus the absence of compulsion.
>
> (Kiesewetter 2017: 100)

Being irrational$_{\text{RESP}}$ presupposes that the person is directly responsible for the attitudes that are evaluated as irrational, and that she is consequently a potential target of legitimate criticism for holding these attitudes. To see this, compare it with uses of 'irrational' that do not imply direct criticizability. Arachnophobia is, in a sense, irrational – a person suffering from it often fears a spider knowing that it is not dangerous. And yet we do not criticize arachnophobes directly for their fears. Rather, we hold them at most indirectly responsible (for not doing therapy, for example). Such uses of 'irrational' do not raise any question for the normative authority of a requirement to be rational. There just is no requirement in place not to fear spiders if the fear is something for which the person isn't directly responsible. Any question for the normative authority of rational requirements arises only if we are responsible for whether we comply with these requirements. Thus, current theories of rationality (should) focus on rationality$_{\text{RESP}}$. For it would be pointless to *require* someone to be rational in cases of irrationality for which one is not directly responsible. To refer to severely pathological forms of irrationality, that is, to those that aren't sufficiently responsive to reasons anymore, I will use the term 'arational'.

3.4 The Argument from Responsibility

A straightforward way of defending the normativity of rational requirements is to argue that their normativity is implied by the fact that violating rational requirements makes us criticizable. I will now proceed by, first, pointing out a *prima facie* problem for any such argument from criticizability. I then modify this argument in a way that avoids this problem. The modified version of the argument builds on the idea that the criticism in question presupposes responsibility – that is, amounts to a form of *blame*.

The *prima facie* problem for any argument from criticizability is that not all forms of criticizing a person imply that they have violated a normative requirement. Take the forms of criticism Parfit mentions ('foolish', 'stupid',

'crazy'). Such expressions might be applied to a person because of some cognitive malfunctioning for which they are not responsible. Calling them stupid might be unfair or unjust, but it isn't clearly *false*. Yet if the person isn't even responsible for the stupidity, then it's not the case that they *ought not* to be stupid. The criticizability-intuition thus can support the normativity of rational requirements only if we take the criticism in question to presuppose direct responsibility for the irrational attitudes – that is, if we take the criticism to be a form of *blame* for the person's attitude.

I thus suggest a modification of the argument from criticizability, which I call *the argument from responsibility*. It starts by pointing out that merely being incoherent is not a good ground for blame. For incoherence might be pathological: if you believe that p and believe that not-p, then you need not be blameworthy for holding your incoherent set of beliefs. In such cases, you might be *stuck* with your incoherence: your beliefs might not be sufficiently responsive to reasons for you to be responsible for them.

The argument proceeds by noting that reasons-responsiveness is a plausible ground for responsibility: in order for you to be directly responsible for your attitudes, they must be sufficiently responsive to reasons. Failing to *correctly exercise* your capacity of reasons-responsiveness can then make you blameworthy. You are responsible for holding (in)coherent attitudes only if your incoherent attitudes are sufficiently reasons-responsive (i.e., you *would* revise them under certain circumstances). Therefore, it seems that $\textit{rationality}_{RESP}$ is plausibly a kind of reasons-responsiveness rather than a kind of coherence.

A worry is that this argument confuses reasons-responsiveness – a capacity that grounds responsibility for attitudes – with responding correctly to reasons – which is a specific way of exercising reasons-responsiveness. It might be true that any attitude that is evaluated as (ir)rational_{RESP} needs to be sufficiently responsive to reasons. But that does not imply that whenever we are (ir)rational_{RESP}, we (fail) to respond correctly to our reasons.

In reply to this objection, let me spell out the argument in some more detail. The objection assumes the falsity of the following principle that allows us to connect the capacity of reasons-responsiveness to individual exercises of that capacity:

> *Norms and Capacities (NC)*. If you are responsible for whether you comply with norm N in virtue of a capacity C, then your praise- or blameworthiness for complying or, respectively, violating N is grounded in a correct exercise or, respectively, failure to correctly exercise C.[16]

If NC is true, and if our responsibility for whether we are (ir)rational is grounded in reasons-responsiveness, then our praise- or blameworthiness for being (ir)rational is grounded in a correct exercise or, respectively,

failure of reasons-responsiveness. On the further assumption that (ir)rationality$_{RESP}$ is *always* praise- or, respectively, blameworthy, it follows that, whenever we are responsible for our (ir)rationality, we respond correctly or, respectively, fail to respond correctly to our reasons. Thus, rationality$_{RESP}$ implies the correct exercise of reasons-responsiveness, and irrationality$_{RESP}$ implies failing to correctly exercise reasons-responsiveness.

NC has initial plausibility when we think about moral requirements: It seems that, *if* our responsibility for complying with moral requirements is grounded in our ability to voluntarily control our conduct, then we are praise- or blameworthy in virtue of our correct exercise or failure of voluntary control. For instance, if I should raise my hand to re-elect the president, but I fail to raise my hand, then I am blameworthy in virtue of the fact that I didn't raise my hand – which is a failure of correctly exercising my capacity of voluntary control. Or consider my responsibility for riding my bike, which is grounded in my capacity to ride my bike; if I run into someone due to my own fault, then I failed to exercise my capacity to ride my bike correctly – which is just an instance of exercising my capacity of voluntary control. My failure to exercise this capacity correctly explains why I am blameworthy.

This is the argument from responsibility. To see its structure more clearly, consider the following version of it, which employs a version of NC that focuses on *irrationality and blame*, rather than on rationality and praise (premise (1)):

1 If you are responsible for whether you comply with norm N in virtue of a capacity C, then your blameworthiness for violating N is grounded in a failure to correctly exercise C.
2 We are responsible for complying with the norms of rationality in virtue of our capacity to respond to reasons (for attitudes).
3 Thus, our blameworthiness for violating the norms of rationality is grounded in a failure to correctly exercise our capacity to respond to reasons (for attitudes).
4 We are blameworthy for violating the norms of rationality$_{RESP}$ (absent excuse).
5 Therefore, if we violate the norms of rationality$_{RESP}$ – that is, if we are irrational$_{RESP}$ – then we fail to correctly exercise our capacity to respond to reasons (for attitudes).

As we will see, there is an important objection to premise (4) that we need to consider. Premise (2) will be defended throughout the book: I show that various blaming responses are appropriate when someone fails to respond correctly to right-kind reasons. It also gains support from recent accounts of direct responsibility for belief and other attitudes (see Chapter 1.1).

Consider briefly the reverse claim of conclusion (5), namely that, if you're rational, then you respond correctly to your reasons. I take it that this claim is true in a trivial sense: *failing* to respond correctly to your reasons makes you *substantively* irrational. This leaves it open whether you might still be *structurally rational* when you fail to respond correctly to reasons: say, your beliefs might still cohere with each other, even though you didn't take up certain perceptual or testimonial evidence. Such cases could well be possible, and I therefore don't think that the reverse claim of (5) can easily be defended.[17]

So, the best chance for me to defend a non-trivial connection between rationality and reasons, I take it, is claim (5), which says that irrationality – *whether structural or substantive* – implies a failure to respond correctly to reasons. That is, responding correctly to reasons does not just trivially imply that one is substantively rational, but it also implies a kind of *structural* rationality. The remainder of this chapter defends my argument for this claim in some detail.[18]

3.5 Blameless Irrationality$_{RESP}$?

Note that (3) says that, *if you are blameworthy for irrationality*$_{RESP}$, then you failed to respond correctly to your reasons. However, there might be some instances of irrationality for which we are responsible yet blameless. In such cases, you might respond correctly to your reasons, although you are irrational, as far as the argument goes. The blameless irrationality at issue might be an irreducible kind of *incoherence-irrationality*. So we need a defense of premise (4): if you are irrational$_{RESP}$, then you are blameworthy. It will be helpful to return to our definition of blameworthiness from Chapter 2.4:

S is blameworthy for φing iff
- a S is responsible for φing, and
- b by φing, S violates a norm with which S (normatively) ought to comply, and
- c S is not excused for violating the norm by φing.

Accordingly, there can be three reasons why one is blameless for irrationality. First, (a) might not hold: we might not be responsible for the attitudes that constitute our irrationality. However, our focus on irrationality$_{RESP}$ rules out such cases as mere cases of *arationality*. Second, according to condition (c), one might be blameless for one's irrationality because one is *excused*. However, if there are such cases, then they do not pose a problem for the argument from responsibility. This is because if the *only* reason why one is blameless for one's irrationality is that one is excused, then (a)

and (b) are still fulfilled: one still failed to respond correctly to decisive normative reasons. So, irrationality that is blameless merely due to an excuse is not a counterexample to the view that irrationality$_{RESP}$ implies failing to respond correctly to reasons.

The most challenging type of case for (4) would be one in which only (b) does not hold, but (a) and (c) are fulfilled: a case in which one is blameless for one's irrationality, even though one is neither exempted nor excused. Rather, one would be blameless because one responded correctly to one's reasons. If there are such cases, then irrationality$_{RESP}$ wouldn't always imply a failure of not responding correctly to reasons.

The literature on rationality contains a range of cases in which we might be irrational while responding correctly to our reasons. In such cases, we are *merely* irrationally incoherent while responding correctly to our reasons. Proponents of reasons-responsiveness views of rationality argue that in all such cases, either the incoherence after all implies a failure to respond correctly to reasons, or else the incoherence is rational.[19] In the present context, the objector would have to argue that these are blameless cases of irrationality$_{RESP}$. This seems plausible if we indeed respond correctly to our reasons in these cases of incoherence.

Before turning to some potential examples, I wish to mention a general worry with blameless irrationality$_{RESP}$. According to the current objection to the argument from responsibility, we are sometimes blameless for irrationality$_{RESP}$ because we didn't violate decisive normative reasons: it's not the case that one ought to be rational. However, the following statement is puzzling: 'You are irrational, but this is totally fine. It is fine not because you are excused or exempted for your irrationality, but because you should not even be rational in this case.' Here the charge of irrationality is supposed to lack its characteristic sting. But it is puzzling how it can lack that sting given that you are neither excused nor exempted from your irrationality. The incoherence at issue would be *completely innocuous*. I do not think that these cases are very interesting if we wish to think about normative questions. I thus think that, whenever you are blameless for your irrationality, where irrationality is understood as a *serious* failure that *can* make you criticizable or blameworthy, then you are either exempted or excused. The only reason why one could be blameless for one's irrationality$_{RESP}$ is an excuse. In cases of excuse, one still fails to respond correctly to reasons. I thus endorse the picture presented in Figure 3.1 below.

Note that the picture captures the idea that blamelessness for irrationality is either due to an excuse or an exemption. In cases of exemption, the irrationality falls outside the scope of direct reasons-responsiveness, and thus, no normative requirement to be rational makes sense. In cases of excuse, the person violates a rational requirement, and is thus irrational, but the person is blameless due to an excuse. Finally, if the person is responsible

for being irrational and is not excused, then the person is blameworthy. This picture allows us to capture the characteristic sting of irrationality ascriptions while still allowing for two ways in which we can be blameless for being irrational: exemptions and excuses. I'll briefly elaborate on both notions.

If a person is exempted, then they're *not responsible* for their norm violation. In the doxastic realm, the most obvious exemption is severe pathology, where a belief lacks any responsivity to counterevidence. For instance, I might believe that everybody is constantly watching me. Although I might then, under certain conditions, be blameworthy for not getting therapy, the belief itself is not rationally evaluable due to its complete unresponsiveness to counterevidence. In this case, I am an unreliable informant about the activities of people around me. But treating me as unreliable will not amount to epistemic *blame* due to my exemption.[20]

Excuses, on the other hand, make room for cases in which a person violates an epistemic norm, is responsible for violating it, and is yet blameless (due to an excuse). In epistemology, *externalists* about epistemic justification can most obviously make room for epistemic excuses (see Littlejohn forthcoming). For externalists, epistemic excuses are cases in which your evidence is misleading (say, because you are deceived by an evil demon), but where you are epistemically rational. According to externalists, you are *not* epistemically justified in holding the belief if, say, it was unbeknownst to you formed by an unreliable process. However, you are still *epistemically blameless* because you're ignorant of your unfortunate epistemic situation. Moreover, by being epistemically rational you still manifest a kind of virtue (Wedgwood 2017, 2023), or a knowledge-conducive disposition (Lasonen-Aarnio forthcoming).[21] Yet some recent suggestions of what counts as an epistemic excuse are compatible with *internalism* about epistemic justification. For instance, Alex Worsnip (2021: 162–164) suggests that epistemic irrationality (which is essentially a perspectivist notion) is excusable in cases of *cognitive overload* (when the amount of evidence cannot be processed) and when one has *practical justification not to revise one's credences* (say, because the house is burning right now).[22]

The objection I consider here argues that there is a mistake in the picture just sketched (and illustrated in Figure 3.1): sometimes, we are blameless for irrationality even though we are responsible for it, and the reason for this is not an excuse. By contrast, according to my view, all blameless irrationality for which we are responsible can be explained by excuses. I have already pointed out that the opponent has trouble capturing the characteristic sting of irrationality – any case of blameless irrationality$_{\text{RESP}}$ without excuse would be *harmless*. However, let's now try to get some potential cases of blameless irrationality$_{\text{RESP}}$ into view. This will allow me to illustrate how our intuition that the person is irrational fades in such cases.

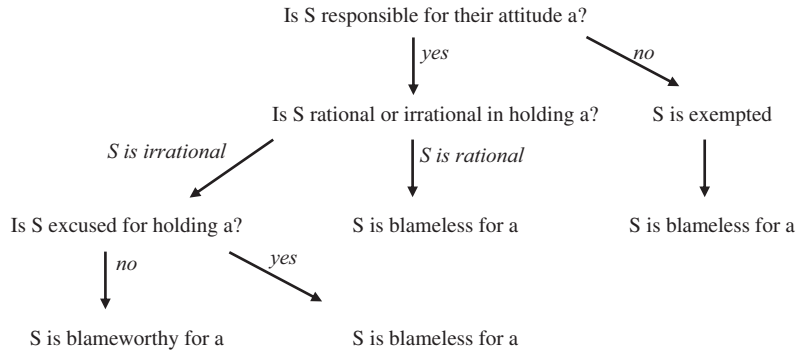

Figure 3.1 Rationality and blame

I focus on the most challenging examples for reasons-responsivist accounts of rationality: cases of higher-order defeat (see Lord 2018: 55–61).

For starters, consider the 'preface paradox'. If I was asked for each of my beliefs whether I think it is true, I might consider the evidence I have for each belief and rationally conclude that each belief is true; but if I was asked about whether I think that *all* of my beliefs are true, the only rational answer seems to be 'no'. I believe p, believe q, believe r, … believe n, but I do not believe that (p, q, r, … n). This is incoherent, and thus could seem irrational. Yet I also seem to be blameless.

However, I am *not* irrational. Rather, the knowledge of my own fallibility defeats my first-order evidential reason for believing (p, q, r, … n). That is, I lack sufficient reason to believe this conjunction. My fallibility gives me a reason for doubting the conjunction, but it does not give me decisive reason against believing each individual conjunct when considering my reasons for believing each one separately. Therefore, there are rational cases of doxastic incoherence. This is because I am rational in holding incoherent doxastic attitudes in some cases of higher-order defeat. So although I am blameless in such cases for my incoherence, these are not cases of blameless *irrationality*.

Other cases of higher-order defeat are more challenging. Consider cases in which you have (misleading) evidence about what your first-order evidence justifies. For example, suppose your scientific supervisor is a renowned expert about p-related issues. You and your supervisor together inquire into whether p, but you fail to gather evidence that would be sufficient for justifying the belief that p. However, your epistemically superior supervisor then tells you – to your surprise and mistakenly – that you *did* uncover sufficient evidence for p. It seems that, given your supervisor's testimony and their epistemic authority, you should believe that your evidence sufficiently supports p. Yet at the same time, given only your first-order evidence, you

should not believe p. It seems that, given your overall epistemic reasons, you should not believe p (you lack sufficient evidence) *and* believe that your epistemic reasons are sufficient for believing p (this is what your supervisor's testimony supports). If you believe this, then you seem incoherent, and irrational. Yet you seem to respond correctly to your reasons, and thus you're blameless. There is no reason to suppose that this is a case of pathology, and so you are still responsible for your irrationality. Thus, this seems to be a case of blameless irrationality$_{\text{RESP}}$ that does not consist in a failure to respond correctly to reasons, but merely in being incoherent.

Such cases are controversial. In the present case, it is unclear whether believing your supervisor is the correct response to your reasons. Maybe you should instead suspend judgment until you understand your supervisor's reasons for telling you that the evidence sufficiently supports p. If the supervisor cannot provide you with reasons for believing this about your evidence, then you might be rational in not believing what they tell you, despite their general epistemic authority. However, if the reason why you lack sufficient first-order evidence for believing p is merely *that it is difficult for you to evaluate your first-order evidence due to your own lack of expertise*, then the testimony of your epistemic superior might support believing that you have sufficient evidence for p *and* believing p. In any case, your reasons do not require you to have an incoherent set of beliefs (see Lord 2018: 59–60).

This brief discussion provides us with a recipe against counterexamples to the argument from responsibility. In any case in which you seem to be responsible for irrationality while being blameless (although you're not excused), you either fail to respond correctly to your reasons after all, and so you're after all blameworthy for your incoherent attitudes, or else you are in fact *rational*, because there are some rational incoherences, as in the preface paradox.[23]

While I cannot argue that all counterexamples end up in one of these two horns, I have outlined a general dilemma for objections to (4). The objection was that there are some cases of blameless irrationality$_{\text{RESP}}$ which aren't failures of reasons-responsiveness. I have suggested that such cases either count as failures of reasons-responsiveness or else that they are not plausibly cases of irrationality$_{\text{RESP}}$. As long as we hold fixed the intuition that irrationality is a *serious* failing that deserves criticism absent excuse or exemption, then there are no cases of blameless irrationality$_{\text{RESP}}$ that cannot be explained by excuses. Instead, all inexcusable irrationality$_{\text{RESP}}$ implies blameworthiness.

3.6 The Coherentist's Replies

Defenders of rationality as coherence à la Broome could react to the argument from responsibility in two ways. First, they could argue that their

theory does not focus on rationality$_{RESP}$, but rather includes cases of irrationality for which we are not responsible, like severely pathological phobias or severely delusional beliefs that are incoherent, as well as more mundane cases of incoherence that aren't very serious failures deserving of blame or personal criticism. However, then their dispute with theories of rationality as reasons-responsiveness would turn out to be only apparent. This would reveal that the two kinds of theories just focus on properties that are not co-extensive, thus talking past each other.

In a sense, reasons-responsivists can of course allow for severely pathological cases of irrationality: if you're stuck with a belief even though your evidence speaks clearly against it, there is an obvious sense in which you don't respond correctly to your evidential reasons, and we might call this irrational. My point here is that it wouldn't be *your failure* that you don't respond to your reasons – you wouldn't be responsible for being stuck with this belief that isn't supported by your evidence. Importantly for the present context, your irrationality wouldn't be blameworthy, since you aren't even responsible for this irrationality: there is no normative requirement here to be rational. This is why I prefer the term 'arational' for such cases. Since reasons-responsivists defend the *categorical* normativity of rationality, they must exclude severe pathologies (see Chapter 3.3, and Kiesewetter 2017: 100): it doesn't make sense to normatively expect the believer who is unresponsive to evidence to revise their unsupported belief, and to blame them if they don't live up to such an unwarranted expectation.

Alternatively, coherentists à la Broome could restrict their theory to (in) coherences for which we are responsible, and which deserve blame absent excuse—that is, to rationality$_{RESP}$. However, given the argument from responsibility, this would commit them to the view that whenever we are incoherent (in the relevant sense), we fail to respond correctly to our reasons. Of course, they could still deny the other conditional: we might sometimes be coherent but fail to respond correctly to our reasons (as when we fail to take up easily available perceptual evidence but nevertheless hold a coherent set of beliefs). I'll return to this view briefly in the next subchapter, and point out that it rests on some idea of 'the Given'.

Finally, coherentists could deny that we ought to respond only to the reasons that we *possess* or that are *available* to us. They would then commit to objectivism about reasons (see Chapter 3.1), thereby implying that normativity and substantive rationality come apart, since the latter supervenes on the mental while the former doesn't (Broome 2020). However, if this was the *only* significant difference between rationality as responding correctly to reasons and rationality as mental coherence, then we can, at least for present purposes, bypass the dispute by focusing on cases in which the reasons that are relevant for what we ought to do are *possessed* or *available*. And of course, theorists of substantive rationality could also deny the normativity of substantive rationality by opting for objectivism about reasons.

In sum, if the argument from responsibility is sound, then theories of rationality as coherence and theories of rationality as reasons-responsiveness either (a) do not talk about the same property (since the latter mean to focus on rationality$_{RESP}$) or (b) their dispute comes down to a dispute between subjectivism or objectivism about reasons. As I have argued, we should focus on rationality$_{RESP}$ when we are interested in rational *requirements*: rationality does not require anything if we are not directly responsible for complying with rational requirements. I thereby rule out the kind of property that some coherentists might be interested in, who can, given *their* focus, plausibly deny that we are responsible for (ir)rationality in *their* sense. This sense might include severely pathological cases and cases of harmless incoherence. Furthermore, we can, for present purposes, bypass the dispute between objectivism and subjectivism about reasons by focusing on cases where the relevant reasons are possessed by the subjects.[24]

3.7 The Disputes about Rationality

Up to now, I have considered three possible ways of contrasting theories of rationality as coherence with theories of rationality as reasons-responsiveness: first, coherentists can be concerned with *all* kinds of (ir)rationality, including severely pathological forms of (ir)rationality and cases of harmless incoherence, while reasons-responsivists *must* restrict themselves to instances of serious (ir)rationality for which we are directly responsible (to make sense of *normative requirements* to be rational); second, coherentists could be objectivists about reasons (and subjectivists about rationality), while reasons-responsivists must be subjectivists about reasons *if* they wish to capture the ideas that rationality supervenes on the mental *and* that it has normative authority (see Chapter 3.1); third, coherentists must argue that some irrational incoherences are not just reducible to failures to respond correctly to reasons, while reasons-responsivists must reduce all irrational$_{RESP}$ incoherences to failures to respond to reasons. Is there another way of contrasting these two positions in the current debate?

We might understand these positions as having different views about the 'Myth of the Given' (Sellars 1956). Reasons-responsivist accounts tend to understand reasons as facts in the external world that are accessible to the subject, and which are thus, as Kiesewetter (2017: 173) puts it, part of the subject's "total phenomenal state": the facts must be perceived, introspected, remembered by the subject, or be the content of a subject's *seeming* or the subject's *intuition* in order to be relevant for the subject's rationality. However, one might worry that, if these facts are perceived or remembered *but not believed*, then it is unclear how they can justify belief. It might seem that facts can serve as reasons for belief only if one believes

these facts. What renders our beliefs rational are then just our other beliefs about facts, even according to the reasons-responsivist account of rationality. But this would make reasons-responsivism collapse into coherentism.[25] To avoid this collapse, reasons-responsivists must maintain that facts accessible in perception are already conceptually structured, and that this enables them to justify and rationalize beliefs (McDowell 1994; Siegel 2017). That is, reasons-responsivists must accept some version of 'the Given' as normative foundation for our beliefs. So the debate about rationality might remind us of the classical epistemological dispute between foundationalism and coherentism about the structure of epistemic justification (Bonjour 1999).[26]

This brief discussion provides us with a *pluralist diagnosis* of the dispute between reasons-responsivism and coherentism about rationality. It shows us that the contrast between reasons-responsivist and coherentist accounts of rationality is far from straightforward, and that we can attempt to draw it by appealing to very different kinds of philosophical questions that hide in the background:

a Should we count pathological attitudes and harmless cases of incoherence as irrational (or rather as *arational* or, respectively, *merely incoherent*)?
b Are only available or possessed reasons relevant for what we ought to believe, desire, feel, and intend (*subjectivism* versus *objectivism*)?
c Are there cases of irrational$_{\text{RESP}}$ incoherence that are not reducible to failures to respond correctly to one's available or possessed reasons (*coherentism* or *dualism* versus *reasons-responsivism*)?
d Can only facts that are believed be relevant for an attitude's rationality (*traditional coherentism* versus *traditional foundationalism*)?

Concerning (a), the coherentist can answer 'yes' if they do not want to focus only on rationality$_{\text{RESP}}$. However, it follows trivially from this position that one ought not always be rational, because severely pathological attitudes are not subject to any normative requirements, and cases of harmless incoherence aren't in any way criticizable or blameworthy. This would simply avoid any substantial debate with the reasons-responsivist. Concerning (b), reasons-responsivists must say 'yes' insofar as they want to preserve *both* the intuition that rationality supervenes on the mental *and* that rationality is normative. They must commit to subjectivism about reasons. Here the dispute with the coherentist comes down to a debate about whether reasons are objective or subjective, in the relevant sense. Concerning (c), the reasons-responsivist must argue that any irrational$_{\text{RESP}}$ incoherence implies either a failure to respond to reasons, or is not a case of irrationality$_{\text{RESP}}$, so that some incoherences are rational$_{\text{RESP}}$. Alternatively,

if one endorses that there is a structural kind of rationality, then one has to at least argue that this structural kind of rationality matters in its own way and isn't reducible to responding correctly to reasons (Worsnip 2021). Concerning (d), the reasons-responsivist seems to be committed to answering 'no', and thus to a version of foundationalism. For otherwise the position would just collapse into coherentism. Here the dispute comes down to the old debate about 'the Given'.

This reveals that many philosophical issues are lurking in the background of the current discussions about rationality – issues which should be kept apart in order to clarify the nature of the dispute between coherentism and reasons-responsivism. This dispute can have its sources in quite different philosophical disagreements, and so it's questionable whether there is just one philosophical debate about rationality. Rather, there are several debates about different issues, which we would do well to keep apart.

For the present purposes, it is enough that I have justified the restriction of my use of '(ir)rational' to cases in which we are responsible for (ir)rationality. This restriction implies that any irrationality *worthy of the name* implies a failure to respond correctly to reasons.

Crucially, the argument from responsibility assumes that our responsibility for complying with rational requirements is grounded in reasons-responsiveness – which was premise (2) of the argument from responsibility. The remainder of the book supplements this argument by defending the view that the relevant capacity C in principle NC when the relevant norms N are rational requirements is indeed our capacity to respond to reasons of the right kind. Trivially, what grounds our *direct* responsibility for rationality is not our ability to control our mind *indirectly* by means of our actions. For this capacity merely explains why we are sometimes responsible for managing our mental life by responding to wrong-kind reasons. It does not explain our direct responsibility to rational requirements, and thus does not ground our direct responsibility for rationality. The main question I pose is therefore whether we are ever directly responsible for responding to our right-kind reasons for attitudes. I defend the view that we are.

I will remain neutral about questions (b) and (d). If the answer to (b) was 'no' – if objectivism about normative reasons and 'ought' was true – then rationality would be normative only if the relevant reasons are available or possessed by the subject (see Chapter 3.1). I am fine with this. If the answer to (d) was 'yes' – if 'the Given' is indeed a myth – then there is no substantial difference between my version of reasons-responsivism and coherentism. For then a subject's reasons would render a belief rational only if these reasons are believed by the subject. Yet both views on rationality would imply that rationality is normative, at least when the relevant reasons are believed. I can thus remain non-committal about (b) and (d).

However, I commit to a view about (c), since I claim that all incoherences that are irrational$_{\text{RESP}}$ imply a failure to respond to reasons. I have discussed how we can defend this view by appealing to the idea that any blameless irrationality$_{\text{RESP}}$ lacks the characteristic sting of our irrationality ascriptions. If the reader isn't fully convinced by my argument, then the rest of the book can be read as defending the normativity of *substantive* rationality against a new challenge, thereby further developing reasons-responsivist accounts in the light of recent debates.

Furthermore, insofar as we are interested in the normativity of coherence, we should also be interested in the normativity of right-kind reasons. This is because if coherence is to be normative, it must sometimes provide us with right-kind reasons for certain responses. For instance, Worsnip (2021) argues that coherence provides is with right-kind reasons for structuring our deliberations, and according to Eva Schmidt (2023), incoherence provides us with right-kind reasons for suspending judgment. If right-kind reasons weren't genuinely normative reasons, then their arguments won't do to defend the normativity of (in)coherence. Any theorist of rationality will thus have an interest in the normativity of right-kind reasons, for any defense of the normativity of rationality must appeal to the idea that rationality provides us with right-kind reasons. Since the remainder of this book will be about the normativity of right-kind reasons, it will also be of interest to anyone interested in the normativity of coherence.[27]

3.8 Summary and Outlook

This chapter has presented an argument that irrationality, at least in the cases where we are responsible for being (ir)rational, implies a failure to respond correctly to one's available or possessed right-kind reasons for attitudes. It follows that if you respond correctly to these reasons, then you're rational. This was the argument from responsibility. One premise is that we are responsible for (ir)rationality in virtue of our capacity to respond to reasons. The remainder of the book will defend this premise: reasons-responsiveness can ground a *direct* responsibility to rational requirements. I argue for this premise by showing that various blaming responses can be appropriate when someone fails to respond correctly to right-kind reasons. The appropriateness of blaming responses implies that one is responsible for one's failure of not responding correctly to one's right-kind reasons. Therefore, our ability of reasons-responsiveness is a plausible ground for direct responsibility for attitudes. This justifies premise (2) of the argument from responsibility. The argument therefore allows us to conclude that irrationality$_{\text{RESP}}$ is a *serious failure*: it implies the violation of a genuinely normative requirement to be rational; for failing to respond correctly to right-kind reasons makes one blameworthy absent excuse.

In the next chapter, we will dive into epistemology and consider recent doubts about the normativity of epistemic rationality – doubts coming from views that argue that epistemic reasons are 'not genuinely normative'. Such doubts turn out to be the main motivation for Indirect Voluntarism: if rationality is not normative, because right-kind reasons aren't normative, then we lack our initial motivation for assuming *direct* responsibility for attitudes; all the remaining *indirect* attitudinal responsibility could plausibly be derived from responsibility for actions and omissions, which are clearly subject to normative requirements (i.e., they are subject at least to those of prudence and morality).

We can now see how the problem of mental responsibility breaks down to the debate about whether rationality is normative. If rationality weren't normative because right-kind reasons weren't normative, then Indirect Voluntarism would be plausible. Our intuition that we are responsible to rational requirements could no longer be motivated by the thought that we're blameworthy merely for being irrational, since then failing to respond correctly to right-kind reasons wouldn't be a mistake worthy of blame or serious criticism. By contrast, if rationality is normative because right-kind reasons are normative, then Indirect Voluntarism must be false: we could be legitimately blamed for irrationality insofar as irrationality implies a failure to respond correctly to right-kind reasons. We could leave it open whether this blameworthiness presupposes a genuine kind of direct non-voluntary control over attitudes, since exercising reasons-responsiveness might not be exercising a genuine kind of control.

I conclude the first part of this book with the claim that it is the normativity of rational requirements, and so the normativity of right-kind reasons, that creates the problem of mental responsibility. To develop an answer to this problem, we must first understand why the normativity right-kind reasons could become questionable. The next chapter works out this new challenge for the normativity of rationality, focusing on epistemic rationality. The task of this book is to understand this specific kind of responsibility to rationality, which will require us to better understand the normativity of right-kind reasons for attitudes.

Notes

1 The wide-scope versions of the standards mentioned above would be 'you ought to [not believe that you have sufficient evidence for p *or* believe that p]', 'you ought to [not believe that you ought to φ *or* intend to φ]', and 'you ought to [not believe that you ought to φ *or* not believe that ψing is a necessary means to φing *or* intend to ψ]' (cf. Kiesewetter 2017: 88). What is peculiar about these standards is that they can be satisfied in more than one way – i.e., by giving up or adopting one of the attitudes mentioned in the standard.

2 Another problem for the normativity of coherence is what the reason *is* that such structural requirements provide us with (Kolodny 2005: 547–551; Kiesewetter 2017: chapter 5): is coherence *non-derivatively* normative, or does it derive its normativity from being conducive to some other value?
3 The distinction between propositional and doxastic rationality shows that I can believe what my evidence supports while being blameworthy because of the way I formed the belief: that is, it seems to open the space for cases of *blameworthy norm compliance*. Such cases might give rise to the worry that blameworthiness and norm-compliance must be detached radically from one another (see Lasonen-Aarnio forthcoming: chapter 6; cf. also Luvisotto 2022). My definition of "blameworthiness" in Chapter 2.4, which states that norm violation is a necessary condition on blameworthiness, would be in trouble. However, we can maintain the connection between blameworthiness and norm violations by focusing on *doxastic* rationality, and more generally by considering responding *correctly* to reasons for actions and attitudes as the central deliberative norm: if we focus on norms of rationality in *this* sense, then there are no cases of blameworthy norm compliance. To take up Lasonen-Aarnio's terminology, we can say that a person who responds correctly to their reasons thereby *manifests success-conducive dispositions*.
4 See Kiesewetter (2017: chapter 8) and Lord (2018: chapter 8). See Broome (2020) for an argument that normativity is not about available or possessed reasons, while rationality is. See Kiesewetter (2020) for a reply that normativity is also concerned with possessed or available reasons, just as rationality.
5 Cf. "*substantive rationality requires structural rationality*" (Wedgwood 2023: 76, italics in original). For Wedgwood, this is because rationality is a matter of holistic coherence with "the Given": if you hold incoherent attitudes, then your overall mental state (including also your non-attitudinal perceptions, memories, certain beliefs you held in the past, etc.) won't be perfectly coherent, and so you won't be ideally substantively rational. The argument I develop reveals *another* sense in which substantive rationality requires structural rationality: insofar as structural irrationality is taken to imply serious criticizability (absent excuse), it must also involves a failure of reasons-responsiveness.
6 The label has its origin in a specific debate about fitting-attitude accounts of value (cf. Rabinowicz and Rønnow-Rasmussen 2004). See Gertken and Kiesewetter (2017) for a detailed discussion of the distinction.
7 Recent discussions question whether epistemic rationality is distinctive of belief by arguing that also actions are evaluable in terms of epistemic rationality (see Flores and Woodard 2023; but see Arpaly 2023 for pushback). I do not deny this. Importantly, however, evaluating attitudes such as desire or intention and *most* actions doesn't primarily happen in terms of *epistemic* rationality. At the very least, epistemic rationality is distinctive of belief and maybe some distinctively *intellectual* actions, such as assertion and inquiry. What I think we should avoid is that all kinds of actions become subject to epistemic rationality just because they can downstream impair one's epistemic performance. Boozing or not getting enough sleep can be practically, but not epistemically, irrational.
8 Schroeder (2021) calls these two features 'earmarks' of right-kind reasons, thereby avoiding a commitment to necessary conditions. This is partly because one might worry that one cannot respond to all right-kind reasons, such as in the famous surprise-party cases (Schroeder 2007). However, as I suggest in Schmidt (2022), we should accept that something is a normative reason for a

person only if this person can, in one way or another, respond to that reason (I call this 'the weakest possible version of the motivational constraint on reasons'). For otherwise we blur the line between the evaluative (what is good) and the normative (what we have reason to do, believe, etc.).

9 See Wedgwood (2002) on belief 'aiming at truth'.

10 The reason to drink the toxin in Kavka's toxin puzzle (1983) is state-given in this sense.

11 Pragmatists will disagree about *when* the value of a belief provides a practical reason for this belief. Moderate pragmatist views, as presented by McCormick (2015), would argue that it is not a reason for this belief if all your evidence speaks against the existence of guardian angels. More radical pragmatists, like Rinard (2015, 2017, 2022), argue that every consideration indicating the value of an attitude is a practical reason for this attitude. Reisner (2008) and Howard (2020) think that evidence becomes irrelevant when the belief is sufficiently important.

12 Cf. esp. Shah (2006) and Hieronymi (2006). According to these positions, state-given considerations could still be evidence for evaluative claims about mental states and thus provide us with epistemic reasons, for instance, *to believe that a certain mental state M is valuable*. But they are not thereby reasons *for M*, but rather to *believe* that M is valuable, or to *desire* M, or to *bring about* M, etc.

13 Cf. Leary (2017: 537–540), McCormick (2018: 641), Reisner (2009: 269–270, 2018: 722), and Rinard (2015, 2019a: 1939–1944, 2019b: 775). Vahid (2022) has recently objected to a related strategy by the new pragmatists which claims that beliefs are merely *caused* by evidence but not *based* on evidence, so that the only proper bases for belief turn out to be practical reasons. I discuss this strategy in Schmidt (2022: 1799–1801), where I point out that it *presupposes* the indirect control strategy mentioned above: if the latter strategy fails, then the former fails as well.

14 For related recent diagnoses of the dialectic between evidentialism and pragmatism about reasons for belief, see Arpaly (2023) as well as Kelly and Cohen (2024).

15 Kolodny (2005: 551) pointed out that facts about an attitude's coherence would be state-given reasons and thus wrong-kind reason for the attitude. He took this to be an argument against the existence of (normative) coherence requirements. Contrast Eva Schmidt's (2023) proposal that incoherence is a state-given yet right-kind reason for suspension, and Worsnip's (2021) proposal that coherence is a right-kind reason for structuring one's deliberations in certain ways.

16 My initial formulation of this principle in Schmidt (2020b) mentioned a 'successful' exercise of C instead of a correct exercise. This formulation was misleading. You might exercise a capacity successfully but in a way that is incorrect, which could make you blameworthy. For instance, you might be said to exercise your capacity for moral agency *successfully* by intending to break a promise and then successfully breaking the promise, which would make you blameworthy. Although successful, this is not a *correct* exercise of your capacity for moral agency, since the correctness of exercising moral agency is determined by moral requirements.

17 However, see Chapter 3.7, where I point out that the position that "the Given" is a myth – i.e., the view that non-attitudinal states like perceptions cannot provide reasons if you don't *believe* their contents – would collapse the distinction between structural and substantive rationality.

18 Ernst (2020) argues that one is irrational if one doesn't respond correctly to beliefs about normative facts, e.g., when one fails to believe that there's a reason against performing an action while also believing that this action causes pain. This is not a common use of the term 'irrational' in the current debate. For the subject might not be in a position to know about the connection between the pain and the normative fact that this pain is a reason not to perform an action. For instance, they might have been indoctrinated so that they believe that the pain of certain individuals doesn't normatively matter. The subject could then be rational both in the sense of coherence and in the sense of responding correctly to all their available or possessed reasons: due to indoctrination, they might not have easy epistemic access to the connection between the pain and the normative fact. If they're nevertheless blameworthy for their failure to believe that there's a reason against the action that causes pain, then Ernst might have to commit to a kind of *externalism about blameworthiness*: even if a person wasn't in a position to know that their action is wrong, they could still be blameworthy for not responding correctly to their beliefs about normative facts. I object against this view in Chapter 6.5. Alternatively, Ernst could hold that human beings are *always in a position to know* that pain is a reason against performing actions that cause it. If that was empirically plausible, then the irrationality Ernst identifies would, as an empirical matter of fact, always be substantively irrational in the sense common to the debate. However, this claim could be subject to empirical refutation. A similar claim has been upheld by Arpaly and Schroeder (2014: 183), who note that a historical slave master always *can* come up with the right conclusion that what he's doing to his slave is wrong. This claim seems empirical to me.

19 See Kiesewetter (2017: chapters 9–10) and Lord (2018: chapter 2) for extensive attempts to reduce coherence requirements of rationality to requirements to respond correctly to reasons. See Worsnip (2021) for an extensive reply.

20 However, I don't wish to deny that pathologies are sometimes reasons-responsive (see Hubacher Haerle 2023), which is why I talk, a bit artificially, about 'severe' pathologies to refer to those who aren't sufficiently reasons-responsive to be properly rationally evaluable.

21 There is a recent debate about whether the category of epistemic blamelessness or epistemic excusability can really help externalists against the 'New Evil Demon Problem', which consists in explaining from an externalist perspective why a subject A who is systematically deceived seems, from an epistemic point of view, to be doing *as well as* another subject B who isn't deceived but is otherwise identical to A, without saying that A and B are equally epistemically justified. For recent externalist replies, see Littlejohn (forthcoming) and Williamson (forthcoming). For more critical voices and further discussions, see Ballarini (2022), Boult (2017), Gerken (2011), Greco (2019), and Madison (2018).

22 See Flores and Woodard (2023: 2558, 2561) for similar and further proposals of epistemic excuses. I deny that a practical justification is an epistemic excuse: one can still be blameworthy when epistemic and practical reasons conflict (see Chapter 5 and Schmidt forthcoming c).

23 For more on rational incoherence, see Field (forthcoming).

24 I suggest in Schmidt (2020b) that issues of responsibility are also central to the dispute about subjectivism and objectivism about reasons, as well as about whether rationality supervenes on the mental. I leave out my argument for this further claim here, which is of no direct relevance to my argument in this chapter.

25 Wedgwood (2017: 12) argues that the distinction between rationality as mental coherence and rationality as responding correctly to reasons is "illusory". However, coherence amongst your attitudes could plausibly differ from coherence amongst your attitudes *and* your possessed reasons (to which you need not take any attitude). See also Worsnip (2021: 11–17) on why Wedgwood's point relies on a notion of coherence that is too broad to capture the idea of structural rationality which is concerned with the former, narrower kind of coherence. In his recent work, Wedgwood (2023) clearly commits to the view that non-doxastic mental states affect the rationality of belief, and so explicitly rejects pure coherentism. His broader conception of 'coherence' is roughly equal with the notion of responding correctly to reasons (which go beyond the reasons that are *believed*). He has by now revised his view that the distinction is "illusory" (see Wedgwood 2023: 75–76). Yet Wedgwood still emphasizes that the distinction is misleading since substantive rationality also concerns the *coherence* of one's overall mental states, including non-doxastic ones, and so is a kind of coherence in a broader sense. I agree.
26 For some explicit connection with this debate see Daoust (forthcoming).
27 It could in principle be argued that there is always a *wrong-kind* reason to be coherent. But this is implausible because wrong-kind reasons paradigmatically indicate the *practical value* of an attitude—and holding coherent attitudes isn't always practically valuable.

Part 2

The Normativity of Epistemic Rationality

4 A Neglected Challenge for the Normativity of Epistemic Rationality

> Faced with a choice between epistemic indiscretion and personal or public disaster, […] we find it natural to side with prudence. The intellectual saint blind to the consequences of an intemperate regard for epistemic virtue is a comic figure.
>
> John Heil, "Believing Reasonably" (1992: 48)

The first part of this book has argued that the problem of mental responsibility comes down to a problem about the normativity of rationality. The problem is not how we can be responsible for our attitudes given that we cannot choose our attitudes. For this problem can easily be solved by appealing to indirect control over attitudes. Instead, the problem is how we can be *directly* responsible for our attitudes, given that we are sometimes irrational merely in virtue of *holding* certain attitudes. As we saw in the last chapters, the charge of irrationality intuitively implies a serious form of personal criticism or blame. And yet irrationality ascriptions are often independent of how a person has *managed* her mental life. Instead, they imply that she did not respond correctly to her reasons for attitudes. It seems that this already implies the normativity of rationality. For surely responding correctly to reasons is what we ought to do.

However, the normativity of rationality does not immediately follow from a conception of rationality as reasons-responsiveness. Recently, there have been doubts about whether reasons of the right kind are genuinely normative reasons.[1] If right-kind reasons are not genuinely normative reasons insofar as they lack the characteristic *normative authority* associated with such reasons, then rationality, understood as correctly responding to right-kind reasons, cannot get any normative authority from right-kind reasons. Therefore, a neglected challenge for the normativity of rationality arises from recent skepticism about the normativity of right-kind reasons.

In this chapter, I will focus on skepticism about the normativity of *epistemic* reasons, which comes from pragmatist and instrumentalist accounts of reasons for belief. Susanna Rinard (2015: 219), for instance, argues

DOI: 10.4324/9781003382973-6

"that only pragmatic considerations are genuine reasons for belief. That is, purely evidential considerations – evidential considerations that are not also pragmatic reasons – do not constitute reasons for belief". Asbjørn Steglich-Petersen and Mattias Skipper argue "that evidence for *p* speaks in favor of believing *p* only in contexts where there is a practical reason to pursue the aim of coming to a true belief as to whether *p*" (2019: 9), and that therefore "it is strictly speaking false to say that evidence *by itself* constitutes a normative reason for belief" (2020: 114). Similarly, Barry Maguire and Jack Woods (2020) have compared epistemic requirements with rules of games: we only have a reason to comply with each if we have a practical (prudential or moral) reason to engage in the relevant practice. That is, I have a reason to move a chess piece according to the rules only if I have a practical reason to play chess; analogously, they argue that I have a reason to believe that p only if I have a practical reason to play what Maguire and Woods call "the game of belief".[2]

I assume here that purely evidential considerations – that is, considerations indicating the truth of a proposition – provide us with epistemic reasons. I thus focus on object-given reasons for belief, that is, those reasons bearing on the truth of a proposition, which are clear-cut reasons of the right kind for belief. One might think that epistemic reasons are not just object-given, but that state-given reasons concerning the value of holding or not holding a belief are also relevant to epistemic rationality (see Chapter 3.2). However, I will not rely on the possibility of state-given reasons of the right kind. Instead, I will argue (in Chapter 5) that purely evidential considerations have normative significance by themselves, even in cases where there is no state-given reason for holding a belief, and even in cases where reasons of the wrong kind seem to pull in another direction. This strategy allows me to defend the normativity of rationality without involving controversial claims about state-given reasons being relevant to epistemic rationality. My hope is that this will be more convincing *dialectically*, that is, within the context of the recent debates on reasons and rationality.

First, however, the task is to understand the challenge that is posed to the normativity of rationality, and thereby to the idea of direct responsibility for holding (ir)rational attitudes, by the epistemologists quoted above, who doubt the normativity of epistemic reasons. The problem cases for the normativity of epistemic rationality, understood as implying responding correctly to epistemic reasons, are twofold: cases of *trivial belief* and cases of *epistemic-practical conflict*. These cases give rise to a neglected problem for the normativity of epistemic rationality: that epistemic reasons might fail to be genuinely normative reasons. This challenge has been largely ignored in the literature on rationality. Furthermore, it has been overlooked that this challenge bears on the discussion about responsibility for belief.

I begin by introducing the challenge arising from trivial belief by means of the familiar clutter avoidance problem (Chapters 4.1 and 4.2), before turning to the challenge arising from normative conflict cases (Chapters 4.3 and 4.4). I will then carve out a common denominator of both challenges: the possibility of epistemic blame (Chapter 4.5). I then contextualize the neglected challenge within the current debate on the normativity of rationality (Chapter 4.6). Finally, I summarize (Chapter 4.7). Chapter 5 then replies to the challenge by defending the possibility of epistemic blame. This will also allow us to refute Indirect Voluntarism by establishing that we are directly responsible for holding (ir)rational beliefs.

4.1 Clutter Avoidance

Epistemic rationality requires of us, very roughly, that we believe what we have sufficient evidence for, and that we refrain from believing what we lack sufficient evidence for. This initial formulation leads to a familiar problem. Consider, first, the following requirement:

> (EN) One ought to believe everything that is sufficiently supported by one's evidence.

Gilbert Harman (1986: 12) points out that (EN) implies that we should clutter our minds with uninteresting implications of our beliefs. My current belief-stock implies the proposition that [I am sitting in my office *or* the moon is made of cheese *or* there is no coronavirus *or* there is no human-induced climate change *or* …]. This disjunctive proposition is true right now while I am sitting in my office. It is true because the first claim, that I am sitting in my office, is true. At the same time, it seems that I do not always believe this disjunctive proposition while I am sitting in my office. Most importantly, it seems that I would not be *blameworthy* or *criticizable* in any sense for *not* believing such disjunctive propositions. It follows, so it seems, that there is no unconditional norm to believe everything that is sufficiently supported by my evidence.

One might doubt that I do not believe the disjunctive proposition. For if I am asked whether I believe it, and I understand the content of the proposition, I will reply that I do believe it. However, even if one thinks that we believe all those disjunctive propositions, one will agree that we do not believe all the implications of our current belief-stock, like certain mathematical or logical implications that are just too hard to figure out. That we do not believe those implications does not make us blameworthy or criticizable in any sense. Furthermore, we *can* imagine a case where I fail to believe such weird disjunctive propositions. Why on earth, we might ask, should any reasonable person care about this so much as to regard me as blameworthy for not believing them?

In reaction to this, we might modify the epistemic requirement so that its violation more plausibly gives rise to serious criticism. We might propose background conditions for *when* we are required to believe what our evidence sufficiently supports. These background conditions should fulfill two criteria:

a They must make it plausible that the subject is, at least normally or in paradigm cases of an epistemic norm violation, blameworthy or criticizable for not complying with the epistemic norm when the background conditions are fulfilled.
b They should not render the norm *practical* rather than *epistemic*.

Call (a) the criterion of *significance*, and (b) the criterion of *content*. (b) makes sense as a criterion on epistemic requirements for our purposes because the normativist wants to defend the normativity of a distinctively *epistemic* kind of rationality. But why (a)?

The guiding idea behind (a) is that the significance of a norm expresses itself in the reactive attitudes that we show toward violations of the norm. For instance, the significance of a moral requirement will make it often – in the absence of an excuse or exemption – appropriate to show resentment or indignation. These emotions are expressions of the normative significance we attach to the moral requirement because they are appropriate in the face of its violation. Similarly, if there are distinctively epistemic requirements that provide us with reasons for compliance, then we should expect there to be distinctively epistemic reactive attitudes that we show toward the violation of those epistemic requirements. In this vein, Antti Kauppinen (2018: 3) understands genuine norms (as contrasted with mere evaluative standards) as "rules that someone is accountable for conforming to in suitable circumstances".

Here is an argument for (a). *Why* is it false that we should clutter our minds with all the implications of our beliefs? If we would accept that epistemic rationality requires us to clutter our minds, then we would constantly violate an epistemic requirement by not drawing all the implications from our beliefs. However, this constant violation would have no further significance: we would not normally be blameworthy or criticizable for failing to believe what we epistemically ought to believe. The problem with this is that the normative force of this 'ought' would then be mysterious: why comply with this norm if we cannot be held legitimately responsible for non-compliance? The epistemic requirement would at best have the force of the norms of etiquette or the rules of a game: we can intelligibly ask *why* we have a reason to comply with the norms of etiquette or rules of a game in a given situation. Such norms do not, by themselves, provide us with reasons. Thus, the best explanation of the intuitive appeal

of Harman's clutter-objection when it comes to trivial implications of our beliefs is that we assume that epistemic norms with normative significance would fulfill (a).[3]

I will return to the connection between epistemic reasons and epistemic blameworthiness in Chapter 5. For now, consider another strategy for finding a plausible requirement of epistemic rationality. Instead of proposing background conditions to (EN), we might rather argue that (EN) is not a central epistemic requirement at all. In response to Harman's clutter-objection, we might argue that, although we are never blameworthy merely for failing to believe what our evidence sufficiently supports, there are *other* epistemic requirements that are purely evidential. Specifically, we might defend the following requirement of epistemic rationality:

> (EN*) One ought not to believe what is not sufficiently supported by one's evidence.

(EN*) is not confronted with Harman's clutter-objection: rather than requiring us to believe plenty of propositions we intuitively have no reason to believe, (EN*) merely *prohibits* us from having certain beliefs. Steglich-Petersen (2018) accepts (EN*) but denies (EN): he thinks that evidence alone determines the *permissibility* of belief (which beliefs I am epistemically allowed to have), but he argues that evidence alone never gives us, as he puts it, "positive reason" to believe a certain proposition. Epistemic rationality, on this picture, determines the space of epistemic permissibility, but it never requires a specific belief – rather, it merely prohibits certain beliefs (see also Whiting 2010, 2013).

However, appealing to a norm of permissibility like (EN*) instead of (EN) won't help to defend the normativity of epistemic rationality against the challenge I spell out here. First, it seems that (EN*) should not be any more plausible to skeptics about the normativity of epistemic rationality than (EN). The norm that we ought *to believe everything* that is supported by our evidence faces the problem that it requires us to needlessly clutter our minds. The norm that we ought *not to believe anything* that is *not* sufficiently supported by our evidence faces the reverse problem: it would require us not to have a lot of beliefs we, it seems, have no reason to give up.

For example, why should I give up evidentially unsupported but beneficial beliefs? We often overestimate our own abilities or the virtues of our significant others. Arguably, this can promote our self-esteem (see Kelly 2003) or benefit our relationships (see Stroud 2006). It seems that such beliefs are blameless as well. A general requirement not to believe what is insufficiently supported by one's evidence seems *too exclusive*. And a requirement to believe anything that is sufficiently supported by our evidence seems *too inclusive*.

Furthermore, cases of trivial belief pose the same problem for (EN*) as they pose for (EN). What if you believe, in the absence of sufficient evidence, that the celebrity gossip in this unreliable magazine is true? Why should it make sense for anyone to blame or criticize you for having this trivial belief, if we stipulate that your trivial belief will have no bad consequences? Such trivial propositions seem to pose a challenge to (EN*) as they do to (EN) – as I will illustrate in some more detail in Chapter 4.2.

One final clarificatory remark: I will call any form of blame that arises from the violation of a distinctively epistemic requirement – i.e., an evidential requirement that does not mention any practical considerations – *epistemic blame*. That is, epistemic blame, if there is such a thing, is a kind of negative reaction that is appropriate in virtue of violations of requirements of a distinctively epistemic kind of rationality, given suitable non-pragmatic background conditions (and, as we will see later, in the absence of an excuse or an exemption).

I now turn to the idea of background conditions on epistemic requirements in some more detail to spell out a dilemma for the normativity of epistemic rationality.

4.2 A Dilemma

One way of developing a background condition on (EN) that might preserve the normativity of epistemic rationality is presented by Benjamin Kiesewetter (2017: 184–185). He responds to Harman's clutter-objection by proposing that epistemic rationality requires us to believe p if p is sufficiently supported by one's evidence and *if one attends to p*. According to this proposal, if I attend to a specific disjunctive proposition for which I have sufficient evidence, then I would be criticizable (because irrational) if I do not come to believe it. Thus, Kiesewetter concludes, there is a sense in which I *ought* to believe it as soon as I *attend* to it. Analogously, we could propose a background condition on (EN*) by saying that if we lack sufficient evidence for p and we attend to p, we ought not to believe p: we would be criticizable if we were to believe p; but we wouldn't be criticizable for believing p if we never consciously considered p – we wouldn't count as irrational for still believing p.[4]

It seems that Kiesewetter's background condition, while doing a good job in fulfilling criterion (b), does not fulfill (a). There are cases where we attend to a proposition that is sufficiently supported by our evidence but where it would not, it seems, make much sense to regard us as blameworthy or criticizable if we, for whatever reason, do not believe it. Take a case in which I come across the latest celebrity gossip in a magazine that I know to be reliable, but I fail to believe the gossip. Assume again that having a belief about the matter is of no importance and that I do not care about

whether the gossip is true. Again, it seems that there is no obvious sense in which I am blameworthy, and that it is false that I ought to believe the gossip.[5]

It thus seems that if it does not *matter* whether we believe an evidentially well-supported proposition, it is false that we *ought* to believe it. However, if we instead propose a background condition on sufficient evidence that implies that it always matters whether we comply with the epistemic requirement, we seem to end up violating criterion (b): if the requirement is only in place when it matters whether we comply with it, then, so it seems, the requirement is no longer distinctively epistemic. It thus seems that there is no background condition on requirements of epistemic rationality like (EN) and (EN*) that fulfills both (a) and (b). The normativist about epistemic rationality is in a dilemma.

Let us provide this dilemma with additional support by considering a background condition that does not fulfill criterion (b). Steglich-Petersen (2011), after discussing a case of a trivial belief that the subject is not required to have although it is well-supported by the subject's evidence (23), presents the following partial analysis of reasons for belief:

> Necessarily, if S has all-things-considered reason to form a belief about *p*, then [if S has epistemic reason to believe that *p*, S ought to believe that *p*] (24).

Here "epistemic reason" can be read as "sufficient evidence for p". The conditional then states that

> *if one has an all-things-considered reason to form a belief about p*, then one ought to believe what one's evidence sufficiently supports.

The italicized if-clause is Steglich-Petersen's background condition for the epistemic standard (EN). Steglich-Petersen could analogously propose a background condition on (EN*):

> *if one has an all-things-considered reason to form a belief about p*, then if p is *not* sufficiently supported by one's evidence, one ought not to believe p.[6]

Steglich-Petersen's "all-things-considered reason to form a belief about p" can, for instance, be a reason for an action prior to the belief.[7] "Forming a belief about p" might refer to the action of *thinking about whether p*: I may have more or less reason to think about whether something is true. I have some reason to think about whether there will be nice weather during the coming days, but I have no reason at all to think about the latest

celebrity gossip (I might even have reason to avoid such thinking). Thus, according to one plausible reading, the truth of "S ought to believe that p" in Steglich-Petersen's analysis is conditional on a practical reason for an action – for instance, the action of actively thinking about whether p. It says that if we have a reason to *bring it about* or to *maintain* that we have a (true)[8] belief about p, and there is sufficient evidence for p, then we ought to believe that p.

Since we only have a reason to bring a belief about when it *matters* whether we have this belief, Steglich-Petersen's proposal does a good job fulfilling our criterion of *significance*. Yet his proposed background condition, and thus the proposed epistemic requirement, is no longer purely *epistemic*, because it includes a practical reason (for an action). His proposal thus fails to fulfill our criterion of *content*.[9]

Thus, while proposing non-pragmatic background conditions on epistemic requirements (à la Kiesewetter) apparently does not result in requirements that fulfill the criterion of *significance*, proposing a pragmatic background condition (à la Steglich-Petersen) results in requirements that do not fulfill the criterion of *content*. If the normativist about epistemic rationality accepts the criterion of significance for epistemic requirements, then they have to defend the claim that compliance with epistemic requirements matters (in a sense) even if we do not equip these norms with a pragmatic background condition. *Prima facie*, it is difficult to see how distinctively epistemic requirements could matter by themselves. Therefore, any normativist will, it seems, end up facing either of two horns of the dilemma:

i Epistemic requirements are purely epistemic, but they fail to be significant.
ii Epistemic requirements are significant, but they fail to be purely epistemic.

We might be tempted to conclude from the dilemma that we should just reject the idea that we are ever *epistemically* blameworthy and that there is any such thing as a distinctively *epistemic* kind of rationality that has normative significance. Since we understood epistemic requirements as substantial requirements to respond correctly to one's epistemic reasons, this dilemma suggests that we should reject the normativity of epistemic reasons.

I think this dilemma points to a serious challenge for the normativity of epistemic rationality. However, I will ultimately propose that epistemic rationality *has* normative significance: (non-)compliance with distinctively epistemic requirements matters, in a sense. This will require me to make sense of a notion of epistemic blame. For the significance of a norm

expresses itself in our reactive attitudes toward violations of the norm (see Chapter 4.1).

Thus, giving a satisfying reply to this first challenge for the normativity of epistemic rationality requires us, in line with our overall investigation, to consider the concept of blameworthiness and responsibility for beliefs. For now, however, we should accept that there is a good case to be made against the normativity of epistemic rationality, even if epistemic rationality is understood as responding correctly to right-kind reasons for belief. For, as we just saw, these reasons might turn out to lack normative significance – and thus fail to be normative reasons.

I will now, throughout the next two subchapters, turn to a second argument that is implicit in many recent epistemological discussions, and that seems to pull us into the same direction of denying the normativity of a distinctively epistemic kind of rationality. As we will see later in Chapter 4.5, both arguments are an instance of a more general challenge for the normativity of epistemic rationality. In Chapter 4.6, I will show how this neglected challenge connects to the current debate about the normativity of rationality.

4.3 Skepticism about the 'Ought' of Epistemic Rationality

Sometimes, there is practical value in adopting a belief that lacks sufficient support by your epistemic reasons. Suppose, for instance, that your friend would be more confident in a job interview if you were to believe that they are the best candidate, even though you have no clue who the other candidates are.[10] You then have a practical reason to *cause* yourself to believe that your friend is the best candidate, at least if you can. Some would argue that you do not merely have a practical reason to *cause* yourself to believe that p, but also a practical reason to *believe* that p. This arguably assumes that beliefs can be based on practical reasons.[11] Here is a pressing issue that arises independently of this assumption. Let's stipulate that believing p is indeed better than not believing p. Your practical reasons could then require you to believe that p, or at least to cause yourself to believe that p, if you can. However, is there then still a normative sense of 'ought'—the *epistemic* sense—in which you *ought not* to believe that p? This question is my focus. My ultimate aim (in Chapter 5) will be to defend

> *The traditional verdict about epistemic-practical conflicts* (TV). Even when, practically, you ought to (cause yourself to) believe that p, it might still be that, epistemically, you ought not to believe that p. In these cases, there is no answer to what you ought simpliciter to believe.

TV has been the standard position within epistemological theorizing: many authors deemed it essential to retain an epistemic dimension of normativity that is distinct from and largely independent of practical normativity.[12] Recent discussions, however, have questioned this verdict. Effectively, TV tells us that there are cases in which we ought, in a sense, to (cause ourselves to) believe that p and that we ought, in another sense, not to believe that p. However, we might want to have an answer about what we ought to believe *simpliciter*. For it seems unhelpful to say that we ought, in a sense, to respond to our reasons in such a way that we end up believing that p, and that we ought, in another sense, to respond to our reasons in such a way that we end up not believing that p. This intuition against TV has recently been illustrated vividly by Lindsay Crawford in the following passage:

> Suppose that after having been riveted by your recent lecture on Pascal's wager in your Introduction to Philosophy class, your student seeks your professional advice about what she ought to believe. She makes a compelling case that there is good evidence that her roommate dislikes her, but she also makes a compelling case to you that she would be quite a bit better off if she refrained from believing that her roommate dislikes her. So, she asks: "Should I believe my roommate dislikes me, because that's what the evidence suggests? Or should I not believe that she dislikes me, because that would make me feel better?" Having just made the distinction in class between theoretical deliberation about what to believe, and practical deliberation about whether to get yourself to have a belief, you might advise her in the following way: "Well, if you're asking whether you should *believe* that your roommate dislikes you, then yes. That said, you absolutely should do what it takes to *get yourself* not to believe that your roommate dislikes you."
>
> (Crawford 2020: 91)

Crawford goes on to point out that your advice would be impossible to follow, and that the student might well wonder whether it is more important for her to believe that p or rather to cause herself not to believe that p.[13] We might then think that the student should somehow weigh or compare her epistemic reasons with her practical reasons in order to determine how she ought to proceed *simpliciter*. After all, having an adequate conception of her roommate's attitude toward her also seems to be important – maybe it motivates the student to address the issue or to look for another place to live. As the case is described, the student 'would be quite a bit better off if she refrained from believing that her roommate dislikes her'. It therefore seems that the student should follow her practical reasons and ignore her epistemic reasons, because the latter seem to be outweighed by the former (Reisner 2008, forthcoming; Howard 2020), or less important

in comparison to the former (Meylan 2021). John Heil formulates this pragmatist intuition vividly as follows:

> From one natural perspective it seems patent that nonepistemic considerations have priority in such cases. Faced with a choice between epistemic indiscretion and personal or public disaster, for instance, we find it natural to side with prudence. The intellectual saint blind to the consequences of an intemperate regard for epistemic virtue is a comic figure. When that blindness affects the well-being of others, he appears callous or worse […].
>
> (Heil 1992: 48)

Whether epistemic reasons can be weighed against practical reasons is controversial (see Berker 2018; Kauppinen 2023). Surely, comparing the *practical value* in the two relevant scenarios is an intelligible and reasonable thing to do for the student in Crawford's case: she could compare the (expected) practical value that would be realized by complying with one's epistemic reasons with the (expected) practical value that would be realized by not complying with them, and then deciding how to proceed on the basis of whichever option realizes more (expected) practical value. Importantly, however, this would not amount to weighing epistemic reasons against practical reasons. Rather, this is straightforward practical deliberation. It just leaves the epistemic side of things out of the picture. At the very least, those who argue that we can weigh epistemic reasons against practical reasons owe us an account of why their proposed procedure amounts to more than just comparing practical values. I'll argue against the view that there is an 'ought' simpliciter that we can reach by weighing epistemic and practical reasons in Chapter 5.5.

However, there is yet another way for questioning TV. Even if TV was right that there is no answer to what one ought to believe *simpliciter*, opponents of TV could question the other claim that TV makes: that there *is* a normatively significant epistemic 'ought'. Maybe epistemic reasons lack any normative significance when the overall balance of practical reasons requires one to act or believe against them, and thus epistemic reasons do not deliver a *normative* sense of 'ought' at all. This would also explain why practical reasons seem more relevant: it is because epistemic reasons aren't normatively significant by themselves.

4.4 Epistemic-Practical Conflicts

To get to the bottom of this version of skepticism about the epistemic 'ought', it will be helpful to consider the general structure of the cases at issue. First, the cases are such that one must choose between compliance

with one's epistemic reasons and compliance with one's practical reasons – i.e., one will either end up believing that p by complying with the overall balance of one's epistemic reasons (having the evidentially supported belief) or end up not believing that p by complying with the overall balance of one's practical reasons (having the beneficial belief). Furthermore, one has *reasonable means available* to choose either option (say, swallowing a belief-inducing pill). Finally, complying with the practical reasons is *obviously more important*. Let's call cases sharing this structure 'epistemic-practical conflicts.'[14]

Like Crawford above, skeptics about the epistemic 'ought' claim that one ought to comply with the overall balance of one's practical reasons in epistemic-practical conflicts. They think that the epistemic 'ought' is either outweighed or even rendered normatively irrelevant by the practical reasons for violating it. Examples abound in the literature. Mantel (2019) describes a case where one has to choose between having a trivial belief about phone book entries that is well-supported by one's evidence but unimportant, on the one hand, and saving the life of a climber, on the other, and she argues that 'it seems clear that in the example I ought simpliciter to save the climber rather than believing that there are exactly 298,304 entries in the outdated phone book' (2019: 222). Reisner (2008) has argued that the epistemic reasons are just normatively irrelevant when the practical reasons to go against them are sufficiently strong. Meylan (2021) argues that we can compare both sets of reasons, and that practical reasons win out when complying with them matters more. Rinard (2017) even argues that having a belief that is well-supported by one's practical reasons is all-things-considered *rational* even when there are *no* epistemic reasons for this belief, and she now explicitly endorses the view that there is no such thing as a distinctively epistemic kind rationality, so that all normativity is just practical (Rinard 2022). This list of recent epistemologists siding with the practical reasons in epistemic-practical conflicts could go on.[15]

Importantly, all these philosophers appeal to cases in which you cannot comply both with your epistemic reasons and with your practical reasons. Given the structure of epistemic-practical conflicts explained above, the following argument against TV then suggests itself

1 In epistemic-practical conflicts, you *cannot* comply with both your epistemic reasons and your practical reasons.
2 Thus, it is not the case that you *ought* to comply with both sets of reasons.
3 But you ought to comply with your practical reasons.
4 So, it is not the case that you ought to comply with your epistemic reasons.

Note that (4) doesn't straightforwardly imply that we must weigh epistemic and practical reasons against each other to reach an 'ought' simpliciter, or that epistemic reasons aren't normative reasons at all. To derive one of these views, we need an inference to the best explanation (see also the similar argument presented by Mantel 2019: 216). We might discuss which view is the better explanation of (4). Is the epistemic 'ought' insignificant because the epistemic reasons are *outweighed* by practical reasons? Or because epistemic reasons aren't normative in the first place? Rather than deciding what's the better explanation (which would be a task for my opponents), I'll instead show (in Chapter 5) that the argument already fails at an earlier stage.

Let's first clarify the implicit assumptions of the argument. First, the step from (1) to (2) involves a version of 'ought implies can'. I won't take issue with this step here. For I want to argue that there's a distinctively epistemic kind of normativity even if we hold on to the relevant version 'ought implies can'. So it's more advisable for me to take issue with the second step of the argument, which I think involves a more questionable implicit commitment.

In this second step, the argument emphasizes that there is clearly a normatively significant sense of 'ought' in which one ought to comply with one's *practical* reasons in epistemic-practical conflicts. After all, (a) the student in Crawford's case would be 'quite a bit better off' if she didn't believe that her roommate dislikes her, and we can assume (b) that there are *reasonable means available* to her to bring about that she doesn't hold that belief—say, by swallowing pill that induces disbelief, or by engaging in reliable strategies of self-deception that don't take up much of her time and energy. This, in turn, makes it difficult to see how there could still be a normatively significant sense in which the student ought to comply with her *epistemic* reasons. For what's the point of telling her that, although she should *practically* do whatever she can to get herself not to believe that p, she still ought *epistemically* to believe that p? Intuitively, if you ought to get a specific haircut, you're also permitted to have that haircut. Similarly, if you ought to cause yourself not to hold a belief, then you're permitted to not hold that belief. It cannot be impermissible to be in a state that you should have caused yourself to be in: belief-states are not an exception (Rinard 2017).[16]

Note first that this argument is convincing if we throughout employ the *practical* sense of 'ought': given all the prudential and moral reasons at play in epistemic-practical conflicts, it's practically better to comply with your practical reasons rather than to comply with your epistemic reasons. However, on this reading of the argument, the conclusion will then merely state that

> (4p) In epistemic-practical conflicts, it's not the case that you ought *practically* to comply with your epistemic reasons.

Now, (4p) is clearly not in conflict with TV. The proponent of TV can just insist that you still ought *epistemically* to comply with your epistemic reasons in epistemic-practical conflicts. Indeed, this might allow the proponent of TV to explain the intuitive plausibility of the argument while rejecting its conclusion: the argument is *sound* if we consistently read the 'ought' as the practical 'ought', but otherwise it's *not even valid* because it switches between different senses of 'ought' in its premises.

However, I don't think that this is the defeat of the argument. Instead, the most charitable reading needn't make use of different senses of 'ought'. Instead, the argument can be understood as drawing on our intuitions about *normative* 'ought's: if there is no genuinely normative sense of 'ought' in which you both ought to comply with your epistemic reasons and your practical reasons, but there *is* some genuinely normative sense of 'ought' in which you ought to comply with your practical reasons, then it's not the case that you ought (again, in any genuinely normative sense of 'ought') to comply with you epistemic reasons. To spell out this argument, we can define this sense of 'ought' the argument employs as the disjunction of all genuinely normative senses of 'ought':

$$\text{ought}_{N} =_{\text{def}} \text{ought}_{\text{Practically}} \text{ or } \text{ought}_{\text{Morally}} \text{ or } \text{ought}_{\text{Prudentially}} \text{ or } \text{ought}_{\text{Simpliciter}} \text{ or } \ldots$$

It doesn't matter for the argument what the correct disjunction of genuinely normative 'ought's turns out to be – whether it includes an 'ought' *simpliciter* that we gain by weighing or comparing epistemic and practical reasons, or whether it includes the moral and prudential 'ought's.[17] Instead, 'ought_{N}' functions as a *placeholder* for any 'ought's *with genuine normativity*.

So rather than questioning the validity of the argument – which won't be a convincing move in the eyes of opponents of TV anyways – we should instead read the argument along such more charitable lines, namely as employing the notion of 'ought_{N}'. We then have to ask whether its premises are true. Note that there's an implicit premise in the last step, namely:

Agglomeration. If you ought_{N} to φ and you ought_{N} to ψ, then you ought_{N} to [φ and ψ].

It's this principle that gets us from (2) and (3) to (4). For *Agglomeration* allows us to derive from the claims (2') that it's *not* the case that one ought to [φ and ψ], and (3') that one ought to φ, that (4') it's not the case that one ought to ψ. Or, to plug in our argument: since it's (2) not the case that one ought to comply with one's epistemic *and* practical reasons, but (3) one ought to comply with one's practical reasons, (4) it's not the case that one ought to comply with one's epistemic reasons. We need *Agglomeration* to derive (4) from (2) and (3) (cf. Rinard 2019a: 1932–1934).

I will argue in Chapter 5 that epistemic-practical conflicts are cases in which *Agglomeration* fails. This is a challenge for anyone who wants to defend TV while maintaining 'ought implies can': in epistemic-practical conflicts, even though it's not the case that you ought to [comply with your epistemic reasons and comply with your practical reasons] (due to 'ought implies can'), it's still the case that you ought to comply with each set of reasons separately.

Of course, one could instead stick with *Agglomeration* and TV by denying 'ought implies can', thereby blocking the step from (1) to (2). Interestingly, anyone who wants to preserve a distinctive and genuine kind of epistemic normativity must either deny the relevant version(s) of 'ought implies can' that the argument employs, or else *Agglomeration*. I'll explore the latter option in Chapter 5. There I will argue that genuine normative 'ought's from the practical domain and the epistemic domain don't agglomerate, thereby defending TV.

I think that the epistemic 'ought' has normative significance, and, what is more, I think that dissolving the epistemic 'ought' into an 'ought' *simpliciter* obscures the normativity of our cognitive lives: it leaves out a central aspect of our social practice of relating to each other as epistemic agents. I return to the latter issue in Chapter 5.5.

However, before turning to my defense of TV in Chapter 5, let us consider the general structure of the challenge for the normativity of epistemic rationality that is arising from the two arguments I have spelled out throughout the last subchapters. This will then allow us to see also how we can generalize the challenge to a problem about the normativity of other kinds of rationality than epistemic rationality.

4.5 The Possibility of Epistemic Blame

We have considered two arguments against the normativity of epistemic rationality. First, in cases of trivial belief, it seems that there is no requirement to be epistemically rational, even if your epistemic reasons are decisive. Second, in cases of epistemic-practical conflicts, it seems that there is no requirement to be epistemically rational, because the practical reasons for violating the epistemic 'ought' – for making yourself epistemically irrational – seem normatively more significant than the epistemic reasons. At the very least, both arguments show us that the normative import of the epistemic reasons in the relevant cases remains unclear.

I suggest that both arguments arise from a common worry. Here is an initial formulation of this worry:

1 Normative reasons for belief matter for what we ought to believe.
2 Epistemic reasons do not matter for what we ought to believe.
3 Thus, epistemic reasons are not normative reasons.

Both of our challenges lend support to premise (2).

The first challenge appeals to cases of trivial belief. It maintains that although one has excellent epistemic reasons to believe that p in such cases, one is not normatively required to believe that p. Instead, it seems that one is normatively required to believe that p only if there is some practical reason for pursuing the aim of truth. Epistemic reasons do not seem to matter by themselves: they only matter insofar as they are conducive to our practical aims.

The second challenge appeals to cases of epistemic-practical conflicts, maintaining that there is no normative sense in which one ought to believe what one's epistemic reasons seem to decisively support in these cases: there is no normative sense in which we ought to be epistemically rational. Epistemic rationality seems to lack normative authority when there are practical reasons against being epistemically rational. Again, the authority of epistemic rationality seems to depend on practical reasons to be epistemically rational.

Overall, it seems that epistemic reasons are not normatively relevant in the right way to count as normative reasons. Maybe epistemic reasons are therefore not normative (*epistemic anti-normativism*) or they are not genuine reasons after all (*epistemic nihilism*).[18]

Importantly, both arguments against the normativity of epistemic reasons rely on the idea that what it means for reasons to 'matter' is tightly connected to blameworthiness or criticizability. Both arguments maintain that in cases of trivial belief or in cases of epistemic-practical conflicts, we are not criticizable or blameworthy in any substantial sense if we fail to comply with decisive epistemic reasons. From this they derive the conclusion that epistemic reasons are not genuinely normative. What is the connection between blameworthiness and normative reasons that these arguments assume? I suggest that it is the following claim:

> *Epistemic Rationality and Blameworthiness* (EB). Epistemic rationality provides us with normative reasons for belief only if we can be blameworthy *merely* for violating the requirements of epistemic rationality.

EB states a very minimal conceptual connection between normative reasons for belief and blameworthiness. It states that it must be *possible* to be blameworthy merely for non-compliance with epistemic requirements. If this is not possible, epistemic rationality does not, independently from a practical reason to be epistemically rational, provide us with normative reasons for belief. That is, EB states that epistemic blame, as defined at the end of Chapter 4.1, must be conceptually possible if epistemic rationality is to provide us with reasons for belief.

Remember that, as I have argued in Chapter 4.1, Harman's clutter-objection gets its grip on us only because we implicitly assume such a connection between normative reasons and blameworthiness. Epistemologists propose background conditions on requirements of epistemic rationality, like Kiesewetter's attending-condition or Steglich-Petersen's practical reason for forming belief about whether p, precisely because they want to make sense of the *normative significance* of these requirements: they want to explain why it *matters* to us whether we are epistemically rational – why we can be blamed or criticized if we fail to comply with them. The point of spelling out a notion of epistemic blame is to understand the normative force of epistemic rationality: why it matters to be epistemically rational.[19]

Yet most importantly, even if we were to reject that we need to be *always* blameworthy or criticizable for being irrational if epistemic rationality is to be normative, this would not refute EB. For according to EB, in order for epistemic rationality to be normative, it must merely be *possible* to be blameworthy for such non-compliance: there must be some possible cases in which we are blameworthy *merely in virtue of the fact* that we violate an epistemic requirement. Therefore, a convincing reply to the neglected challenge for the normativity of epistemic rationality should make it intelligible how we could be blameworthy in the problem cases at issue: how we can be blameworthy in cases of trivial belief and in epistemic-practical conflicts. For in these cases, the blameworthiness at issue cannot be reduced to blameworthiness for violating practical norms: in trivial cases, there are no practical reasons to be epistemically rational; and in conflict cases, there are decisive practical reasons against being epistemically rational. If there is still a sense in which we are *epistemically* blameworthy in these cases when we comply with our practical reasons but not with our epistemic reasons, then this shows that epistemic reasons are normative even when it's not the case that we *practically* should comply with them.

I will address potential objections against EB in the next chapter. There I will defend the view that skeptics about the normativity of epistemic rationality are right in assuming EB, but that they are wrong in their claim that there is no distinctively epistemic kind of blame: even in trivial cases, as well as in epistemic-practical conflicts, we can be blameworthy. This reveals the normative significance of epistemic reasons and rationality. For now, I will contextualize the challenge for the normativity of epistemic rationality within the more general debate on the normativity of rationality.

4.6 Are Right-Kind Reasons Normative?

Rationality has been prominently understood as mental coherence (see esp. Broome 2007, 2013). Kiesewetter (2017) calls the norms that correspond to this kind of rationality 'requirements of *structural* rationality', because

they are concerned with the relation among, or structure of, our mental states. Structural rationality is supposed to require, for example, to intend what one believes one ought to do; or not to believe what one believes one lacks sufficient evidence for. These are norms to avoid certain combinations of attitudes. It has been pointed out at length that, if we want to save the idea that these *structural* requirements are normative requirements (that is, if we want to say that we always have a *normative reason* or *ought* to comply with these requirements), then we face several problems which seem to be unsolvable (see Chapter 3.1).

The argument I spelled out in the present chapter poses a challenge for the normativity of rationality that has been neglected in this literature. The challenge arises even if we adopt an account of rationality that is more promising for preserving the normativity of rationality than rationality as coherence, such as rationality as responding correctly to reasons. The challenge is to defend the normativity of so-called 'right-kind' reasons. The challenge is just as pressing for any defense of the normativity of rationality as other challenges are – such as showing that irrational incoherences guarantee a failure to respond to reasons or defending subjectivism about normative reasons.[20] I will now explain this challenge in some more detail.

Kiesewetter (2017, 2020) defends a view according to which rationality consists in responding correctly to one's (possessed, accessible, or available)[21] reasons. The argument for the normativity of rationality seems straightforward once such a view is established:

> If rationality consists in responding correctly to reasons, then rational requirements could be understood as inheriting both their content and their authority from the content and authority of the relevant reasons.
> (Kiesewetter 2017: 160)

That is, according to this conception of (attitudinal) rationality, what rationality requires of us is just what we ought to believe, desire, feel, or intend:

(RO) Rationality requires of us to φ if and only if we ought to φ.

One objection to this conception of rationality comes from objectivism about 'ought': sometimes we cannot know whether we ought to φ, but we are still rational if we respond correctly to our possessed or accessible reasons while ignoring reasons that are unavailable to us (Broome 2007: 253). Defenders of rationality as reasons-responsiveness argue that we should reject the first assumption of this "quick objection", as Broome calls it: we can always know what we ought to do, because what we ought to do is determined by the reasons that are *possessed* by us or *available* to us

(Kiesewetter 2017: chapter 8; Lord 2018: chapter 8). This is subjectivism about 'ought' and reasons. It is in line with the idea that not doing what we ought to do normally implies personal criticizability. For if one could have known that φing was impermissible, then one is, at least when one lacks an excuse, criticizable for φing.[22]

I stay neutral here about whether we could plausibly defend (RO) against the quick objection by rejecting objectivism about reasons, as Kiesewetter and Lord do. However, even by endorsing subjectivism instead, one does not escape the challenge I set out for the conception of rationality as correctly responding to reasons. This is because the challenge arises from a more fundamental assumption in the debate about the normativity of rationality

(RC) We always ought to respond correctly to our reasons.

How can anyone reasonably call (RC) into question? It might be argued that (RC) is an obvious analytical truth. Doing what you ought to do just *means* that you give the response (or one of the responses) that your reasons favor most, and to give this response for those reasons that favor it. To give this response for these reasons is to respond 'correctly' to them. Thus, to do what you ought to do just means that you respond correctly to your reasons.

However, note that, if we spell out how the claim is understood by its main proponents, then it is not trivial anymore:

(RC*) We always ought to respond correctly to our right-kind reasons.

The argument I have spelled out throughout this chapter calls (RC*) into doubt. For it questions the normativity of epistemic reasons – i.e., right-kind reasons for belief. As we will see later (in Chapter 6.1), the challenge can be generalized to *all* right-kind reasons.

To see how we might doubt (RC*), note that it is denied by *pragmatists* about what we ought to believe. For pragmatists, there will be cases where our right-kind reasons all favor a specific response but in which it is not true that we ought to give the response. For example, our epistemic reasons might all favor the belief that God does not exist, but since – as we can stipulate – it would be better for us to believe in God than not to believe in God (no matter whether God actually exists), our practical or wrong-kind reasons for belief favor believing in God (see Pascal 1670: §233; James 1896). It might be true, according to pragmatists, that we ought to believe in God, even though the correct response to our *right-kind* reasons (here: the *epistemic* reasons) would be not to believe in God. Currently, there is a debate about whether we can weigh or compare right- and wrong-kind

reasons to determine what one ought to believe, all-things-considered, in such cases (for proposals, see Reisner 2008, forthcoming; Howard 2020; Meylan 2021). If we sometimes ought all-things-considered to believe what is favored by our practical or wrong-kind reasons, rather than by our epistemic or right-kind reasons, then there is a sense in which (RC*) is false. So, pragmatists will deny (RC*) in this sense.

Importantly, we need not be pragmatists to call (RC*) into doubt. More generally, we might doubt (RC*) because we doubt the *normative force* of epistemic reasons and other reasons of the right kind. However, this is also a main motivation for pragmatism. For if epistemic reasons aren't normative, then wrong-kind reasons remain as the only candidates for normative reasons for belief. If wrong-kind reasons weren't reasons for belief – but were rather, say, only reasons for actions of *bringing about* or *maintaining* states of belief – and if epistemic reasons weren't normative reasons at all, then there wouldn't be *any* normative reasons for belief whatsoever. To avoid this implausible conclusion, we could just endorse pragmatism and accept wrong-kind reasons as genuine normative reasons for belief.

(RC) is trivial if we read 'reasons' as 'normative reasons'. But (RC*) isn't trivial. For we can doubt the normative force of right-kind reasons. That is, we can think that sometimes, we should *simpliciter* not respond correctly to our right-kind reasons, because we ought to comply instead with the verdict of our wrong kind reasons.[23] Alternatively, we might think that right-kind reasons aren't even normative reasons. Instead, they derive their normativity from wrong-kind reasons that favor complying with our right-kind reasons. According to such a view, if there is no practical value in holding a true belief about p, then epistemic reasons – which merely indicate truth – have no normative force.[24] While the first kind of view could still allow for a *pro tanto* normativity of rationality, the second view would deny that rationality is normative in any way—at best, it can be said to be *derivatively normative* whenever there are wrong-kind reasons to be rational.[25]

If the overall verdict of one's right-kind reasons lacks normative force, then the norms of rationality could not, as Kiesewetter puts it, 'inherit' such normative force from right-kind reasons. It would then become mysterious how rationality, understood as responding correctly to right-kind reasons, can be normative. Therefore, we need to make the normative force of right-kind reasons *intelligible* rather than taking it for granted, if we want to defend the normativity of rationality.

One could object that there is an uncontroversial sense in which (RC*) is true, whether or not epistemic reasons are 'normative': one ought *epistemically* not to believe in God if one's epistemic reasons for belief in God are insufficient. So, epistemic reasons are normative for *epistemic* rationality (Paakkunainen 2018). I don't think this response is very convincing. I

agree that, if this is all that defenses of the normativity of rationality wish to defend, then (RC*) cannot be doubted. However, they in fact wish to defend more, and moreover I think that they *should* aim to defend more. For defenses of the normativity of rationality endorse the idea that irrationality is *personally criticizable* (see Chapter 3). Importantly, the criticism that is appropriate when someone is irrational is supposed to be different from merely criticizing a system for malfunctioning and from criticizing a bad move in a game. The idea is that one doesn't merely fail *relative to the game of rationality* when one violates rational requirements. Rather, one fails independently of whether one cares about being rational and independently of whether one has practical reason to be rational. Regarding epistemic rationality, what the debate asks is whether epistemic rationality is an independent source of normative reasons, such as prudence or morality, rather than merely a standard according to which we can rank beliefs. For such standards might lack significance independently of our prudential or moral reasons for scoring high in the relevant ranking. Asking for the normativity of epistemic rationality is to ask whether criticizing a belief as 'irrational' has more significance than criticizing a move in chess as bad, relative to the standards of good chess. This is why arguing that one always ought *epistemically* to respond correctly to epistemic reasons isn't yet sufficient for defending the normativity of rationality.[26]

Next, one might object that the challenge I have spelled out here has already been met. For instance, Kiesewetter (2021) defends the normativity of epistemic reasons. However, the features of normative reasons he identifies – providing partial justification, being premises in good reasoning, and being good bases for responses – won't convince someone with pragmatist inclinations. Pragmatists would just deny that epistemic reasons provide partial justification in the absence of a wrong-kind reason to comply with your right-kind reasons. And they would argue that the fact that right-kind reasons are good for reasoning or basing doesn't imply that we *ought* always to conform with the verdict of right-kind reasons. It isn't obvious how to meet such pragmatist replies, without ending up throwing intuitions at each other. I therefore think that addressing this neglected challenge for the normativity of rationality requires us to more directly engage with the arguments that motivate the views that are incompatible with the normativity of rationality, understood as responding correctly to right-kind reasons. It requires us to defeat the pragmatist at their own game, as it were.

Thus, if right-kind reasons weren't normative reasons, then neither of the two dominant approaches to rationality – as a kind of reasons-responsiveness and as a kind of coherence – would vindicate the normativity of rationality.[27] Moreover, if right-kind reasons *were* normative, but can still *conflict* with wrong-kind reasons in some cases, and if there was a

meaningful question of whether to φ *all reasons considered* in such conflict cases, then rationality as responsiveness to right-kind reasons could at best be *pro tanto* normative.

In their extensive defenses of the normativity of rationality, Kiesewetter and Lord spend much space engaging with two kinds of challenges for the normativity of rationality:

a Reducing any intuitively irrational incoherences among one's attitudes to failures to respond to reasons (Kiesewetter 2017: chapters 9 and 10; Lord 2018: chapter 2) and
b Arguing that the notions of 'ought' and 'reason' are subjective or perspective-dependent (Kiesewetter 2017: chapter 8; Lord 2018: chapter 8).

Neither author considers the normativity of right-kind reasons within these books, however. Yet if the doubts about their normativity are well-motivated, then the current literature on the normativity of rationality neglects an important challenge. Indeed, the challenge seems *at least as important* for defending the normativity of rationality as are (a) and (b).

Indeed, the whole modern debate, beginning with works from Parfit, Scanlon, and Kolodny, takes for granted the normativity of epistemic reasons for belief and right-kind reasons for intention and desire. Yet pragmatist accounts of epistemic evaluation are older than the recent debate on rationality (see Meiland 1980; Stich 1990). Nevertheless, such views never informed this debate. As Laura Callahan (2023: 6) observes, the literature on rationality assumed that the normativity of epistemic reasons just doesn't fall within the scope of their discussions. However, asking the normative question about rationality ('why be rational?') is to ask about the broader significance of epistemic and other kinds of rational evaluation to our lives – just as asking the normative question about morality ('why be moral?', see Korsgaard 1996) is to ask about the broader significance of morality to our lives. The challenge coming from pragmatist and instrumentalist accounts of epistemic evaluation should therefore fall within the scope of the debate on the normativity of rationality.[28]

Although I have focused on epistemic reasons in this chapter, I will later generalize the challenge by considering how it bears on right-kind reasons for attitudes other than belief (see Chapter 6.1). For now, we remain in epistemology.

4.7 Summary

According to reasons-responsivist views of rationality, rationality consists in responding correctly to the (possessed or available) right-kind reasons.

Recent defenses of the normativity of rationality tend to assume that this immediately implies that we always ought to be rational, because we always ought to respond correctly to our right-kind reasons. However, this follows only if right-kind reasons are normative reasons. Recent epistemological discussions have questioned this assumption by appealing to cases of trivial beliefs that are decisively supported by epistemic reasons, and by appealing to epistemic-practical conflicts, in which decisive epistemic reasons seem to be normatively irrelevant, and where only wrong-kind reasons for belief seem to be relevant for what we (normatively) ought to (cause ourselves to) believe. It seems that epistemic rationality does not provide us with normative reasons independently of wrong-kind reasons to be epistemically rational. If this is right, then rationality isn't normative. Even if rationality always provides us with right-kind reasons for being rational, rationality might not be normative because right-kind reasons might not be normative.

In the next chapter, I argue that this challenge only arises because we misinterpret the normative significance of epistemic reasons. Properly understanding the normativity of epistemic reasons as being relevant for our evaluations of epistemic character allows us to preserve the normativity of epistemic rationality. This, in turn, will underwrite the idea that we are directly responsible to epistemic rationality, because it implies that we can be blameworthy for mere epistemic irrationality. Indirect Voluntarism (see Chapters 1–2) cannot explain this direct blameworthiness for epistemic irrationality. By defending blameworthiness for epistemic irrationality, the next chapter defends the normativity of epistemic rationality against the recent challenge spelled out in this chapter *and* it refutes Indirect Voluntarism.

Notes

1 See Paakkunainen (2018), Kiesewetter (2021), and Schmidt (2024a, forthcoming c) for critical diagnoses of these doubts.

2 For recent accounts with similar implications, see Bondy (2018), Cowie (2014), Mantel (2019), McCormick (2015, 2020), Papineau (2013). Of these, only McCormick identifies explicitly as a pragmatist about reasons for belief (like Rinard). Instrumentalists, like Steglich-Petersen and Skipper, argue that they can avoid a commitment to pragmatism – i.e., to the existence of practical reasons for belief *per se* – by saying that it is always *evidence* that motivates our beliefs (rather than practical reasons). Instrumentalists merely claim that evidence *gains its normative authority* from practical considerations (see Steglich-Petersen and Skipper 2019: 11; also Cowie 2014: 4004–4005). I am not here interested in subtle differences between the versions of pragmatism and instrumentalism I consider. I am only interested in their implication that epistemic reasons are not 'genuinely normative'.

3 For the sake of brevity, I use the notion 'blameworthiness' in what follows to cover both the notion of blame and other forms of personal criticism. See Kiesewetter (2017: chapter 2) on the notion of *personal* criticism as contrasted with, say, criticism of a knife for not being sharp. One might think that

'blameworthiness' is essentially a moral notion. As I will argue in Chapter 5, epistemic blame is often a *sui generis* kind of blame. This might warrant the label 'criticism' instead of 'blame' for the *sui generis* forms of epistemic blame. However, I take this to be a mere terminological issue. Importantly, Chapter 7 argues that epistemic blame can take the same forms as moral blame.

4 Kiesewetter is not concerned with background conditions on (EN*). But his view might naturally be extended to (EN*) in the way described. As argued at the end of the last subchapter, a background condition on (EN*) is as important as a background condition on (EN) to make it intelligible that these norms have reason-providing force. Suppose, for instance, that the belief that p is a practically beneficial belief that you have held for as long as you can think, but you lack evidence for p. Why give up this belief if it benefits you? With Kiesewetter's background condition, we could say that you should give up this belief *from an epistemic point of view* as soon as you consciously consider whether p. This is more promising than saying that you should give up the belief even if you have never consciously considered whether p, and instead simply because you lack evidence for p.

5 See Rinard (2015: 220) and Steglich-Petersen (2011: 23) for this verdict about structurally analogous cases.

6 It is important to note that Steglich-Petersen does not think that this background condition is necessary for (EN*) to express an epistemic requirement. However, as I have argued in Chapter 4.1, he thereby ignores that (EN*) is faced with very similar challenges as (EN): Why give up a lot of beneficial beliefs that are insufficiently supported by one's evidence? And why blame or criticize anyone for believing something that is insufficiently supported by their evidence if it doesn't matter at all whether they believe it? It seems that Steglich-Petersen's instrumentalist framework commits him to a background condition not only for (EN), but also for (EN*): without assuming pragmatic background conditions, both norms seem questionable as genuine normative requirements.

7 Note that the reason for forming a belief about p cannot itself be an epistemic reason for belief, because epistemic reasons for belief favor believing a specific proposition – they favor believing that p or believing that not-p. Steglich-Petersen's reason to form a (true) belief about p, by contrast, does not favor believing a specific proposition. It merely favors having a true belief about a matter, whatever this belief turns out to be.

8 In later works, Steglich-Petersen accepts that the reason to form a belief about p must in fact be a reason to form a *true* belief about p (see esp. the formulations of the norms of belief in Steglich-Petersen and Skipper 2019, 2020).

9 One might wish to interpret Steglich-Petersen's view without committing to the idea that the practical reason is always a reason for an action of managing one's beliefs (causing, maintaining, etc.). However, as pointed out in endnote 7 above, the "reason to form a belief about whether p" cannot be an epistemic reason for belief. It must thus be practical. If the practical reason does not favor an action, it must favor the state of believing itself. This would, however, commit Steglich-Petersen to pragmatism about reasons for belief. Whether this reason favors actions of managing beliefs or rather beliefs themselves, the condition renders the requirement practical rather than purely epistemic.

10 See Stroud (2006) on a plea for doxastic partiality. Other cases of this structure include overestimating one's abilities, thereby increasing one's performance (Hazzlet 2013: 44–52), discarding statistics about divorce rates (Marušić 2015), and religious and other meaning-making beliefs (McCormick 2015: 52–65).

11 For the traditional arguments against this assumption, and thus against practical reasons for belief, see Shah (2006) and Hieronymi (2006), as well as Chapter 3.2 on the dispute between pragmatism and evidentialism about reasons for belief. For recent defenses of practical reasons for belief, see McCormick (2015, 2020), Rinard (2015, 2019a; 2019b). For replies, see Arpaly (2023), Kelly and Cohen (2024), Schmidt (2022), and Vahid (2022).

12 Cf. Heil (1992), Feldman (2000: 680–681), Kelly (2003: 619), Pojman (1993). Evidentialists about reasons for belief commit to TV, arguing that practical reasons are reasons to *cause* belief (or to *desire* a belief, etc.), while the only reasons to *believe* are provided by evidence; see Hieronymi (2006), Kelly (2002), Shah (2006), Skorupski (2010), Way (2016). Recent proponents of TV are Berker (2018), Christensen (2021: 514), and Wedgwood (2017: 41–46).

13 We must interpret 'to cause a belief' broadly so as to encompass all kinds of activities and omissions that we can engage in for practical reasons and that have some kind of foreseeable effect on our doxastic attitudes, like diverting our attention, avoiding evidence or avoiding active reflection, and engaging in long-term projects of self-deception.

14 In Schmidt (forthcoming c), I dubbed them "doxastic dilemmas" to emphasize the impossibility of complying with both one's practical reasons and one's epistemic reasons. This is just a matter of terminology.

15 Cf. Cowie (2019), Howard (2020), Leary (2017), Maguire and Woods (2020), McCormick (2015, 2020), Steglich-Petersen and Skipper (2019), and others.

16 Cf. Parfit (2011: 432), who claims that 'ought's to be in states are nothing but 'ought's to bring yourself into the state, or to desire to be in the state. I discuss this reductive view of 'ought's to be in states in Schmidt (2022: 1810–1816). Importantly, Parfit rejects the view that 'ought to believe' just functions like 'ought to be in a state', and so rejects Rinard's view that beliefs are not an exception. I ultimately agree: intentional mental states *are* an exception, since they're responsive to reasons. Evaluating them like a non-mental state obscures an important facet of their normativity.

17 It's widely held that you can be *all-things-considered* or *practically* justified not to do what you morally or prudentially ought to do, and so you might not be *blameworthy* for not doing it. However, even if we tie genuine normativity to blameworthiness, as I propose that we should, we can still say that moral and prudential reasons and 'ought's are normative: by weighing them, we gain the practical 'ought', and we're clearly blameworthy if we violate this 'ought' (at least absent excuse or exemption). See Kauppinen (2023: 138–142) for discussion. See also Chapters 2.3–2.4. I assume that practical reasons include normative prudential and moral reasons, and that they can be weighed to derive the overall practical 'ought'.

18 Both positions are distinguished by Kiesewetter (2021). I will deal with them as a package by arguing that epistemic reasons are normative reasons, and thus won't discuss which one would be the more plausible conclusion to draw. As Kiesewetter points out, nihilism faces the worry of being very revisionist, while anti-normativism faces the serious challenge of explaining what kinds of reasons epistemic reasons are if they are not normative. I won't repeat his well-argued worries. Instead, I will offer a novel positive alternative to anti-normativism *and* nihilism throughout Chapters 5–7.

19 A recent account of epistemic criticizability which is motivated in this way is Kauppinen (2018). On the side of the skeptics about the epistemic 'ought', McCormick (2020) argues that all blame for belief is ultimately prudential or moral. She *concludes from this* that there is no distinctively epistemic

normativity. Thus, she explicitly reaches her view by accepting EB (see McCormick 2020: 30). The problem with McCormick's argument is that she does not consider the epistemic forms of blame that have been spelled out by Boult (2021, 2024c), Brown (2020), Kauppinen (2018), and Piovarchy (2021) to which I return in Chapter 5. She thinks of blame mainly in moral terms.

20 Remember that, even if we wish to retain the normativity of coherence requirements (like Worsnip 2021), we must argue that coherence requirements provide us with right-kind reasons (see Chapter 3.7). Therefore, also coherentists who wish to preserve the normativity of rationality must rely on the normativity of right-kind reasons.

21 Kiesewetter employs the notion of *availability*, while Lord (2018) favors the notion of *possession* to characterize these reasons, which in turn presupposes *access* to reasons. As explained in Chapter 3.1, I ignore these subtle differences. What matters is that *rationality supervenes on the mental* (see Wedgwood 2017: chapter 7): two subjects with identical (non-factive) mental states cannot differ in their rationality (say, because one subject is systematically deceived). Rationality, maybe in contrast to justification, is an essentially *internalist* concept. Reasons that a subject cannot be aware of cannot make a difference to their rationality.

22 More precisely, one will either be criticizable for not doing what one thinks one ought to have done (that is, for *akrasia*), or for not having known what one ought to have done (that is, for culpable ignorance). It is important to note that the 'ability to know' that is necessary for blameworthiness needs to be adequately specified. The fact that we can *in principle* know something that is important to know does not always make us criticizable for failing to know it, for there may still be no *reasonable* way to come to know it.

23 For defenses of this view with regard to epistemic reasons, see Howard (2020), Meylan (2021), Reisner (2008, forthcoming).

24 See Côté-Bouchard and Littlejohn (2018), Cowie (2014, 2019), Grimm (2009), Mantel (2019), Maguire and Woods (2020), McCormick (2020), Papineau (2013), Steglich-Petersen and Skipper (2019, 2020) and Rinard (2015, 2017, 2019a, 2019b; 2022).

25 Kiesewetter (2021) distinguishes two readings of the second view: *epistemic anti-normativism*, according to which epistemic reasons are reasons, but not normative reasons (which gives rise to the challenge of explaining what kind of reasons epistemic reasons are), and *epistemic nihilism*, according to which epistemic reasons aren't reasons at all (which Kiesewetter takes to be too revisionist). I won't be concerned here which of the views would be more plausible. The distinction is also explicit in the work of Jonas Olson, who has earlier defended nihilism (Olson 2011, 2014) but later anti-normativism (Olson 2018).

26 Cf. Côté-Bouchard (2017: 412–413), who argues on similar grounds that appeals to *epistemic value* don't suffice to defend the authority of epistemic norms. The same goes for reasons-based conceptions of epistemic norms.

27 Remember that also coherentists that wish to vindicate the normativity of rationality must claim that we always have right-kind reasons for being coherent in the relevant sense. See Chapter 3.7.

28 Wedgwood (2017, 2023) defends the view that rationality is a virtue—a claim which could explain why we should care about rationality (Kiesewetter 2024), and why irrationality is criticizable, given that it is a central vice. Yet Wedgwood also doesn't address the recent doubts about the normativity of epistemic reasons and other right-kind reasons coming from pragmatism.

5 Blameworthiness for Epistemic Irrationality

> [O]ur control over […] actions yields only an *indirect* control over belief and […] such indirect control cannot underwrite the *direct* responsibility of belief to epistemic norms.
>
> David Owens, *Reason without Freedom* (2000: 87)

The challenge from the last chapter can be summarized in the following argument:

1 Epistemic rationality provides us with normative reasons for belief only if we can be blameworthy *merely* for violating requirements of epistemic rationality (EB).
2 We cannot be blameworthy merely for violating requirements of epistemic rationality.
3 Thus, epistemic rationality does not provide us with normative reasons for belief.

Premise (1) assumes a weak connection between normative reasons and blameworthiness. I will discuss the premise in some more detail in this chapter by addressing common objections. Premise (2) was supported by cases of trivial belief and by cases of conflict between epistemic reasons and practical reasons (of the wrong kind),[1] where the latter favor having a beneficial belief that is ill-supported by epistemic reasons, or the practical reasons disfavor a harmful belief that is well-supported by epistemic reasons. In each case, it seemed that practical reasons do the normative work: if there are no practical reasons to comply with the epistemic reasons (trivial attitude cases), then it seems that there is no normative sense in which we ought to have the belief that is decisively supported by epistemic reasons; and in cases where there were decisive practical reasons not to follow the epistemic reasons (conflict cases), it seems that there is no normative sense of 'ought' in which one ought to comply with the

DOI: 10.4324/9781003382973-7

epistemic reasons. This is not what to expect if epistemic reasons were normative reasons.

According to the view expressed in (3), we have normative reasons to be epistemically rational only if there is some practical reason to be epistemically rational – that is, some reason indicating the practical value of having a rational belief. This view argues that even if epistemic rationality does not *by itself* provide us with normative reasons for belief, it often *appears* as if it provides us with normative reasons because we *normally* have a practical reason to be – or at least to try to ensure that we are – rational: rationality normally helps us to attain the truth, which is practically valuable most of the time. (3) expresses this pragmatist-instrumentalist view about the normativity of epistemic rationality that claims that we only ought to be epistemically rational if there are practical reasons to be rational – epistemic reasons are normatively impotent by themselves.[2]

I have already pointed out in the last chapter that (1) has a high *prima facie* plausibility because it assumes a very minimal connection between normative reasons and blameworthiness: if we are never blameworthy merely for failing to comply with norm N, then it seems that N at best has the conditional normativity of game rules or etiquette; that is, we then should comply with N only if there is some independent practical reason to comply with N. However, some doubts might arise about whether there is any such tight connection between normative reasons and blameworthiness. I will first defend (1) against these main worries (Chapter 5.1). In order to reply to the new challenge for the normativity of epistemic rationality, I then have to argue against (2). This will require me to defend the possibility of blameworthiness for mere epistemic irrationality. I first defend the possibility of blameworthiness for epistemic irrationality, and especially how we can be epistemically blameworthy in cases of trivial belief and in epistemic-practical conflicts (Chapters 5.2–5.4). I then argue that this epistemic blameworthiness reveals the normativity of epistemic rationality, although a kind of normativity that is, in a sense, evaluative (Chapters 5.5). Finally, I argue that this refutes Indirect Voluntarism about doxastic responsibility (Chapter 5.6). In the next chapter, I then turn to the rationality of attitudes other than belief, generalizing the challenge as well as my reply.

5.1 Normative Reasons and Blameworthiness

The first objection to premise (1) points out that children or some non-human animals can act for reasons but cannot be blameworthy if they fail to do what is decisively supported by their reasons, because they are not responsible agents. Analogously, some children and animals might be considered as having normative reasons for belief even though they cannot

be blameworthy merely for violating epistemic norms. Thus, (1) does not seem to hold.

The objection can easily be met by pointing out that (1) does not imply that *everyone* can be blameworthy merely for violating epistemic norms. As I have explained in the previous chapter, (1) only states that it must be possible to be blameworthy in a distinctively epistemic sense if epistemic rationality is to provide us with normative reasons for belief. If there are subjects who are not fully responsible agents or have not yet developed into fully responsible agents, then they might also be exempted from epistemic blame, even if there is a sense in which they can believe on the basis of epistemic reasons. But that does not count against the idea that those who are fully responsible for their actions and beliefs must sometimes be subject to epistemic blame if epistemic rationality is to provide us with normative reasons.

Secondly, and more seriously, one might wish to deny (1) if one is an objectivist about the meaning of 'ought' and 'reasons'. Objectivists deny a close connection between failing to do what one ought to and being blameworthy – a connection usually utilized or argued for by subjectivists.[3] Objectivism about practical reasons states that 'S ought to φ' means that φing is the best option, no matter whether S is in a position to know or has cognitive access to whether φing is the best option (see Chapter 3.1 on subjectivism and objectivism). Yet, surely, we are not always blameworthy for failing to do what is objectively best, especially in cases where we are non-culpably ignorant about what is best. It seems that, for objectivists, there is no tight connection between blameworthiness and violating normative reasons.

However, (1) is uncontroversial for objectivists. This is so again because (1) states a very *loose* connection between reasons and blameworthiness. Even if we grant the objectivist that we can be completely ignorant of our reasons, there will be at least some possible cases in which we are blameworthy for failing to give the response that is best. When we focus on actions, such cases will be cases where we either act against our better knowledge of what is best or where we culpably fail to know what is best and do the wrong thing as a result of our ignorance. Analogously, if we are objectivists about reasons for belief – for instance, if we think that we ought to believe what is true rather than what is supported by our evidence – then we can still argue that we are at least *sometimes* blameworthy merely because we fail to have a true or correct belief: namely, in some of the cases where the evidence was accessible to us, and yet we did not believe what it supported. So even if we spell out epistemic norms in objectivist terms (for instance, 'one ought to believe only what is true'), this does not yet give us an argument against the idea that such norms provide us with epistemic reasons only if we can be blameworthy merely for violating

them. The accessibility of the evidence does not render the epistemic norm practical, and thus does not compromise the idea that we are sometimes blameworthy *merely* for violating epistemic norms.

Finally, one might want to object to (1) by adopting a permissivist epistemology. Permissivism states, roughly, that our total set of evidence permits more than one set of doxastic attitudes to take toward each (or at least some) proposition(s).[4] According to a permissivist, it could be true that when we have sufficient evidence for a proposition, it is both epistemically permissible to believe it as well as epistemically permissible not to believe it (and suspend judgment instead). Such an account might seem to be exactly the conclusion to draw from Harman's clutter-objection (see Chapter 4.1): we are not rationally *obligated* to believe everything that our evidence sufficiently supports, and thus we are not blameworthy for not believing implications that are trivial or very complicated. But we are *permitted* to believe propositions with sufficient evidential support.

However, it is not straightforward how permissivism could pose a problem for (1). The premise states that epistemic rationality provides us with normative reasons for belief only if we can be blameworthy merely for violating an epistemic norm. To deny this, the permissivist would have to argue that the possibility of epistemic blame is not a necessary condition for the normativity of epistemic rationality. They might do so by claiming that we are not blameworthy for violating norms of permissibility, and yet these norms provide us with normative reasons.

But it is difficult to see why we shouldn't be blameworthy for violating norms of permissibility. If something is permitted only under a certain condition, then it is *not* permitted – and thus *prohibited* – if this condition is not fulfilled. That is, that I am only permitted to believe that p if p is sufficiently supported by my evidence implies that I *ought* not to believe that p whenever p is *not* sufficiently supported by my evidence (see claim (EN*) in Chapter 4.1). Thus, even if we understand the epistemic norms only as permissions, this will still require us to make sense of the idea that we sometimes *ought* to be in certain doxastic states under certain conditions. Epistemic norms would then require us to *withhold* belief in propositions that are not sufficiently supported by our evidence (given suitable non-pragmatic background conditions). Such a normative requirement would still be puzzling if we could not be blameworthy merely for violating it. Thus, if one formulates epistemic norms in terms of permission rather than obligation, one does not thereby formulate a problem for (1).

This confirms that (1) is uncontroversial, mainly because it rests on a very loose connection between normative reasons and blameworthiness. The premise can be accepted both by proponents and opponents of the normativity of epistemic rationality, by objectivists, and by permissivists alike, and it is thus a hinge around which the debate can progress. If the

premise is plausible when it comes to epistemic rationality, then we should also accept it when it comes to the distinctive rationality of other kinds of attitudes. For remember that epistemic reasons are right-kind reasons for belief, and that other attitudes have their own distinctive rationality determined by right-kind reasons for these attitudes. There is no reason to assume that the normativity of other right-kind reasons doesn't have such a minimal connection to blameworthiness. I return to the generalized version of our challenge in Chapter 6.

It is important to see, however, that (1) might be false if we assume that the relevant sense of 'blameworthy' must be a paradigm form of *moral* blame. A person's epistemic failure does not obviously give rise to emotions like resentment, indignation, or guilt. If someone believes that candidate X will win the next election because the flight of the birds gave them a sign, then our reactive attitudes are not, or not necessarily, of that moral kind. But this does not yet rule out that there could sometimes be a distinct kind of *epistemic* blame appropriate in such cases even when the moral reactions aren't. This is also why (2) is not trivial: it claims that there *is no such thing* a distinctively epistemic kind of blame.[5]

I will now turn to the form of epistemic blame that can legitimately arise due to our epistemic answerability, and to my proposal of how this concept of epistemic blame might help us to understand the normativity of epistemic rationality. It will allow us to make sense of epistemic blameworthiness both for trivial belief and for epistemically irrational belief that one ought practically to (have caused oneself to) have.

5.2 Epistemic Reactive Attitudes and Epistemic Normativity

Pamela Hieronymi's and Angela Smith's approaches to the nature of blame (Hieronymi 2004, 2019; Smith 2013) are in line with Thomas Scanlon's (1998, 2008) account. According to this family of views about the nature of blame, blaming someone need not mean that one feels emotions like resentment or indignation toward the person. Rather, we might blame someone merely by *modifying our relationship* with them in a certain way. We might blame a person without feeling any hostility toward them, for instance, by just ceasing to be friends, or by no longer providing special support to the blamee, or by not taking pleasure in their successes, or by not valuing their opinions in the way we did before, or by developing a general sense of distrust toward them. Recently, Cameron Boult (2021, 2023, 2024c) in particular has applied these accounts to the epistemic domain. For my purposes, the main claim that I adopt from accounts of epistemic blame is that there are distinctively *epistemic* responses to norm violations which need not amount to moral blaming responses, and which in turn can help us to identify when we react to violations of distinctively

epistemic norms. So my overall argument need not rely on Boult's specific view, even though I prefer his account to others.[6]

Importantly, not all relationship modifications count as instances of blame. First, one might modify a relationship in a positive way, say, when one is so fond of someone that one wants to be closer friends with them; or when a parent finds out that their child committed a crime and, in response to this, the parent cares even *more* about their child (Smith 2013: 137). Secondly, even negative relationship modifications can happen without blame – as when people who live in different places just drift apart. Scanlonian approaches to blame thus owe us an account of what makes a negative relationship modification an instance of blame.[7]

I propose that relationship modifications count as instances of blame if and only if they are responses to blamee's moral or epistemic vice. I do not count as blaming my friend by judging and treating them as unreliable, and by modifying my expectations accordingly, if I am aware that their unreliability is due to factors that do not stem from their faulty character. Such factors might include their newborn child that forces them to spontaneously cancel on me, or their depression that is the cause for their unreliability. Such factors do not give me a reason not to trust them but merely a reason not to *rely* on them.[8] By contrast, it might be legitimate to blame my friend if their unreliability indicates that they do not care about the friendship as much as I can reasonably expect of them as their friend. In this case, they are not fully honest about their attitude toward our friendship. Reducing my trust in them, and thus modifying my relationship with them negatively in response to their vice of dishonesty, can be legitimate.

To illustrate this notion of blame, consider Smith's (2005) discussion of a case from George Eliot's *Scenes of Clerical Life*, where Captain Wybrow fails to notice that Miss Assher never takes jelly, which, according to Smith (2005: 243), "suggests to Miss Assher that she does not yet occupy a distinctive place in his overall emotional and evaluative outlook". Assher might legitimately modify her expectations toward Wybrow, and thus modify the relationship negatively, in response to Wybrow's not caring about the relationship as much as she can reasonably expect of him. According to the proposed account, this can count as blame even if it does not involve emotions of resentment or indignation.

Negative relationship modifications in response to vices can plausibly count as blaming responses because they are legitimate only when directed toward responsible beings. This is because we can only have the specific relationships that are presupposed by these reactions with fully responsible beings. Computers, children, and animals cannot display full-blown vices that give rise to blaming reactions. Their misbehavior can only give rise to impoverished analogs of these reactions. For instance, I might 'not trust' a dog in the sense that I do not know whether they could bite me. The dog's

behavior might be unreliable, but it won't give me a reason to blame the dog, since the dog's behavior is not grounded in a vice (on the assumption that dogs cannot have full-blown vices like fully responsible beings). This indicates that negative relationship modifications in response to vices presuppose a subject's responsibility for their character and attitudes. At the same time, the appropriateness of these reactions does not presuppose that the subject could have managed their character or attitudes: these reactions merely presuppose that they are indicative of an underlying vice, independently of whether it originated in the subject's voluntary conduct. The potential 'coolness' of these reactions does not, *pace* Wallace (2011), count against them as genuine blaming-reactions (see Boult 2021).

One could object that negative relationship modifications merely count as ways of *holding responsible*, but not as ways of *blaming*, since they are too 'soft' or 'cool' to count as genuine blame (Wallace 1994, 2011; Wolf 2011; Menges 2017). This raises the question of how full-blown moral blaming responses like resentment and indignation relate to Scanlonian relationship-modifications. For present purposes, it is sufficient to simply note that if we hold a person responsible in the Scanlonian sense, this can provide the basis for a positive or a negative relationship modification in response to virtue or vice. For instance, we might want to be closer friends because of the person's virtues, or we might reduce our involvement with the person due to their vices. I take it to be of secondary interest whether we *call* these reactions 'praise' and 'blame'. At the very least, they seem to be positive and negative ways of holding a person responsible. This is sufficient for these reactions to reveal the normative significance of a requirement that we mark as violated by modifying our relationships. However, my overall argument won't rest content with such 'soft' or 'cool' reactions (see esp. Chapter 7).[9]

As David Owens (2000: 124) puts it, after discussing the epistemic vice of gullibility: when I display a vice indicating a flaw in my character, then "I cannot be trusted to think and feel as I ought". The normativity of these 'ought's is revealed, according to the view I propose, by the fact that violating them impairs our relationship to others in specific ways so that it becomes appropriate to negatively modify one's relationship – for instance, by reducing one's presumption of trust. This impairment exists even if the person had no reasonable opportunity to manage their vice: as long as the epistemic vices are still genuine *vices* (rather than severe pathologies), non-culpable violations of epistemic norms that reveal a person's epistemic vice can impair our epistemic relationships, and thus give rise to legitimate suspension of epistemic trust and other ways of modifying one's epistemic relationship that count as blame.

Thus, if we allow for a concept of blame in terms of impaired relationships, then we might be able to make room for something like a distinctively

epistemic kind of blame. In an initial attempt, we might state that if we are blameworthy *morally* as soon as our relationship to our *moral* community is impaired, then we are blameworthy *epistemically* as soon as our relationship to our *epistemic* community is impaired. This impairment might matter in specific ways for how we should relate to one another: whether we believe the other person, whether we provide them with information, and whether we engage with them in rational discourse.

One problem with this initial formulation is that one's moral or epistemic community can be epistemically or morally flawed, and thus one might end up impairing one's relationship with them by being morally or epistemically virtuous. I think this problem can be remedied by appeal to Boult's formulation of the position: one is blameworthy epistemically only if one falls short of the *normative ideal* of an epistemic relationship – or, in my preferred terminology, only if one displays an epistemic vice. The epistemically virtuous person won't fall short of this ideal even if they live in an epistemically flawed community. Thus, members of the community won't have a reason to reduce their epistemic trust in the virtuous person. By appealing to the normative ideal of an epistemic relationship, we can explain why being dogmatic or gullible, engaging in wishful thinking, or being biased can make one epistemically blameworthy even in epistemic communities that reward such vices. These vices are all, as Boult puts it, problematic ways of exercising one's epistemic agency that makes one fall short of the normative ideal and thus warrant suspension of one's presumption of epistemic trust.[10]

5.3 Blameworthiness for Trivial Irrationality

What is the verdict, according to this account of epistemic blame, about our blameworthiness for trivial belief that is insufficiently supported by one's epistemic reasons? The normativist about evidence has two strategies available. I suggest that both of them should be combined.

First, they can argue that even violations of epistemic norms in trivial matters *sometimes* indicate a general flaw in the epistemic character of a person. As Boult (2021: 526) puts it, "[s]o long as I modify my intentions and expectations towards them, in a way made fitting by the judgment (however implicit) that they've impaired the general epistemic relationship, then I count as epistemically blaming them." That is, if your friend tends to believe some celebrity gossip that they have read in a magazine they know to be unreliable, then this might give you a (*pro tanto* or *prima facie*) reason to suspend epistemic trust in them. Presumably, this could mean that you should suspend your trust in some situations when it comes to matters of importance because you now have some evidence that their epistemic character is flawed.

Secondly, the normativist can just grant that violations of epistemic requirements do not *always* make it appropriate to suspend trust. For they need not argue that such violations *always* make one epistemically blameworthy. In order to disprove premise (2), it is enough to show that we are *sometimes* blameworthy merely in virtue of the fact that we fail to properly base our beliefs on our evidence. More generally, violating a reason-providing norm need not amount to displaying a criticizable vice. Compare the idea that someone's morally wrong action is not necessarily blameworthy. We all act wrongly from time to time, and we all violate epistemic norms from time to time. We can usually *excuse* each other for occasional lapses and do not regard these lapses as having any significant consequences for our interpersonal relationships. Yet moral wrongs and violations of epistemic requirements are lapses nevertheless – that is, they are violations of norms that provide us with reasons for compliance.

Seeing that reducing epistemic trust is an appropriate negative response to an epistemic vice and that it marks the impairment of an epistemic relationship provides us with a plausible starting point for understanding the significance of epistemic normativity. It allows us to meet the challenge for the normativity of evidence presented in the last chapter. This challenge claims that the absence of a distinctively epistemic kind of blame rules out the normativity of evidence. I have proposed that we can meet this challenge by appealing to recent accounts of epistemic blame. They provide us with ideas about what this blameworthiness could consist in, without having to interpret it as always taking the same form as moral blame. By building on Boult's account, I have suggested that epistemic blame consists in marking impaired epistemic relationships by reducing epistemic trust in response to a person's epistemic vice.[11]

5.4 Blameworthiness in Epistemic-Practical Conflicts

Let us now turn to the second shape of our challenge: the argument that epistemic rationality loses its significance when it clashes with practical norms, as when you practically ought to cause yourself to have an epistemically ill-supported belief about your friend (see Chapter 4.4). I'll argue that you can still be epistemically blameworthy when you violate the 'ought' of epistemic rationality in such cases. This will allow me to defend what I have called

> *The traditional verdict about epistemic-practical conflicts* (TV). Even when, practically, you ought to (cause yourself to) believe that p, it might still be that, epistemically, you ought not to believe that p. In these cases, there is no answer to what you ought simpliciter to believe.

How do we make sure that the possibility of epistemic blame reveals the normativity of the 'ought' of epistemic rationality? According to Antti Kauppinen, normatively significant norms are "rules that someone is accountable for conforming to" (Kauppinen 2018: 3). To argue against premise (2) of the challenge I meet in this chapter, it's sufficient to show that some forms of blame are appropriate *merely* for violating a requirement of epistemic rationality. We sometimes *do* blame others for, say, violating non-normative rules of etiquette or norms of good chess playing. But remember that, if we do so, then we do so because we think that the person normatively ought to comply with etiquette or chess norms in this context: there might be *moral reasons*, for instance, to comply with etiquette or chess norms in the specific situation. A genuinely normative 'ought', by contrast, need not be backed up by reasons from another domain to give rise to legitimate personal criticism or blame when it's violated. This is what John Broome (2013: 27) means when he says that non-normative 'ought's are at best *derivatively* normative (see Chapter 2.3).

Note that, since our focus is on instances of epistemic-practical conflicts in which the person successfully complies with their practical reasons but not with their epistemic reasons, any sense in which the person is blameworthy could only be epistemic: their blameworthiness must arise due to their failure to respond correctly to her epistemic reasons, thus revealing the normative significance of the corresponding norm of epistemic rationality. Again, focusing on epistemic blameworthiness is helpful *dialectically*: since both opponents and proponents of TV accept some connection between normativity and blameworthiness (or criticizability), the question of whether a subject in epistemic-practical conflicts can be blameworthy (or criticizable) for violating the epistemic norm while complying with the practical norm is a central hinge of the debate around which the debate can progress.[12]

With these clarifications in mind, let's turn to my main argument for the normativity of epistemic reasons in epistemic-practical conflicts. Suppose that you comply with your practical reasons in an epistemic-practical conflict, and you consequently violate a norm of epistemic rationality. If there's still a sense in which you're blameworthy, then this must be due to your violation of this epistemic norm – it's the only norm you've violated. Let's start with what I take to be the clearest case of epistemic blameworthiness in an epistemic-practical conflict:

Dogmatic Dan

Dan gets accepted in his community only if he disregards scientific evidence about p. Given the high social costs, it is practically reasonable for Dan to comply with the pressure within his community and to adopt

their belief-forming practices. As a result, he ends up having epistemically unjustified beliefs that manifest dogmatism.

I claim that epistemic blame toward Dan can be appropriate, although he ought to have caused himself to be blameworthy: he had decisive practical reasons to allow himself to become a person who holds beliefs that manifest dogmatism. Clearly, dogmatism is the kind of disposition that makes it appropriate to modify one's epistemic relationship with Dan: to reduce one's trust in him, to be reluctant to engage in rational discourse or share information with him, and to desire that he hadn't believed badly. Therefore, the fact that one ought to have caused oneself to violate an epistemic norm isn't always a justification or excuse for violating it. *Some* epistemic-practical conflicts are cases where the subject causes their own epistemic vice for excellent practical reasons. In these cases, the person is practically required to violate a requirement of epistemic rationality, but they aren't completely justified nor fully excused: they're still blameworthy in an important sense. Thus, cases of practically required epistemic vice show that the epistemic 'ought' has normative significance.

A worry with this argument is that Dan is blameworthy for his *vice* but not for his unjustified beliefs. This objection assumes that Dan is blameworthy *only* for being dogmatic, and not *also* for his unjustified belief. However, we can be blameworthy for the *manifestations* of our vices in our actions and attitudes. Dan can become *less blameworthy* if he, despite his dogmatism, still revises one of his unjustified beliefs in response to epistemic reasons against it.

As an analogy, consider a coward who on occasion doesn't act cowardly although he is confronted with danger. In such a case, the coward makes some extra effort not to let his actions be influenced by his vice of cowardice—say, he resists his impulse to run away. So, even if we grant that he is blameworthy for his cowardice, he seems to be *less blameworthy* if he sometimes manages not to act cowardly. We aren't at the mercy of our vices: it is often up to us whether we allow them to manifest. This implies that we are often blameworthy not only for our vices themselves, but at least *also* for their manifestations in our actions and attitudes.

How could this thought carry over to *epistemic* vices? Clearly, beliefs aren't just passive states, like headaches or tickles. Rather, they are often exercises of our epistemic agency insofar as they are our responses to our epistemic reasons.[13] Suppose that Dan, despite his dogmatism, still revises one of his unjustified beliefs in response to counterevidence. In this case, Dan is keeping his epistemic vice in shackles. He is *less blameworthy* than if he didn't revise the belief. Thus, Dan's degree of epistemic blameworthiness isn't just a function of his epistemic viciousness. It also depends

on whether he revises his epistemically unjustified beliefs in response to epistemic reasons.[14]

We might even wish to say that someone who less frequently acts or believes badly is thereby also *less vicious*: a coward who often controls his cowardice is less of a coward than someone who doesn't; a dogmatic person who often actively revises unjustified beliefs is less dogmatic than someone who doesn't. That is, the fact that they less often act or believe badly (due to their own agency) implies that they are less vicious.[15] When Dan revises an unjustified belief on the basis of counterevidence, he *therein* becomes less dogmatic. In any case, Dan's blameworthiness partly depends on how often (and how severely) he violates epistemic norms of belief—at least insofar as it is *up to Dan* to revise his beliefs.

Suppose that we ask Dan *why* he has one of his dogmatic beliefs. If his belief isn't severely pathological but rather responsive to evidence to a sufficient degree, then he is *answerable* for his belief (Smith 2005; Hieronymi 2006, 2008). Suppose that he tells you that he was brought up with this belief and always had more important things to do than to reconsider it. If this was the explanation of why he holds his belief, then his practical reasons for *remaining* in this belief-state were decisive: he ought to have behaved in a way so that he remains in the belief-state; for he had no reasonable opportunity to get rid of this belief. In a sense, his belief is non-culpable (see Chapter 7 and Schmidt 2024b). While such non-culpability is a *practical justification* for remaining in the belief-state, it doesn't render him *epistemically* blameless. Rather, Dan's upbringing ingrained a disposition toward holding epistemically unjustified beliefs. We *rightly expect* him to drop these beliefs, and we'll reduce our epistemic trust if he doesn't. That he had decisive practical reasons to stick with the belief neither implies an epistemic *justification* nor a *full excuse* for the belief.

What about cases in which a person holds an epistemically unjustified belief *without* thereby manifesting an epistemic vice? It doesn't matter for my argument whether such a person is blameworthy: what matters is that it's *possible* that (a) the person does the right thing, practically, (b) fails to believe the right thing, epistemically, and (c) is still blameworthy for not believing the right thing. *Dogmatic Dan* shows that there are such cases.

We might endorse a strong connection between character and blame. According to George Sher, for instance, "the force of many excuses is precisely to imply that the agent did *not* manifest the relevant character flaw" (Sher 2002: 383). While this doesn't imply that one is always excused, and thus never blameworthy, when acting out of character,[16] it at least suggests that appealing to virtue or to a lack of a vice *often* amounts to a good excuse. However, being excused means that one isn't blameworthy but *would* be blameworthy absent excuse. So even if one would often be excused in epistemic-practical conflicts because one's epistemic failure doesn't always

indicate a bad epistemic character in these cases, the epistemic norm violation would still be normatively significant: one *would* be blameworthy absent excuse.[17]

Alternatively, one could hold that epistemic mistakes in epistemic-practical conflicts are never excusable. Instead, they always increase the person's overall degree of epistemic blameworthiness, at least slightly. Whether we should say this will depend on what we should count as an epistemic excuse, and on our specific account of epistemic blame. Yet no matter what we say about these issues, *Dogmatic Dan* already shows that we must reject Agglomeration (see Chapter 4.4): although it's not the case that Dan ought to comply with both the epistemic 'ought' *and* the practical 'ought', he is still blameworthy for not complying with the epistemic 'ought'. I conclude from this that both the epistemic 'ought' and the practical 'ought' retain their normative significance even if it's not the case that we ought to comply with both of them in conflict cases. We therefore must hold that the epistemic reasons that give rise to the epistemic 'ought' in epistemic-practical conflicts are genuinely normative reasons.

5.5 Epistemic Rationality as Evaluative Normativity

Here is a puzzling aspect of my proposed analysis of epistemic-practical conflicts. This aspect can be expressed by the following question: What is the point of blaming people epistemically when they are faced with epistemic-practical conflicts? I will now suggest that in doing so, we mainly express a normative standard of epistemic *evaluation*. Holding each other responsible to this epistemic standard enables forms of epistemic-social relating in response to an agent's cognitive behavior. This reveals the distinctive normative significance of the epistemic 'ought'. After explaining this main idea (Chapter 5.5.1), I will develop it into a sketch of a theory of epistemic normativity by addressing three objections: First, is the epistemic 'ought' genuinely normative, according to my proposal? I argue that it is, despite being 'merely evaluative' (Chapter 5.5.2). Second, can't we still ask what we ought to believe *simpliciter*, or all epistemic reasons and practical reasons considered? I argue that we can't, on pain of otherwise obscuring the normative situation (Chapter 5.5.3). Third, doesn't my view imply that epistemic normativity is ultimately grounded in practical normativity? I argue that it does have such an implication, and yet epistemic norms are normative independently of practical reasons to comply with them (Chapter 5.5.4).

5.5.1 Guiding Actions and Evaluating Doxastic States

The purpose of *practical* norms is not merely evaluation – figuring out whom to trust morally, so to speak. Their function is also, and maybe primarily,

guiding our conduct. A moral demand not to lie to others has significance for how we should conduct our lives. Yet, violating such a demand will also bear on how we should morally evaluate the person who did so. By contrast, the epistemic norm not to believe what is insufficiently supported by our evidence won't have this function of guiding our conduct. Since most of our beliefs are formed as automatic responses to our perceived evidence (see Strawson 2003, and Chapter 2 of this book), there is no point in epistemic norms guiding our doxastic conduct. Our doxastic conduct – when and how we should actively inquire into a matter, for instance – seems instead to be governed primarily by *practical* norms: some issues are worth inquiring into, others aren't, and some deserve more diligence than others.[18] By contrast, the primary purpose of the epistemic norms that govern our *states of belief* is evaluation. While epistemic evaluations will often have consequences for how we should relate to one another in our attitudes and actions (whom to trust, whom to engage with in discussion), they do not *primarily* tell us what we ought to do – for they are (at least *also*) concerned with the evaluation of a mostly involuntary state of belief.

By regarding a subject who violates an epistemic norm as epistemically blameworthy, we make a judgment about how we should relate to them epistemically in our attitudes and actions in future interactions. Having a practice of a distinctively epistemic evaluation of beliefs and believers allows us to keep track of how to engage with whom for our collective epistemic enterprises. These evaluations will provide us with reasons for adopting certain kinds of attitudes and stances toward our fellow epistemic citizens. Constant violations of epistemic 'oughts' by the same agent provide us with reasons to reduce epistemic trust.

To illustrate the position I articulate here, consider how the present account applies to the case described by Crawford (see Chapter 4.3). I agree with Crawford that your advice to the student – that she ought epistemically to believe that her roommate dislikes her but that she ought practically to cause herself not to believe it – seems unhelpful. But this is only because you fail to clarify the significance of the epistemic 'ought'. Its normativity manifests in the epistemic evaluation of agents rather than the guidance of their doxastic conduct. When it comes to guidance, there is only one thing to say to the student: "If it is really better to feel good than to have an adequate conception of your roommate's attitude to you (something you might wish to doubt), then you should (make yourself) believe that she does not dislike you." However, this answer does not exhaust everything that can be said about the student's normative situation. You should additionally make the student aware of the epistemic norms she would violate and clarify the point of these norms: "But note that if this is how you proceed with your doxastic life, you might end up being a wishful thinker, and thus others might have a reason to reduce their epistemic trust

in you: you will regularly violate epistemic norms, thereby affecting your standing within your epistemic community."

The distinction between norms for guidance and norms for evaluation helps the proponent of TV to explain why we *tend to think* that epistemic reasons are irrelevant in epistemic-practical conflicts: we unwittingly switch from the question of what we ought epistemically to believe to the question of how we ought practically to conduct our cognitive life. This is understandable because the focus of epistemic-practical conflicts is intuitively on the conduct-question rather than the question of epistemic evaluation. When asking 'What ought we to believe?', our question is ambiguous between 'How ought we to proceed with our cognitive conduct?' and 'How are we to be evaluated epistemically if we proceed in one way rather than another?'. Both questions are interesting in their own right, and their answers will sometimes point in opposite directions.

I will now flesh out this conception by addressing three pressing issues: whether evaluative normativity can be 'genuine' normativity, whether epistemic and practical reasons can be weighed to derive what one 'ought' simpliciter to believe, and whether the foundations of epistemic normativity are ultimately practical.

5.5.2 Evaluative Normativity as Genuine Normativity

One might worry that, according to the proposed view, the epistemic 'ought' is at best *evaluative* rather than genuinely normative (Rinard 2022: 4, 10). According to this objection, our practice of evaluating epistemic agents in terms of their epistemic blameworthiness is just concerned with descriptive facts about how an agent's belief relates to their evidence, and thus only with evaluating how well they respond to evidence in automatically forming beliefs. One might instead wish to reserve the term 'genuine normativity' for norms with the primary function of guiding our voluntarily controlled actions.

I agree that the epistemic domain might not be 'genuinely normative' in this specific sense: epistemic norms might indeed not guide our actions by which we can influence our beliefs. Yet some now argue that epistemic norms govern our distinctively intellectual actions, such as inquiry, reasoning, or assertion.[19] In any case, the sense of 'genuinely normative' I am interested in here applies to domains that give rise to interpersonal criticism or blame when its norms are violated, and thus to domains that matter for how we should relate to one another.[20] In evaluating someone as epistemically (un)trustworthy, for instance, we don't merely evaluate the person as an unreliable indicator of truth. Rather, we judge their epistemic character as displaying an epistemic vice, thereby epistemically blaming

the person. Furthermore, epistemic norms *can* guide our beliefs, even if it turns out that they don't guide our actions: epistemic reasons can *directly motivate* our beliefs, in the sense that we can base our beliefs on epistemic reasons by taking them to normatively favor a belief we hold (cf. Kiesewetter 2021). These two criteria—liability to personal criticism and direct motivation—characterize normative reasons. So it wouldn't be an obstacle to genuine epistemic normativity if the epistemic 'ought' was ultimately 'merely' evaluative and not action-guiding, but only belief-guiding. Rather, evaluating each other's doxastic states in epistemic terms is essential for figuring out whom to trust and whom to epistemically engage with and how. This is arguably also the central and maybe primary function for epistemically evaluating intellectual actions, such as inquiry or assertion (rather than guiding them).

5.5.3 Against Weighing Epistemic and Practical Reasons

The second worry is that the opponent of TV (see Chapter 5.4) can agree that the epistemic 'ought' has its distinctive normative significance. After all, there are two possible claims involved in TV that they could reject: either that there is a normatively significant epistemic 'ought' that pulls in another direction than the practical 'ought'; or else that there is no answer to what one ought to believe *simpliciter*. Thus, the opponent of TV might just accept the first claim, which I defended here, but then deny the second. That is, they might just grant that there is a normatively significant epistemic 'ought', but then argue that there is *also* an 'ought' *simpliciter* in epistemic-practical conflicts that we gain by giving the epistemic and the practical 'ought' their proper normative weights, thereby determining what we ought to believe, all reasons considered. Let's call the resulting view in this objection the

> *Tripartite Analysis (TA)*. In epistemic-practical conflicts, there is something one ought epistemically to believe, something one ought practically to (cause oneself to) believe, and something one ought simpliciter to believe. Each 'ought' has its own distinctive normative significance.

TA is a version of the weighing view because it claims that the 'ought' simpliciter can be derived by weighing or comparing epistemic and practical reasons. For the sake of argument, I grant the proponent of TA that there might be some plausible mechanism of weighing or comparing epistemic and practical reasons to arrive at an 'ought' simpliciter. My argument will be that, even if there was such a mechanism, we should still be hesitant to introduce a third 'ought' simpliciter alongside the epistemic and the practical 'ought': an 'ought' simpliciter is either superfluous or it even obscures the normative situation.

I will now present this dilemma for TA. The dilemma has the following two horns:

a Either the practical 'ought' and the 'ought' simpliciter always require the same response; then the 'ought' simpliciter is superfluous,
b or the practical 'ought' and the 'ought' simpliciter sometimes require different responses; then introducing an 'ought' simpliciter obscures the normative situation.

Consider (a) first. It is unclear what we gain by adding an 'ought' simpliciter that always algins with the practical 'ought'. We've established that the epistemic 'ought' and the practical 'ought' have normative significance on their own. Given this, an 'ought' simpliciter that never conflicts with the practical 'ought' seems rather superfluous. The view might attempt to express that the practical 'ought' has some kind of *priority* in doxastic dilemmas. I agree that it has priority in the sense that it would be *practically better* to comply with the practical 'ought'. Yet this trivial observation hardly justifies introducing a third normative domain.

If (b) was true, then the 'ought' simpliciter would sometimes go against the practical 'ought'. Christopher Howard (2020: 2234), for instance, argues that we sometimes ought simpliciter believe what is best supported by epistemic reasons even though it would be practically better not to believe it. In his case, borrowed from Kelly (2002), one gains decisive testimonial evidence about the ending of a movie that will ruin one's cinematic experience. Howard claims that one ought simpliciter to believe what the testimonial evidence supports, because the practical costs of belief are not yet above a certain threshold below which only the epistemic reasons are relevant (see also Reisner 2008). Above this threshold, only the practical reasons matter.

This proposal renders the normative significance of the 'ought' simpliciter dubious. For on Howard's view, one's practical reasons can require one to violate the 'ought' simpliciter. Suppose that you have easy means to cause yourself not to believe the truth about the movie ending (say, swallowing a belief-preventing pill). Given that it is practically better for you not to have a belief about the ending of the movie, you ought practically to cause yourself not to hold this belief. Now, what is the point of saying that you still *ought simpliciter* to hold the belief although you ought practically to cause yourself not to hold it? At best, this emphasizes that you will *fail as an epistemic agent* if you follow your practical reasons instead of your epistemic reasons. But this verdict can be captured just by appealing to the epistemic 'ought'. On the other hand, if you decide *not* to cause yourself not to have the belief about the movie ending, although you could easily do so, then you are plausibly *prudentially* criticizable for failing to do so

(you might regret not having taken the pill). Again, no matter how the case turns out, you violate an 'ought' that has normative significance. This normative situation is captured by TV. There is no need for an 'ought' simpliciter.

In reply to this second horn, adherents of TA might argue that there are still cases other than epistemic-practical conflicts in which we need an 'ought' simpliciter to explain the normative situation. Selim Berker (2018: 443–445) mentions cases in which it is practically equally good to disbelieve p or to suspend judgment about p, while one is epistemically required to disbelieve p (since one's evidence against p is excellent). According to Howard (2020), one ought simpliciter to disbelieve p in these cases, even though one's practical reasons leave it open whether to disbelieve or to suspend judgment: now the epistemic reasons can tip the balance toward disbelief. Contrast this verdict with the view I propose. According to my proposal, all we can say in this case is that practically, it doesn't matter whether you suspend judgment or disbelieve, while epistemically, you should disbelieve. But isn't there a need for an 'ought' simpliciter here?

I do not see why. First, these cases are very specific: the practical reasons for two doxastic options are equally balanced while the epistemic reasons only favor one of these options. Introducing a third normative domain next to the practical and the epistemic just because of such rare cases seems exaggerated. Second, the proposed view can capture these cases: you fail epistemically if you suspend judgment about p or believe p, and you fail practically if you believe p. Each failure could make you blameworthy absent excuse or exemption. However, you can just easily comply with both the practical and the epistemic 'ought' *by disbelieving p*. So of course, if you want to avoid being blameworthy either practically or epistemically, then you should ensure, just from a plain practical perspective, that you disbelieve. Maybe you do not care about your epistemic norm violation. If so, then either you are epistemically excused, or your not-caring might reveal a flaw in your epistemic character, thus rendering you epistemically blameworthy. Again, there is no need to introduce a domain 'simpliciter' next to the epistemic and the practical domain. For we can give a satisfactory description of the normative situation without it.

Reverse cases are discussed by Jaakko Hirvelä (2023: 1809–1810), who argues that sometimes believing and suspending are epistemically permissible but one *practically* ought to believe. He claims that there's a further sense in which one *ought* to believe what one practically ought to believe, rather than suspend judgment. Again, I think we can fully capture these cases (if they're possible) by saying that one practically ought (to cause oneself) to believe but one would be epistemically permitted to either belief or suspend. Our blaming-responses will vindicate this result.

We might wonder why a third domain 'simpliciter' should be *more* normatively significant than either the epistemic or the practical domain, given that each of these domains is already significant in its own right: each domain can give rise to blameworthiness when its verdicts are violated, even without normative backup from other domains. The theoretical purpose of introducing a third normatively significant domain that houses the 'ought' *simpliciter* next to the practical and the epistemic domains is dubious.

I conclude that introducing an 'ought' simpliciter that results from weighing epistemic reasons against practical reasons is either superficial because its verdict is always identical to the practical 'ought' (the first horn of the dilemma), or it introduces a dubious *third kind of normativity* that is distinct from practical and epistemic normativity (the second horn). Again, the problem with this third kind of normativity is that its purpose is unclear, given that we have already established that both the epistemic and the practical 'ought' each have their own normative significance, which allows us to explain the normative situation in epistemic-practical conflicts and other cases. In epistemic-practical conflicts, no matter what we do, we fail to live up either to epistemic norms or to practical norms, and we can still end up blameworthy for violating each of them (cf. Kauppinen 2023; Schmidt forthcoming c).

5.5.4 The Practical Foundations of Epistemic Normativity

The final worry is that if epistemic normativity is indeed relevant for the evaluation of believers, then how can it not depend on practical or wrong-kind reasons? Ultimately, we seem to care about our practice of evaluating believers in epistemic terms because this practice has some practical utility (roughly, enabling us to collectively gain important knowledge, or to collectively understand the world better). So shouldn't we just refrain from epistemically blaming people for epistemic norm violations whenever doing so lacks utility, and therefore doesn't the normativity of epistemic reasons again depend on wrong-kind reasons to be epistemically rational?

In response to this, I grant that epistemic normativity must ultimately have a pragmatic foundation. Consider, for instance, David Owens' (2017b) account. He argues that we are justified to engage in our epistemic practices because it is practically valuable to *be subject* to epistemic norms. Owens distinguishes this value from the value of *conforming* to epistemic norms: while it is not always practically valuable to conform to epistemic norms, it is always of practical value *to be subject* to epistemic norms. For Owens, this value consists in enabling a kind of valuable emotional engagement with the world. What matters for our purposes here is his idea that there is practical value in *being subject* to distinctively epistemic norms.

Importantly, a pragmatic justification of our epistemic practices along Owens' lines is compatible with the view that epistemic norms have normative significance even in the absence of wrong-kind reasons to comply with them. To see this, compare Owens' view with Dennett's position concerning the justification of punishment (Dennett and Caruso 2021: 119–127). Dennett argues that a practice of punishment can be externally justified only by consequential considerations, such as deterrence, or resocialization. However, at the same time, he argues that our particular judgments within this practice – about who is to be punished and how – are justified by desert-based considerations, such as who deserves to be punished, or whom it is unfair to punish. Desert-based considerations are *internal* to this practice in the sense that they are used to justify instances of punishing, but consequential considerations are the *external justifiers* for having this purely desert-based practice. That is, even if in a particular case there are no good consequences for punishing an individual, the punishment might still be justified by desert-based considerations.

While I do not claim here that Dennett is right about reasons for punishment, his position will help us to understand how we can accept a pragmatic foundation of epistemic normativity while still acknowledging that epistemic reasons have their *distinctive* normativity. For Owens' position must analogously be read as proposing an external justification of our internal epistemic practice. It *pragmatically* makes sense for us to adhere to a practice where we evaluate our beliefs in distinctively epistemic terms. Combined with my sketch of an account of epistemic blame in the previous subchapters, we can say that there is value in showing reactive attitudes of suspending or increasing epistemic trust in response to violation of or compliance with epistemic norms. The value might lie, roughly, in keeping track of who is a reliable source of information, whom we can share our knowledge with, and whom we can engage in fruitful discussion with, so as to get more reliably to the truth collectively (Williams 2002; Dogramaci 2012). Even if our beliefs are epistemically justifiable only by distinctively epistemic reasons, our overall practice of epistemic evaluation must be justified on pragmatic grounds. For it is hard to see how an external justification of a whole practice could be anything other than pragmatic in nature (in the broadest possible sense of 'pragmatic').[21]

I have argued that epistemic norms are evaluatively normative in the following sense: their purpose is to figure out how to relate to other epistemic agents, depending on how well they fare in complying with epistemic norms. I have furthermore addressed three objections to clarify this picture. First, I have argued that 'genuine' normativity is plausibly the kind of normativity that is tied to personal criticism or blame, and I have granted that epistemic normativity is not directly relevant for our doxastic conduct – here instrumentalists and pragmatists are right insofar as

practical reasons determine how we ought to *manage* our beliefs. What they overlook is the important dimension of normative epistemic evaluation. Second, I have argued that if one accepts the normative significance of the epistemic 'ought', then one should accept TV. For if we instead add a distinctive 'ought' simpliciter to the epistemic 'ought' and the practical 'ought', then we will obscure the normative situation by artificially introducing a third normative domain. Finally, I have argued that my account is compatible with *pragmatic foundations* of epistemic normativity without committing to the view that the normativity of epistemic reasons depends on practical reasons to be epistemically rational in particular cases. Since my focus was on epistemic reasons that are provided by evidence, my defense of TV shows how evidential considerations can provide us with genuinely normative reasons.

5.6 The Refutation of Indirect Voluntarism

Indirect Voluntarists argue that we are blameworthy for belief only if we violate requirements to *cause* beliefs: they argue that responsibility for belief can be made intelligible by the fact that we have indirect control over belief by means of various ordinary activities that have some foreseeable effect on our beliefs (Meylan 2013; Peels 2017). According to these epistemologists, we *never* blame people merely for having a belief that is not properly based on epistemic reasons. Rather, we blame them for their belief only if their belief is culpable, that is, if they failed to manage and influence their doxastic life in the way they should have done: we blame them for their beliefs because, for example, they did not attend *earlier* to certain evidence or did not investigate properly. Responsibility for belief turns out to be completely analogous to the responsibility we have for other consequences of our actions (Meylan 2017). We can fail in our doxastic lives only if we have these lives as a result of not *acting* the way we ought to.

We can now see that these accounts cannot explain *epistemic responsibility*. The requirements of epistemic rationality directly govern *belief*, not (only) action.[22] They require us to properly base our beliefs on our evidence. This can be illustrated by epistemic-practical conflicts. If I can avoid disaster by being epistemically irrational, then I *should ensure* that I violate an epistemic requirement. For failing to be epistemically rational is not a huge cost here – or even not a cost at all. Yet a failure of epistemic rationality is still a failure. It is not a failure to act the way I ought to have acted. For if I'm epistemically irrational in this case, then I still *did* what I ought to have done. That is, I caused myself to commit a *purely doxastic* failure: a failure in believing, not acting. I complied with a practical norm to act but violated an epistemic norm to believe. Given that we are responsible for complying with epistemic norms, as I have argued here by

showing that we can be epistemically blameworthy in conflict cases, these cases show that we cannot explain our responsibility for epistemic rationality by pointing to actions prior to belief. As Owens (2000: 87) has argued earlier, "our control over […] actions yields only an *indirect* control over belief and […] such indirect control cannot underwrite the *direct* responsibility of belief to epistemic norms." This chapter confirms his point.

Indirect Voluntarists must therefore deny that there is such a thing as epistemic blame: we are never blameworthy merely for epistemic irrationality. The most straightforward way to do so is to argue that blame is always a *moral* notion: only reactions like resentment, indignation, or guilt can count as blame. According to this move, the relationship modifications that I discussed above do not amount to genuine blame.

However, given (1) – the connection between normative reasons and blameworthiness – this leaves us puzzled as to why epistemic reasons are *normative*. Indirect Voluntarism cannot explain epistemic normativity: the view must deny that there is a distinctively epistemic blaming practice in which the normativity of epistemic reasons is revealed. For this practice implies that we are sometimes blameworthy for epistemic norm violations even if we had no practical reason to perform actions that make us comply with the epistemic norm (trivial cases), and even if we had decisive practical reasons *against* making ourselves comply with epistemic norms (epistemic-practical conflicts). In these cases, we are not blameworthy in virtue of acting wrongly, but in virtue of believing irrationally. Indirect Voluntarism therefore cannot capture the distinctively epistemic side of our normative lives.

Furthermore, as I will argue in Chapter 7, purely rational failure can even warrant *moral* blame. This is plausible especially in cases where irrationality causes moral harm. This argument will strengthen my case against Indirect Voluntarism, and it will favor an alternative picture of responsibility for attitudes that is broadly rationalist. However, as I argue in the next part, we should also accept a voluntarist *face* of responsibility for attitudes.

5.7 Conclusion

This chapter has replied to the neglected challenge for the normativity of epistemic rationality. I have argued that a person can be epistemically blameworthy for being epistemically irrational both in trivial cases as well as in cases of epistemic-practical conflicts. In both kinds of cases, the blameworthiness results from a failure to respond correctly to epistemic reasons. This reveals the evaluative normativity of epistemic rationality: epistemic reasons matter to us because the way we respond to them reveals our epistemic character, which in turn provides others with reasons

to epistemically relate to us in certain ways (for instance, by modifying epistemic trust). If we allow for concepts of blame in terms of relationship impairments instead of restricting blame to passionate reactive attitudes, we open the space for a face of responsibility for belief that does not presuppose indirect control over beliefs. Indirect Voluntarists cannot account for this *epistemic* face of doxastic responsibility. For here we can be blameworthy for a belief even though we responded correctly to our reasons for *managing* this belief.

This concludes the (narrowly) epistemological part of our investigation. However, Part 3 will still be relevant for epistemology. Chapter 6 applies the discussions from the last two chapters to the rationality of attitudes more generally, and then it carves out the two faces of responsibility for attitudes that also apply to belief. That is, Chapter 6 takes us from epistemology to an ethics of mind. Chapter 7 then argues that rational failure can warrant moral blame. This will also apply to epistemic failures, and thus bring into view a hitherto unnoticed normative force of the requirements of epistemic rationality and other kinds of rationality.

Notes

1 I will just stipulate for my purposes here that all practical reasons are reasons of the wrong kind. There might be state-given reasons of the right kind, and some of them might even be provided by practical value (as defended by Schroeder 2021), but I do not call these 'practical reasons' here. See Chapter 3.2 for more on these issues.

2 See, for instance, Rinard (2015: 219) as well as the references in introduction to Chapter 4 for these views.

3 For some discussion and an argument for subjectivism that builds on the notion of praiseworthiness, cf. Lord (2018: chapter 8). See Kiesewetter (2017: chapter 8) for a good overview of the debate and another case for subjectivism.

4 The denial of permissivism is often discussed as the Uniqueness Thesis, as introduced by Feldman (2007). On epistemic permissivism and some of its problems, see White (2005).

5 However, I will revisit the possibility of moral blame for epistemic failure in Chapter 7. I argue there that, at least sometimes, we are morally blameworthy for mere epistemic failures and other kinds of rational failure. Here my point is merely that the argument in this chapter does not rely on this stronger claim.

6 Similar accounts have been worked out by Kauppinen (2018, 2023), who also presents an account of epistemic criticism as a form of distrust, and by Brown (2020) and Piovarchy (2021), who both share the spirit of Boult and Kauppinen in that they regard epistemic blame as being neither a mere negative evaluation nor a kind of strong reactive emotion (like resentment). Brown mainly builds on Sher's (2006, 2009) account of blame as a disposition organized around a belief-desire pair in order to spell out an epistemic kind of blame, and she accepts that epistemic blame can manifest in rebuking someone for their epistemic failure. Piovarchy utilizes Vargas (2013) view of blame as

having the purpose of agency-cultivation. Among these theorists, only Boult puts relationship modifications at center stage. See Boult (2023) for a detailed discussion and critique of Brown's account.

7 In response to Smith's parent-case, Boult (2023) argues that only reactions that are grounded in the judgment that the person is blameworthy count as blame. However, a problem with this proposal is that it does not tell us what grounds our judgments about blameworthiness. My proposal (that I explain in the next paragraph above) avoids this problem: blame is simply an immediate and appropriate response to the blamee's vice. Furthermore, it seems that the parent in Smith's case regards their child as blameworthy without blaming them. I think Smith's case can be met simply by restricting blaming responses to negative relationship modifications. I take it that Smith's (2013) proposal that blaming responses are essentially expressions of one's moral protest (or, for my purposes, one's epistemic protest) is compatible with my proposal that they are responses to vice.

8 See Hieronymi (2004) for more on the difference between mere reliance and genuine trust.

9 Recently, Magalotti (forthcoming) has argued that the coolness of epistemic blame would make it impossible for us to phenomenologically grasp epistemic blameworthiness. I agree with her argument, and my reaction is to say that epistemic blame and blame for rational failure isn't always cool (see also Schmidt 2024b). This also implies that epistemic criticism at least sometimes amounts to genuine blame (namely, when it isn't cool). For defenses of cool blaming reactions as genuine blaming responses, see the recent works that argue that these reactions count as blame because they go hand in hand with, or consist in, a kind of motivation – a desire that the blamee had not 'believed badly' (Brown 2020), a protest against the blamee's action or attitude (Smith 2013), or just generally the motivation to change one's relationship with them by modifying one's expectations and intentions (Boult 2023). On recent skepticism about epistemic blame, see Smartt (2023). See Boult (2024a) for a convincing reply.

10 I take it that Boult does not use 'epistemic agency' as referring to indirect voluntary control over beliefs. Plausibly, one can be dogmatic, gullible, a wishful thinker, or biased even if this vice was not under one's indirect voluntary control. Rather, beliefs – including irrational ones – are often involuntary responses to one's environment (see Chapter 2, and Strawson 2003). Boult's notion of epistemic agency is more plausibly understood in terms of Hieronymi's (2009b) notion of evaluative control, to which I return in Chapter 6.

11 At this point, the close connection between blameworthiness and vice, or responsibility and character, that I am defending, might seem dubious to some readers. However, I here ask them to be patient until Part 3 of the book, where I return more explicitly to issues of responsibility. The connection will become more plausible there.

12 For statements by proponents of TV concerning a connection between criticizability and the normativity of (epistemic) reasons, see, e.g., Boult (2024c: chapter 1.4), Kauppinen (2018, 2023), Kelly (2003: 628), Kiesewetter (2017: chapter 2), Paakkunainen (2018: 135), and, for statements by opponents concerning such a connection, see Grimm (2009: 253–255), Mantel (2019: 223), McCormick (2020), Rinard (2022: 7), as well as Maguire and Woods (2020)'s distinction between mere 'operative criticizability' and 'robust criticizability' (the latter is at issue here).

13 I here mean a very minimal notion of epistemic agency according to which beliefs are often an agent's responses to epistemic reasons while brute states like headaches aren't. This makes beliefs candidates for things for which we could be directly responsible—rather than merely indirectly by managing our beliefs through actions and omissions. See Boyle (2011) and Hieronymi (2006, 2008, 2009a, 2009b) for more substantive accounts of epistemic agency.

14 Cf. Boult (2024), who analyses degrees of epistemic criticizability as a function of epistemic justification and agent culpability. Boult acknowledges that epistemic viciousness factors into how epistemically criticizable a person is, insofar as reduced culpability for a belief—say, when a person's resistance to evidence is partly explained by trauma—also implies reduced epistemic criticizability. Yet epistemic criticizability is also influenced, on Boult's view, by how epistemically unjustified the belief itself is.

15 Plausibly, we can conceive of two equally vicious people where one of them manifests their vice less frequently in their actions or beliefs because of environmental luck—just think of the coward who is rarely in danger, or the dogmatic person who is rarely confronted with counterevidence to their beliefs. By adding "due to their own agency" in brackets, I put these cases aside. It's controversial whether environmental luck can make one less blameworthy.

16 See Kauppinen (2016) and Sher (2002) on this Humean claim. The absence of vice doesn't always seem to amount to an excuse: especially a wise person might be blameworthy for an unjustified belief because she had more control over it than an epistemically vicious person (see McCormick 2015: 93–94; 103–104; cf. also Sher 2002: 385). However, here we might say that the wise person in one instance still manifests an epistemic vice.

17 The notion of epistemic excuses has mainly played a role in externalist accounts of epistemic justification (see Littlejohn forthcoming; Williamson forthcoming). However, some recent suggestions are compatible with internalism: Worsnip (2021: 162–164) suggests that epistemic irrationality is excusable in cases of cognitive overload (when the amount of evidence cannot be processed) and when one has practical justification not to revise one's credences (say, because the house is burning right now). For reasons given above, I deny that the latter is always an excuse. See Flores and Woodard (2023: 2558, 2561) for similar and further proposals of epistemic excuses.

18 This is not to deny that epistemic norms play a guiding function within inquiry or reasoning. However, even within inquiry, practical reasons can influence to which evidence we should attend, how we should weigh this evidence, and even whether we should inquire diligently or carelessly. Cf. Friedman (2020) for discussion.

19 For an insightful defense of epistemic norms for actions, see Flores and Woodard (2023), who also appeal epistemic criticism. See also Boult (2024c: chapter 5) on how epistemic blame might imply that some actions, such as assertions, are subject to epistemic norms. For some recent pushback, see Arpaly (2023).

20 See also Kauppinen (2018, 2023). I am here motivated primarily by a use of 'normative' that is central to the debate on the normativity of rationality (see Chapter 3).

21 Interestingly, Piovarchy (2021) provides a pragmatic justification of our practice of epistemic blame. Given that, as I have argued, this practice reveals the normative significance of epistemic norms and reasons, the present proposal is nicely complemented by his view about epistemic blame. Various recent social accounts of the sources of epistemic normativity have developed similar ideas,

see Chrisman's (2020, 2022) Hobbes-inspired proposal, Goldberg's (2019) appeal to the legitimacy of social expectations that others comply with certain epistemic norms, as well as Boult (2024b), Dyke (2021), Fleisher (forthcoming), Hannon and Woodard (forthcoming), and Wei (2022).

22 Again, there might be epistemic requirements on, say, actions of inquiry and assertion. I do not wish to deny this. The important issue is that epistemic rationality also or even primarily governs our states of belief.

Part 3

Foundations of an Ethics of Mind

6 A Hybrid Account of Mental Responsibility

> [W]e are not merely producers of our attitudes, or even guardians over them; we are, first and foremost, inhabiters of them.
>
> Angela M. Smith, "Responsibility for Attitudes: Activity and Passivity in Mental Life" (2005: 251)

Let us channel our discussions about epistemic rationality back into the debates on the normativity of attitudes more generally. We saw, in Part 1, that the problem of responsibility for attitudes arises only because our attitudes are responsive to reasons. If attitudes weren't responsive to reasons, then responsibility for attitudes would be on par with responsibility for brute sensations, like pain. That is, our responsibility for attitudes could always be derived from our responsibility for actions and omissions that influence our attitudes. Indirect Voluntarism would be true. However, attitudes seem to be responsive to right-kind reasons, among which are epistemic reasons. Responding correctly to right-kind reasons is not always under our voluntary control, whether direct or indirect. Indirect Voluntarism could not explain our responsibility for responding to right-kind reasons – that is, it could not explain our responsibility for rationality.

Chapter 4 then showed us how we might deny that we are ever responsible merely for being epistemically (ir)rational. A recent movement in epistemology doubts the normativity of epistemic reasons. In this chapter, I will first argue that the cases that motivate this movement can be generalized (Chapter 6.1). This leads to doubts about the normativity of all right-kind reasons for attitudes. If right-kind reasons weren't normative, then we wouldn't be accountable for complying with right-kind reasons. What we are responsible for, ultimately, is only to manage our attitudes properly by responding correctly to wrong-kind reasons.[1] This argument generalizes our problem for the normativity of rationality spelled out in Chapter 4.6. For even if rationality is understood as responding correctly to right-kind reasons, rationality would not be normative if right-kind reasons were not

DOI: 10.4324/9781003382973-9

normative. This constitutes a *prima facie* strong case both in favor of Indirect Voluntarism and against the normativity of rationality.

However, Chapter 5 already turned the tables. It presented a strategy for defending the normativity of right-kind reasons, while focusing on epistemic reasons. The main idea is that we are sometimes blameworthy *merely* for failing to respond correctly to epistemic reasons – that is, we are sometimes blameworthy *merely* for being epistemically irrational. The recent literature on epistemic blame helped us to spell this out: we sometimes modify our epistemic relationships in response to epistemic vices. While the ways we modify our relationships need not amount to *moral* blame, they can often amount to *epistemic* blame. In this chapter, I argue that this account can be generalized to an account of blameworthiness for rational failure: if we regularly fail to have attitudes that are rational, then we often display blameworthy vices in virtue of this irrationality, and relationship modifications are appropriate in response to these vices (Chapter 6.2). This blameworthiness for rational failure reveals that rationality has normative significance. For it shows us that blame is sometimes appropriate even in cases where wrong-kind reasons – those reasons that do not bear on the rationality of an attitude – would require us to be irrational. Therefore, the blameworthiness in these cases must result from failing to respond correctly to right-kind reasons. This in turn reveals that there is a genuine normative requirement to comply with our right-kind reasons.

Finally, this chapter spells out my solution to the problem of mental responsibility by drawing a picture of responsibility for attitudes that arises from this defense of the normativity of rationality (Chapters 6.3–6.6). The picture is in line with recent rationalist accounts that ground our direct responsibility for attitudes in reasons-responsiveness. Yet it allows for a distinct indirect responsibility that is grounded in indirect control. By connecting these two faces of responsibility to the corresponding forms of normativity (right-kind reasons and wrong-kind reasons), the present *hybrid* account also explains the indirect voluntarist's driving intuitions. Chapter 7 then argues that irrationality can warrant moral blame, thus supporting my view that reasons-responsiveness grounds a *genuine* kind of responsibility.

6.1 Generalizing the Challenge

Let us return to our challenge for the normativity of epistemic reasons in Chapter 4. Cases of trivial belief and cases of epistemic-practical conflicts seemed to show that epistemic reasons do not matter in the right way to be genuinely normative reasons. Since epistemic reasons are just one kind of right-kind reasons, we would expect this challenge to generalize to all

kinds of attitudinal rationality. Let us, again, start with an initial formulation of our worry:

1 Normative reasons for attitudes matter for what we ought to believe, desire, feel, intend, etc.
2 Right-kind reasons do not matter for what we ought to believe, desire, feel, intend, etc.
3 Thus, right-kind reasons are not normative reasons.

Since reasons of the right kind are the ones relevant to the rationality of an attitude, (3) implies that rational requirements to believe, desire, feel, intend, etc., are not normative. The important premise for this generalization is premise (2). We can see its initial plausibility if we consider cases analogous to cases of trivial belief and epistemic-practical conflicts – that is, cases of trivial attitudes and conflicts between right-kind and wrong-kind reasons for attitudes.

Consider, first, desires. There is some initial difficulty in finding cases of wholly trivial desires. My desire to scratch my neck because it is itching is not clearly a desire that does not matter. The fact that my neck is itching can be an excellent reason both of the right kind *and* of the wrong kind to desire to scratch my neck. This is because the desire to scratch is both directed at a desirable action *and* might itself have the good consequence that I scratch my neck, thereby getting rid of an unpleasant sensation of itching. One might think that it *always* matters whether I have desires that are supported by right-kind reasons because desires with desirable content help us to achieve what is desirable. It would follow that we cannot construct a case of a trivial desire as we have constructed cases of trivial belief, and thus right-kind reasons for desire might more plausibly be genuinely normative reasons. For it seems that we cannot formulate the same challenge for the rationality of desire as we can formulate for the rationality of belief.

However, note first there are clearly *counterproductive* desires. Suppose that your itching neck provides you with a reason to desire to scratch your neck (it would be pleasant for the moment), but that you should resist acting on this desire, because scratching would just make the itching worse in the long run. Although your desire to scratch is rational (after all, the momentary pleasure *is* a good consequence of scratching your neck), it might be rational for you *to desire not to desire* to scratch your neck. It could even be rational *to get rid of this desire* by telling yourself (falsely) that momentary pleasure is *completely* worthless. Would you be blameworthy if you were successful in getting rid of the desire? Intuitively, it does not seem so. And yet your right-kind reasons to desire to scratch your neck were decisive: the desire is a fitting response to

the momentary pleasure of scratching. Therefore, it seems that right-kind reasons for desire are normatively irrelevant when there are wrong-kind reasons not to have the desire. Thus, even if there were no truly trivial desires, we can doubt the normativity of right-kind reasons for desire by constructing conflict cases.

Furthermore, there *are* truly trivial desires – desires that do not matter *at all* – which are supported by right-kind reasons for desire. Suppose that X is desirable but that what you desire is *impossible*. In such a case, your right-kind reasons are still decisive for desiring X. Suppose, for example, that you know that it is desirable to walk on Pluto: you could enjoy an awesome otherworldly landscape when walking there. Yet, for some reason, you do not desire to walk on Pluto. Are you blameworthy for lacking this desire? After all, you *cannot* walk on Pluto. What, then, is the point of desiring it? Not desiring it might rob you of the pleasure of imagining how nice it is to walk there. But we might well stipulate that you have better things to do than engaging in such imaginative projects, or that they just aren't fun for you. In cases where your desire does not have any benefit, it is unclear why anyone should regard you as criticizable for lacking the desire. Thus, intuitively, right-kind reasons for desire don't seem to matter independently of a practical reason to pursue desiring what is supported by right-kind reasons (that is, desiring what you know to be desirable). This is analogous to the challenge for the normativity of epistemic reasons from Chapter 4, where it seemed that epistemic reasons do not matter independently of a practical reason to comply with one's epistemic reasons.

Next, consider intentions. Initially, we face a similar difficulty as we did with desires. The intention to brush your teeth this morning is not wholly trivial: without it, you would not have brushed your teeth. Similarly, my intention to scratch my neck because it is itching matters to some degree (assuming in this case that scratching isn't bad in the long run, but it is rather what I should do, all things considered). I might plausibly be prudentially criticizable for not having such intentions. It thus might seem that right-kind reasons for intention always matter, and so there are no trivial cases when it comes to intention because right-kind reasons for intention always indicate that having the intention contributes to performing a good action.

To get a truly trivial intention into focus, consider an action which you ought to perform *in the future*. Suppose that attending a conference *in a year* would be the right action for you. The reasons for attending might include, for example, the opportunity for rich academic exchange, for presenting your ideas, and for making important contacts. It is rational for you to intend *now* to attend the conference *in a year*. Yet there is *nothing lost* if you do not yet intend to go to the conference. You might

be *akratic* right now: you know you should attend the conference, but you do not intend to attend. However, it is still a year until the conference takes place, and thus another year until it *matters* whether you intend to attend the conference. You have plenty of time to overcome this akrasia. Since it does not yet matter whether you intend to attend the conference, it seems that you are not blameworthy for being akratic. That is, you are not blameworthy for not intending what you ought to do – that is, for not having an intention that seems to be decisively supported by right-kind reasons. Again, it seems, intuitively, that the right-kind reasons for intention only matter if there is some wrong-kind reason to comply with the right-kind reasons – which is often absent in the case of future-directed intentions.

Is it plausible that you have decisive right-kind reasons for intending to attend the conference that only takes place in a year? Kiesewetter (2017: 190–192) argues that your right-kind reasons are not decisive in this case. He argues that you do not *yet* have decisive right-kind reasons to intend to attend the conference. Rather, intending to attend *becomes* decisively supported by right-kind reasons as soon as you *must* intend to attend to ensure that you *will* attend – say, because now is the time to prepare your travel. At some point, you must form an intention, or else you won't attend the conference.

The main problem with Kiesewetter's view is that it doesn't explain why the intention suddenly becomes decisively supported by right-kind reasons when you must intend in order to perform the right action. Let us assume that, as the conference draws nearer, and you need to take steps in order to ensure that you attend, nothing relevant to the deontic status of your *attending* has changed. No further reasons to attend have appeared on the horizon. Nevertheless, you now have, according to Kiesewetter, decisive right-kind reasons to intend to attend the conference. You would be blameworthy if you fail to intend this, now that you must. But how can right-kind reasons *become* decisive while remaining the same set of reasons? Kiesewetter's view implicitly assumes that right-kind reasons for intention gain their normative force only when there is some wrong-kind reason to comply with them—here, the wrong-kind reason that you must form an intention to ensure that you attend. This amounts to granting that right-kind reasons for intention have no such force *on their own*. I thus conclude, *pace* Kiesewetter, that the future-intention case is analogous to the trivial belief case in relevant ways.

Furthermore, there are conflicts between right-kind and wrong-kind reasons for intention. Kavka's (1983) toxin puzzle is such a case that involves a beneficial intention that lacks support by right-kind reasons. Conversely, there are also cases in which it would be bad to intend to do something but where doing it *is* decisively supported by right-kind reasons. For instance,

you might have a decisive reason to go to the beach tomorrow. But suppose that, if today you intend to go to the beach tomorrow, then you will suffer immensely. This doesn't affect your reasons to go to the beach tomorrow, since you won't suffer in virtue of going there, but rather in virtue of your intention today. You can safely adopt the intention only tomorrow (if you can) and thereby avoid the harm today. Plausibly, you should actively *ignore* your right-kind reasons for intention when you ask what you ought to intend today.

Here the normative import of the right-kind reasons for intention seems unclear. If there are no wrong-kind reasons to comply with right-kind reasons, then, intuitively, right-kind reasons seem to be normatively insignificant. In some cases, there might still be some wrong-kind reasons to comply with your right-kind reasons, but these will often be outweighed by the wrong-kind reasons against compliance. Again, the right-kind reasons don't seem to have any normative significance independently of the wrong-kind reasons to comply with the right-kind reasons. The challenge for the normativity of epistemic rationality seems to carry over to the normativity of the rationality of desire and intention.

Finally, consider cases of various emotions. There can be conflicts between right-kind and wrong-kind reasons for emotion. Sometimes, fearing a danger can be rational but counterproductive. In these cases, it seems that rationality is normatively relevant only insofar as there is something good about being rational. If there is nothing good about being rational – say, one's fear does not help one to avoid danger, and it is rather just disturbing and distracting – then it seems that the mere fact that your fear is rational does not have any normative significance in the situation at hand. Again, it seems that the rationality of fear is only normative if there is a wrong-kind reason to fear rationally.

Things are a bit trickier with emotions that imply pleasure, like happiness. Although here conflict cases can arise (when rationally feeling happy has bad consequences), it is hard to see how feeling happy can lack *any* support by wrong-kind reasons, thus being truly trivial happiness. It seems that one always has a wrong-kind reason to (cause oneself to) feel happy because feeling happy is pleasurable. However, we might imagine a person who is in a depressed mood. For this person, it is *impossible* to make themselves feel happy. If you only have a normative reason to do something if you *can* do it, then this person has no normative reason to make themselves feel happy – that is, they don't have any wrong-kind reason to (cause themselves to) be happy. Now, suppose that they experience a joyful event that rationally requires them to feel happy. In this case, it seems the person has decisive reasons of the right kind to be happy without having any wrong kind of reason to (cause themselves to) be happy. Again, it

seems that the person is blameless.[2] Again, the right-kind reasons seem to be normatively irrelevant by themselves, in the absence of a wrong-kind reason to comply with them.

In the next subchapter, I will argue that some emotions, such as admiration and love, cause trouble for the intuition that right-kind reasons for emotion are normatively irrelevant. However, the purpose of the present subchapter is merely to generalize the challenge for the normativity of rationality. For now, I rest content with having argued that there are cases of desires, intentions, and emotions that are structurally analogous to cases of trivial belief and other cases that are structurally analogous to epistemic-practical conflicts: there can be decisive right-kind reasons for attitudes that are not favored by any wrong-kind reasons (trivial attitude cases), and there can be cases in which the right-kind reasons for an attitude seem to be rendered normatively irrelevant by decisive wrong-kind reasons not to comply with them (conflict cases between right-kind and wrong-kind reasons for attitudes). Thus, the cases can be generalized to all attitudes, and every time, it seems that only the wrong-kind reasons do the normative work, while the right-kind reasons are normatively irrelevant by themselves.

Genuinely normative reasons would sometimes make a person blameworthy if these reasons are decisive and one fails to comply with them (at least in the absence of excuse or exemption). Since a mere failure to comply with right-kind reasons for attitudes doesn't seem to make a person blameworthy, they don't seem to be normative reasons. That is, if rationality was normative, then subjects would sometimes be blameworthy when they lack trivial attitudes that are supported by decisive right-kind reasons, and when these subjects comply with their wrong-kind reasons but not with their right-kind reasons in conflict cases. Again, the following connection between normative reasons and blameworthiness is assumed in the challenge:

> *Normativity and Blameworthiness* (NB). Reasons of kind K are normative reasons only if we can be blameworthy or personally criticizable merely in virtue of failing to respond correctly to decisive reasons of kind K (which we possess, or which are available, or accessible).

That is, right-kind reasons are normative reasons only if we can be blameworthy merely for failing to respond correctly to decisive right-kind reasons (see Chapter 5.1 for a defense). (NB) is compatible with the view that right-kind reasons aren't normative although we are *often* blameworthy when we don't respond correctly to them. This blameworthiness would then just derive from not complying with wrong-kind reasons

that decisively favor compliance with right-kind reasons in these cases (cf. Kauppinen 2023: 141). In such cases, one wouldn't be blameworthy *merely* for not responding correctly to right-kind reasons. Trivial attitude cases and epistemic-practical conflicts seem to show that we cannot be blameworthy *merely* for violating the requirements of rationality. It seems false that we should comply with rational requirements for their own sake. Rather, it seems that we should only comply with them if there is some wrong-kind (or practical) reason to (ensure that we) comply with them.

This is the generalized challenge for the normativity of rationality (see also Chapter 4.6). It arises from cases that are structurally analogous to the ones that question the normativity of epistemic reasons within recent epistemological discussions.

Thus, the challenge for the normativity of rationality is a challenge not only for the epistemic rationality of belief but for the rationality of attitudes in general. I now argue that we can defend the normativity of rationality by appealing to the possibility of blameworthiness for mere rational failure, which I will understand in analogy to epistemic blame. This will require me to explain how we can be blameworthy in the cases of mere rational failure that I have just described. Blameworthiness for rational failure reveals the normativity of rationality.

6.2 Meeting the Generalized Challenge

This time, I will start out by considering blameworthiness for irrational emotions because there are certain types of emotions where it is very intuitive that we can be blameworthy merely for failing to comply with the requirements of emotional rationality – that is, merely for failing to respond correctly to our right-kind reasons for these emotions. This blameworthiness reveals the genuine normativity of the right-kind of reasons for emotion, and thus of emotional rationality, which is, like epistemic rationality, relevant for interpersonal evaluations.

Take *admiration* and *love*. You can be blameworthy merely because you do not *admire* features of the world that are *admirable*, or merely because you do not *love* someone who is *worthy of your love*. I here understand admiration and love as stable dispositions rather than as occurrent feelings. In a case where I am confronted with something admirable (say, I am standing in front of the ancient pyramids) but where I do not *feel* admiration (say, due to my being stressed out from traveling), I am not *necessarily* blameworthy (I might be excused due to stress). Yet I am blameworthy if I do not have a *disposition* to feel admiration for what is admirable under usual circumstances, one that I fail to manifest in appropriate circumstances, and where I lack this disposition not just due to

pathology, stress, or other excusing or exempting conditions. Similarly, if I do not feel love for my partner in a specific moment (say, because I am stressed), this does not *always* make me blameworthy. Yet if I do not have any *disposition* to feel love for them, and so I fail to manifest my love for them regularly, then I am blameworthy for this lack of love for someone who deserves my love.

Again, skeptics about the normativity of right-kind reasons might object that we are not blameworthy *merely* because we fail to respond correctly to right-kind reasons for these emotions. Rather, when something is admirable, or someone is worthy of love, this implies that it is *practically valuable* to feel admiration or to love the person. According to this view, the emotionally irrational are blameworthy only when they have failed to respond to wrong-kind reasons for feeling (or for causing themselves to feel) the relevant emotion. Plausibly, in the cases of admiration and love just described, there are wrong-kind reasons for admiration and love that derive from the practical value of feeling admiration and love. Therefore, the skeptic about the normativity of right-kind reasons for emotions can attempt to trace blameworthiness for emotions back to a failure to respond to the wrong-kind reasons in these cases.

To reply to this legitimate objection, we can describe versions of the cases in which my wrong-kind reasons end up favoring, on balance, *not* admiring or *not* loving the person, and yet I can be blameworthy for lack of admiration or lack of love. These are cases of conflict between right-kind and wrong-kind reasons for these emotions. Suppose that you could bring about an immense amount of good if you successfully cultivate a disposition *not to admire* the ancient pyramids when you are in front of them, or a disposition *not to love* a person who is worthy of your love. Suppose, furthermore that you have a reliable method available for cultivating such a disposition. Clearly, your wrong-kind reasons now require you to cause yourself not to admire and not to love. Still, you will be blameworthy. The inability to acknowledge the admirable and lovable is a defect in character that has specific normative consequences for your relationship with others and thus provides others with reasons to modify their relationship toward you because you fail to acknowledge the admirable and lovable. Why explore the world with someone who cannot admire its admirable features? Why be in a romantic relationship with someone who cannot appreciate a person as being worthy of their love? Even though these defects in character might not warrant responses like resentment, indignation, and guilt, especially when one had excellent wrong-kind reasons for developing them, they still warrant blame insofar as it can be rational not to get involved in specific ways with the people who suffer these defects: certain types of relationships with them are necessarily impaired.

It is important to note that I do not argue for some kind of *elitism* of those who are rational. Saying that relationships with the (regularly) irrational are impaired does not imply that we are never allowed to enter such relationships or that we should not enter other types of relationships with people who are (regularly) irrational. Arguably, we are all (regularly) irrational in certain ways. Rather, my claim is much weaker: that someone is (regularly) irrational is a *pro tanto* reason not to enter certain *types* of relationship with the person. The types of relationship will depend on the kind of irrationality at issue.

Here is a sketch of the general account of blame for rational failure I propose. Seeing a person's disposition to be irrational with respect to a certain topic is to discover their vice. As a result of this discovery, we might legitimately reconsider our relationship with them. We might no longer want to be friends with the person. We might cease to promote their personal projects or not take pleasure in their successes. For we see that they adopt their aims and choose their means or form their beliefs only because they are greedy, weak-willed, cowardly, intemperate, ungenerous, unjust, gullible, or dogmatic. We might also no longer care as much about the opinion of the person whose viciousness we have discovered. We might doubt their judgments because of a general sense of distrust we develop toward them. These are all reactions we can only show to fully responsible beings. This is because we can only have the relationships that are presupposed by these reactions with fully responsible beings. Neither computers nor children can display such defects in character that give rise to the reactions described above. I take this form of blame, insofar as it is legitimate, to be grounded in the person's irrationality. For a fully rational person cannot be blameworthy: they could not do any better while remaining rational. It would be irrational to expect them to do any better, given their epistemic situation.

Let us also apply this account to other kinds of attitudes to illustrate its fruitfulness and broader applicability. In Chapter 6.1, I also discussed fear and happiness. If fear is itself unhelpful or counterproductive, can there still be a sense in which one is blameworthy for not feeling fear? Yes, if one has a disposition not to fear what is fearful, that is, if one is reckless. If fear is a manifestation of this vice, then the fact that one regularly fails to respond correctly to right-kind reasons for fear will provide other people with a reason to modify their relationship. Better not to go on adventure with the reckless person. Similarly, someone who regularly fails to be happy about joyful events might be considered a grouch. The grouch is different from the depressive person, who is, due to pathology, exempted from any rational requirement to be happy. Both depressive people as well as grouches might not always amount to good company. Yet only the

grouches fail to live up to our legitimate expectation that one be rational in one's emotions and thus are blameworthy for their unhappiness.

Concerning desires, I have discussed in Chapter 6.1 a trivial desire to walk on Pluto. Is a person blameworthy for not desiring to walk on Pluto? Again, we can grant that they're not blameworthy if this lack of desire does not indicate a general character flaw. However, if one regularly fails to desire things that are good, one is irrational in one's conative attitude. One's desires are not adjusted to the valuable features of the world. This, in turn, will give other people reasons to distrust one's evaluative capacities. Someone who regularly fails to have rational desires or has irrational desires will not give proper weight to what is important, and thus certain engagements with the person will not be appropriate due to this evaluative deficiency. Again, relationship modifications in response to this vice are appropriate.

Finally, I have discussed cases of future intention and conflicting reasons for intention. In cases of future intention, my intention to attend an important conference in a year seemed *trivial*: having this intention right now seemed valueless; it only seemed to matter that I have this intention as soon as I must take the first steps to make sure that I will attend the conference. Yet, as I have argued in Chapter 6.1, this intention is still supported decisively by right-kind reasons even before I should take these steps. Yet I seemed blameless for not intending it at this point. This seemed to support the view that reasons for intention are not normative.

Again, I can grant that I am not always blameworthy for failing to intend what I ought to do. However, I would be blameworthy if I had a *disposition* never to intend what I ought to do, and if I regularly manifested this disposition. It would be difficult to make any plans with me, to say the least. Intentions manifest in behavior, and if I do not intend to, say, meet up with you tomorrow night although I promised you to meet up with you, then you should not trust me to keep my promise. This is so even if I ultimately were to decide to meet up with you.

One might object that a person who never intends to do what they ought to do until the time of action has come, but then always forms the intention just before they should perform the action, is a completely reliable person and not blameworthy in any way.

In reply, I maintain that we cannot even coherently conceive of a person who *never* intends what she does until the time of action has come. If someone always does what they think they ought to do, then we must ascribe to them the intention to do what they think they ought to do for a significant period of time before the action. Thus, this objection fails because it assumes an unintelligible scenario. A person who fails to have rational intentions is a paradigm case of someone you should not trust. The

person is blameworthy in virtue of regularly failing to be rational in their intentions. This blameworthiness reveals the normativity of the rationality of intention, or of right-kind reasons for intention.

Finally, consider a case of conflicting reasons for intention that is structurally analogous to the toxin puzzle (Kavka 1983): your partner would be happy if you would *intend* to marry them, even though they do not actually want to marry you, and you have reasons not to marry them. If you manage to hold an irrational intention to marry them, you would not be blameworthy – or so it can seem. However, suppose that your intention is stable over time, so that it becomes an entrenched irrationality that leads you to action, like trying to convince your partner to marry you, or even making wedding plans. Surely, such a dispositional intention is blameworthy. Again, you are blameworthy *because* your intention is irrational. For everything you do on the basis of your intention seems to be legitimate in light of your intention: *if* your intention was rational, your actions would be rational as well. It is your intention that you should revise in light of your right-kind reasons. Your failure to respond to these reasons makes you blameworthy: it gives your partner reasons to adjust their relationship toward you.

My arguments here take the same structure as my argument in favor of the normativity of *epistemic* rationality (see Chapter 5.4). I grant that one instance of irrationality can be excused. However, if irrationality becomes dispositional, and thus develops into a character flaw, it gives other people reasons to modify their relationship in response to your vice. This is so independently of the wrong-kind reasons for you not to have this vicious disposition. Even if you have decisive wrong-kind reasons to cultivate a vice, you will still end up having a vice if you are successful in cultivating it, which in turn makes you blameworthy for having this vice, at least if you regularly manifest it in actions or attitudes. While you might not deserve strong reactions of resentment, indignation, or guilt, certain relationship modifications that still count as blame are appropriate.

I thus conclude that we sometimes *are* blameworthy merely in virtue of violating rational requirements. This blameworthiness in turn reveals that there is a normative requirement that is violated when we fail to respond correctly to our right-kind reasons: there is a sense in which we always *ought* to be rational – and in exceptional cases we're blameless due to being excused. The relevant sense of 'ought' is an evaluative sense: we have reasons to modify our relationships toward the person who violates the rational 'ought', which is a way of evaluating them negatively. Since we cannot always decide, directly or indirectly, whether we have rational attitudes, rational requirements do not presuppose that we have voluntary control over our attitudes. They merely presuppose that we are *responsive to right-kind reasons* to such a degree that our attitudes do not count as

cases of severe pathology. As we will see in Chapter 6.3, this amounts to *being answerable* for our attitudes.

This concludes the main part of my defense of the normativity of rationality by appealing to blameworthiness for mere rational failure. At this point, some might still worry whether the kinds of relationship modifications I appeal to are genuine forms of blame. In reply to this worry, I argue in Chapter 7 that mere rational failure can give rise to moral blame in cases where violating a rational requirement leads to moral harm.

6.3 Two Faces of Mental Responsibility

For now, let us consider what the defense of the normativity of rationality implies for our problem of mental responsibility. I will begin by first getting our motivating problem into view before spelling out how the present account of blameworthiness for rational failure helps us to solve it. As we will see, we are responsible for rationality in two distinct ways.

Philosophical thought about responsibility is traditionally structured by taking *actions* and their *consequences* to be the kinds of things for which we are responsible.[3] This way of thinking about responsibility highlights a difference between two modes of being responsible: direct and indirect responsibility. On the one hand, our actions are exercises of voluntary control, and we perform our actions for reasons. Their causal consequences, on the other hand, are not exercises of control, and they are not performed for reasons.[4] Our responsibility for consequences originates in our responsibility for actions which cause them.[5] This allows us to say that we are directly responsible for our actions, but only indirectly responsible for their consequences. Whenever we are responsible for a consequence, our responsibility can be *traced back* to our responsibility for prior actions (see Fischer and Tognazzini 2009).

If we try to put attitudes within this traditional framework, we are faced with the puzzle that motivated the present inquiry. On the one hand, it seems that we lack direct control over our attitudes. In this respect, attitudes behave similarly to mere consequences of our actions. Indeed, it seems undeniable that attitudes *can* be consequences of our actions, and that we thus have, to a certain extent, *indirect* control over them: we can manage our emotions through, say, meditation, and we can form justified beliefs by, for example, proper investigation. Yet it is hard to see what a more direct control over attitudes is supposed to look like. If we lack direct control over our attitudes, we might think that we can be only indirectly responsible for our attitudes – in the same way as we are only indirectly responsible for consequences.

On the other hand, our attitudes are within the 'space of reasons': they are subject to evaluations to which brute consequences of actions could

	X=actions	*X=attitudes*	*X=consequences*
Do we exercise control over X?	Direct control --	? Indirect control	No direct control Indirect control
Is X reasons-responsive?	Responsive to reasons	Responsive to reasons	Not responsive to reasons
Are we responsible for X?	Directly responsible	?	Merely indirectly responsible

Figure 6.1 The peculiar status of attitudes between actions and mere consequences

never be subject. Attitudes cannot only be evaluated as better or worse to have (as consequences can be evaluated as better or worse). Rather, we think of our attitudes as rational or irrational. In this respect, it seems that our attitudes do not behave like mere consequences of our actions: if I fall from the roof due to my carelessness, my broken leg will be my fault, but my broken leg is not irrational. Given our attitudes' presence in the space of reasons, they seem to be much more like actions themselves rather than their consequences. We are tempted to conclude from this second line of thought that we must also be *directly* responsible for our attitudes (*in addition* to sometimes being indirectly responsible for them). This peculiar status of attitudes is represented in Figure 6.1.

Figure 6.1 illustrates how attitudes are both dissimilar and similar to actions. They are dissimilar insofar as they are not themselves exercises of control – or, as it is sometimes put, we seem to lack *direct control* over our attitudes. However, attitudes are also similar to actions insofar as both are responsive to reasons. In this respect, attitudes are quite *unlike* mere consequences of our actions, which are not responsive to reasons. This peculiar status of attitudes creates the question mark in the last row of Figure 6.1: are we directly responsible for them, as we are directly responsible for reasons-responsive entities like actions, or are we merely indirectly responsible for them, as we are merely indirectly responsible for mere consequences? The problem of mental responsibility arose due to this peculiar status of attitudes as reasons-responsive entities – subject to rational requirements – that are not under direct control (see Chapter 2).

We can dissolve this ambiguity by acknowledging two distinct agentive capacities that ground our responsibility for attitudes: indirect voluntary control and responsiveness to right-kind reasons. To illustrate the distinction between the two capacities and corresponding faces of responsibility, let us contrast two cases. In the first case, a person complies with all his

duties regarding the indirect management of his attitudes, but then fails to respond correctly to his right-kind reasons. In the second case, a person fails in his duties to manage his attitudes, but still responds correctly to his right-kind reasons. Here are the two cases:

Caring Todd

Todd was mildly offended, and he reacts with an inappropriate degree of anger. His reaction is irrational and harmful to another person. We learn that Todd has been taking anger-management classes for years, and that he does various exercises in his daily life that were recommended to him by specialists to help him control his anger. Whenever he misses out on a class, and does not do his exercises, he will occasionally show an irrational degree of anger. He knows about this tendency. However, there is no further way for him to control these angry impulses. This week, Todd had to take care of his sick mother. Let us assume that this took up almost his whole time and energy, but it was nevertheless his moral duty. As a result, he could not do his exercises or attend his class without neglecting his moral duty. As a result, he becomes angry.

Devastation

Tony offends Tom. It would be rational for Tom to become angry in response to this offense. However, if Tom gets angry, then Tony would notice Tom's anger and Tony would feel terribly devastated. Tom knows this and he can manage his anger by taking a deep breath. However, Tom does not take a deep breath and instead gets angry. Tony ends up being terribly devastated. Let us assume that it would have been better, all things considered, if Tony didn't feel terribly devastated, and Tom knows this normative fact about the situation.

The difference between Todd and Tom is that they can justify their attitudes in different ways. While Todd can offer reasons for neglecting his anger-management, and thus can *historically justify* himself for being angry, Tom cannot do so. This is because Tom could have easily managed his anger but failed in making an effort of will (that is, taking a deep breath instead of allowing himself to become angry). Thus, Tom is not historically justified, but Todd is historically justified. Yet Tom might justify his anger in *another sense*. For Tom's anger is a rational reaction to the situation. Tom seems thus to be justified in the *answerability-sense* of responsibility. That is, Tom can justify his anger by giving sufficient right-kind reasons

for his anger: he was offended by Tony. Todd, by contrast, cannot do so, because his anger is disproportional to the situation, and he is thus not justified in his anger in the answerability-sense.

This allows us to distinguish between *two faces of blameworthiness*: the direct blameworthiness$_A$ for attitudes (answerability-blame) and the indirect blameworthiness$_H$ for attitudes (the blame of historical responsibility) which can be distinguished from each other by the following conditions which are necessary for one kind of blameworthiness, but not for the other

> *Blameworthiness$_A$*. S is *blameworthy$_A$* for their anger only if S's anger is irrational given the offense – that is, only if the anger is insufficiently supported by right-kind reasons for the anger.
>
> *Blameworthiness$_H$*. S is *blameworthy$_H$* for their anger only if S did not perform or omit an action S ought to perform or omit, and only if performing or omitting this action could have avoided or changed the anger – that is, only if S's performing or omitting the relevant action is insufficiently supported by wrong-kind reasons for the anger.

Given these definitions, Todd is blameworthy$_A$, but not blameworthy$_H$, whereas Tom is blameworthy$_H$, but not blameworthy$_A$. Thus, the cases help us to tease apart two faces of responsibility or blameworthiness. This implies that there is indirect responsibility for attitudes (due to our direct responsibility for actions by means of which we sometimes control our attitudes) and direct responsibility for attitudes (due to our answerability for reasons-responsive attitudes): If you are blameworthy$_A$, then this means that certain forms of blame are appropriate in virtue of your *answerability* for attitudes (that is, direct responsibility for attitudes). If you are blameworthy$_H$, then this means that certain forms of blame are appropriate in virtue of your *historical responsibility* for attitudes (that is, indirect responsibility for attitudes). In the current literature on responsibility for attitudes, especially Pamela Hieronymi and Angela Smith, who have recently elaborated on rationalist views that explain direct responsibility for attitudes, seem to be committed to these two different faces of being blameworthy for attitudes.

According to Smith, both kinds of responsibility are, fundamentally, matters of *being answerable*. If you are responsible for something, then you are in a position to *respond* to requests for justification (Smith 2005, 2015a). The requests for justification are then *appropriate*, *intelligible*, or *correct*. We can ask people for their reasons for believing that there will be nice weather tomorrow, for their reasons for intending to go to the concert, or for their reasons for being angry about this silly remark. When doing so, we do not merely request a *causal* explanation of how those mental

states came into existence. Nor do we ask for the reasons the people might have had to bring themselves into the relevant states by previous actions. Rather, we are asking for reasons which *justify* the mental states in question. These requests are appropriate because the mental states reflect – or are supposed to reflect – our rational judgment about what is true, worthwhile, right, and so on (Smith 2005: 250–264, 2015b: 103–104).

Hieronymi similarly argues that in holding an attitude (as well as in acting), we ordinarily reveal our answer to a certain question (Hieronymi 2006: 53–54, 56, 2014: 12–17): in believing that p we reveal our answer to the question of whether p is true; in intending to do A we reveal our answer to the question of whether to do A;[6] in desiring x we reveal our answer to the question of whether x is good (in at least one respect); in fearing x we reveal our answer to the question of whether x is dangerous. We reveal, as Hieronymi claims, an important piece of our mind, or the *quality of our will* (Hieronymi 2014: 15). Thus, those attitudes by their very nature imply that we are answerable for having them, that a request for justification is appropriate, intelligible, or correct. Indeed, these attitudes just *are* our answers to the relevant questions. In contrast to mere headaches, or in contrast to attitudes that were implanted by an evil scientist, they tell us something important about ourselves, about our overall outlook on the world and our place within it. This is similar to Smith's condition of judgment-reflection.

Both Smith and Hieronymi thus claim that there is no need to suppose that we have voluntary control over attitudes, direct or indirect, in order to explain why we are responsible for them. Rather, for them it is *rational control by judgment* (Smith) or *evaluative control by 'answering a question'* (Hieronymi) that grounds our direct mental responsibility.

Importantly, I think that their accounts should allow for *two faces of responsibility*. Consider again the case of responsibility for having a headache. We could be responsible for our headache insofar as we caused it by previous actions or omissions and could, at the time of the actions or omissions, be expected to foresee that the headache would occur as a consequence of our actions or omissions. Thus, I might have omitted to take a pill which would have prevented the headache. We can say that I had *managerial* control – that is, indirect voluntary control – over my headache, because I had the ability to manage whether I would have it by attending to the regularities of the physical world and using them for my aims:

> We exercise this sort of control when we manipulate some ordinary object to accord with our thoughts about it. We typically control ordinary objects by performing intentional actions which affect that object

> in the way we intend. Of course, our control over such objects is never unlimited. We are subject to physical and temporal limitations, to limitations of skill, and to luck. Importantly, the degree to which we exercise control over some object is measured not by the absence of such limitations – as though we would exercise greater control over our coffee cups if they did not obey the laws of physics – but rather by our ability to navigate, manage, and make use of those limitations in order to accomplish our purposes. In fact, in many cases, exactly those features that seem to limit our control also make such control possible. We can control ordinary objects at all only because we know they observe certain regularities.
>
> (Hieronymi 2006: 53)

We have the same kind of control over our mental states. We can learn the regularities by which we come to have, for example, anger, and then set out to manage this anger by, for example, avoiding certain situations, or engaging in meditation. The fact that we can manage our anger through actions makes us *indirectly* responsible for our anger. Or, to put it in Hieronymi's terminology, the anger *falls in our jurisdiction*.

It is important to see, however, that this is not the only sense in which we can be responsible for our anger if we adopt Hieronymi's (or, for that matter, Smith's) conception of responsibility. We are also *directly* responsible for our anger insofar as it embodies, or reveals, our answer to the question of whether there are good reasons to be angry (for example, that somebody offended us). The fact that our anger falls into our jurisdiction makes us responsible for our anger in one sense; the fact that it embodies our answer to the relevant question, or reveals our judgment about right-kind reasons, makes us responsible in another sense:

> Our responsibility for and agency with respect to the relevant actions and attitudes thus have two distinguishable aspects: we are answerable for them, insofar as they embody our answer to certain questions, and they fall into our jurisdiction, insofar as we are expected to manage and control them through our actions.
>
> (Hieronymi 2014: 24)

The most common worry with a view that allows for these two faces of responsibility is that answerability is not a genuine kind of responsibility. However, note that I argue here that right-kind reasons *are genuine normative reasons*: not responding correctly to right-kind reasons can give rise to genuine blaming responses. If this is right, then answerability *is* a kind of responsibility. To see this clearer, let us consider the two kinds of

attitudinal normativity. This will give more substance to my hybrid solution to the problem of mental responsibility.

6.4 Two Concepts of Mental Normativity

Introducing two faces of responsibility will lead to cases in which it seems intuitively unclear how we are supposed to evaluate the overall blame- or praiseworthiness of a person. We seem to end up with scenarios where a person is blameworthy in one sense, but not in another. I gave examples of such cases in Chapter 6.3 (*Caring Todd* and *Devastation*). This might be a puzzling implication of accounts that allow for both direct and indirect responsibility for attitudes. Is there no answer to whether the persons in the relevant cases are blameworthy, all things considered? And is there no answer to what attitude they *ought* to have, all things considered?

The account of blameworthiness for rational failure I wish to defend can answer these questions, thereby capturing both rationalist and indirect voluntarist intuitions. We lose sight of the distinctive normativity of rationality if we deny that the protagonists in *Caring Todd* and *Devastation* each violate a normative requirement while still complying with another. Although Todd complies with his practical requirements of attitudinal management, and thus with his wrong-kind reasons for emotion, he fails to comply with a rational requirement to respond correctly to his right-kind reasons for emotion. Todd is displaying an emotionally vicious character that makes him blameworthy to some degree. And although Tom displays emotional virtue by being appropriately angry about an offense – that is, by responding correctly to his right-kind reasons for emotion – he is still violating a practical requirement by failing to manage his anger, which affects his blameworthiness.

Note that our two faces of responsibility – historical responsibility grounded in indirect control and answerability grounded in direct reasons-responsiveness – are connected to right-kind and wrong-kind reasons for attitudes. Our answers to requests for justifying our attitudes will, depending on the *kind* of request, mention different *kinds* of reasons: when asked why I failed to manage my anger, I can only justify myself by citing wrong-kind reasons for managing my anger; when asked why I am angry, I will ordinarily justify myself by citing right-kind reasons for being angry. The first kind of answer could contain my reasons for my omission of anger-management. I might, for example, say that I had better things to do than to avoid the relevant situation where I knew I would feel angry (my mother was sick, I had to take care of her, and thus I had no time to attend to avoiding those situations), or I might say that I was justified in getting myself into a situation where I would become angry (say, taking

part in a protest against forms of injustice that make me angry). By citing wrong-kind reasons, I could justify being angry in this historical sense. The second kind of answer, however, might include mentioning that I was badly offended by someone, or that injustice is happening that rationally *should* make us angry, and such an answer might thus rationalize my anger by giving right-kind reasons. This kind of answer can justify me in feeling angry in the answerability-sense.

According to the view I propose, there is no univocal answer to whether one is *justified* in feeling angry, all things considered, in our cases. Our subjects are justified only in one sense, but not in the other. In *Devastation*, Tom failed to manage his anger, and he might be morally blameworthy for this failure, given that he knows that Tony would feel terribly devastated as a consequence of being a target of his anger. Others might feel indignant about Tom and tell him that he should have just taken a deep breath to manage his anger. This could be true although Tom's anger is justified by right-kind reasons: Tom's failure to respond correctly to his wrong-kind reasons for anger by managing his anger properly grounds his moral blameworthiness. Similarly, *Caring Todd* might still be blameworthy for feeling irrational anger, although he managed his anger to the best of his abilities: he responded correctly to wrong-kind reasons but not to right-kind reasons and is thus blameless in the answerability-sense.

However, the fact that there are two kinds of *justification* for an attitude doesn't imply that the two faces of blameworthiness at issue are fundamentally different. Indeed, it is plausible that blameworthiness$_H$ and blameworthiness$_A$ *both* affect the emotional intensity of appropriate reactive attitudes. This observation leads me to endorse a hybrid view.

Intuitively, we seem justified in feeling a bit indignant about Tom, who could have just managed his anger instead of making another person feel terribly devastated. Todd, by contrast, made an effort to manage his anger to the best of his abilities, but he still ended up being emotionally irrational. According to the account of blameworthiness for irrationality that I have proposed, it is appropriate to modify one's relationship toward Todd in certain ways. These relationship modifications need not amount to passionate blaming emotions, like indignation, and so there is a sense in which Todd might be *less blameworthy than* Tom. The fact that Todd did his best to manage his mind should, at the very least, have some mitigating effect on our blaming responses. I will return to this issue in Chapter 7.2, where I argue that a person like Todd *could* still be morally blameworthy despite this mitigating effect of his efforts in attitudinal management on our blaming responses. However, this won't change the fact that both right-kind as well as wrong-kind reasons bear on the same *concepts* of responsibility

and blameworthiness. Although answerability and historical responsibility are grounded in different agential capacities, they do not always warrant fundamentally distinct reactive attitudes.[7]

So I maintain two claims. First, there is an ambiguity in our concept of attitudinal *justification* – attitudes can be justified indirectly or directly, by wrong-kind reasons or by right-kind reasons, and no justification is reducible to the other. Instead, each kind of justification has its own normative significance: wrong-kind reasons serve to guide our mental conduct, while right-kind reasons serve for character evaluation. This is an implication of my argument that right-kind reasons have their distinctive normative force. Second, however, this does not imply that blameworthiness$_H$ and blameworthiness$_A$ are incommensurable forms of blameworthiness. Instead, our reactive attitudes seem to be appropriately affected by both kinds of justification, and our overall blameworthiness for an attitude seems to be a function of how well we do in responding to our right-kind reasons *and* in responding to our wrong-kind reasons.

Thus, we can make sense of the two *faces* of responsibility by acknowledging two distinct *concepts* of normativity. While there is only one concept of responsibility, its two faces give rise to two corresponding concepts of normativity: there are reasons that guide our conduct and reasons that are relevant to character evaluation, and thus to how we ought to relate to one another in our attitudes and actions.

I will now address two worries that one might have with the view I presented here. First, aren't we sometimes blameworthy due to *objective wrongs*, rather than in virtue of failures of our reasons-responsiveness? That is, maybe our capacity for reasons-responsiveness is too narrow to fully explain mental responsibility. Second, is our capacity for indirect control at all relevant to explaining mental responsibility? That is, can't we explain all our intuitions merely by reference to the degree of reasons-responsiveness of an attitude? I start with the first worry before turning to the second. My reply to the second worry will further support my claim that we should endorse a *hybrid view* of mental responsibility, rather than grounding all mental responsibility merely in reasons-responsiveness.

6.5 Objective Reasons and Blameworthiness

One might argue that not all blame is grounded in a failure to respond to reasons. For sometimes reasons are inaccessible to a subject, and yet they might fail to provide a satisfying answer to our request for reasons, given that relevant reasons are available to us, but not to them. In these cases, one might think that a person can be blameworthy$_A$ without failing to respond correctly to their reasons, since these reasons were inaccessible to

them. Consider again a case of someone who could not have reasonably avoided ending up with a problematic attitude:

Ronja the Racist

Ronja was raised in a racist community. The people she lived with regularly made racist remarks and did not allow opposing opinions to co-exist within their community. Furthermore, Ronja never had any contact with people outside her community and barely had access to general education. As a result of her unfortunate history, Ronja holds racist attitudes.

How would the answerability-theorist describe this case? First, they would point out that since Ronja holds a full-blooded racist attitude, she is answerable for it: we can request her reasons. If she cannot give an adequate justification, she seems *blameworthy*$_A$. Yet Ronja could not have reasonably managed her racist attitude. She is not blameworthy for any actions that led her to this attitude. It seems that answerability-theorists must admit that she is *blameless*$_H$ for her racism. We can imagine Ronja as being rational: her racist attitudes would be justifiable by her (apparent) right-kind reasons, at least from her own distorted perspective.[8] We might think that she is then still blameworthy$_A$ because she is unable to provide a justification that is acceptable *from our perspective*. This would imply that someone who is rational can still be blameworthy. It would imply *externalism* about mental blameworthiness.

However, it is hard to see how someone could be blameworthy if they responded correctly to the reasons that were accessible to or possessed by them. Being blameworthy or criticizable in the sense at issue seems to presuppose that we failed in our *own* lights – that is, that we failed to comply with the reasons accessible to or possessed by us. We can see this by considering how the reactions we show toward *rational Ronja* aren't the same as the ones we show toward *irrational Ronja*. If there is indeed a rational racist in our world, then this person will have a radically different set of beliefs (or *Weltbild*) from ours. For given the facts that we know, holding racist beliefs is irrational. Thus, due to the radically different set of beliefs of the rational racist, it would be appropriate to adopt Strawson's (1962) *objective stance* toward them: they become a problem to be dealt with rather than someone who is blameworthy. If our reasons for rejecting their view do not move them to reconsider their racist attitude, then we should deal with them in non-rational ways – not by reasoning, but by ensuring in some way that they do not harm others or spread their attitude, and by changing the societal conditions in which rational racism came

about. The latter would amount to engaging in *epistemic environmentalism* (Ryan 2018).[9]

Thus, being blameworthy presupposes that one fails in one's *own* lights – it requires that one fails to respond correctly to one's *own* reasons. Even if Ronja was raised within a racist community, which partly explains why she holds the racist attitudes she holds, she might still have sufficient right-kind reasons available to her that render her attitudes irrational. We might then conceive of her case as a case of 'social implantation' of an irrational attitude – that is, as analogous in certain respects to the implantation of an attitude by science-fiction surgery. An important difference, however, is that Ronja's irrationality is part of her overall character that developed over the course of her lifetime, while a literally implanted attitude is like an alien intruder in one's mind that occupies it suddenly.[10] Although Ronja is blameless_{H}, she could still be blameworthy_{A}. Thus, since a person is blameworthy_{A} only if they are irrational, and not if their attitude lacks sufficient support by objective reasons that are inaccessible to them, $\text{blameworthiness}_{A}$ is grounded in a failure to respond to accessible reasons. This is a commitment to *internalism* about rationality and blameworthiness. If we would go externalist about rationality and blame, then we run danger of changing our topic.

This also helps us to capture another case discussed by Smith – the case of the abused criminal. She argues that a criminal who was abused as a child

> is fully answerable for, and accountable to us for, his crimes, but the fact that we as a moral community allowed him to suffer such terrible abuse as a child is part of the 'answer' he is likely to give when we demand that he justify his conduct; while such an excuse does not get him fully off the moral hook, it should make a difference to how we respond to him morally.
>
> (Smith 2015b: 114–115)

The case of the abused criminal is similar to Ronja's (and Todd's) case: we can conceive of the case in such a way that the criminal had no reasonable opportunity to avoid his crime, as Ronja (or Todd) had no reasonable opportunity to avoid her (or his) irrational attitude. Yet it seems that we do not want to let him 'fully off the moral hook', and this is not merely for pragmatic reasons (say, in order to deter others). Suppose his crime was a cruel murder. His actions, even though they might have *seemed* justified to the criminal himself at the time he committed the murder ('I have a right to pay back society for what they have done to me'), were still genuine activities located within the space of reasons, which he conducted intentionally, and which might *in fact* have lacked support from his accessible or possessed reasons. Given these assumptions about the case, we still want

to regard the criminal as blameworthy for his attitudes even if he lacked any reasonable opportunity not to commit his crime: his intention was irrational, and this irrationality can make it legitimate to react with moral blame, as I will argue in Chapter 7.

6.6 The Relevance of Indirect Control

Here is a worry about the significance of indirect voluntary control for our responsibility practices. Maybe indirect voluntary control is *not at all* relevant for our blaming responses. Instead of granting two different agential capacities that ground two faces of responsibility, as I have argued we should, Owens (2000, 2017a: intro.) argues that our responsibility for beliefs, desires, and emotions is *wholly* a matter of our responsiveness to (right-kind) reasons.[11] He argues that in cases where a person lacks indirect control over their attitude, it is usually also the case that their attitude is not fully responsive to reasons. In such cases, the blameworthiness of a person for their attitude is indeed diminished in some way, but this is due to the lack of reasons-responsiveness, rather than due to the lack of indirect control. Thus, Owens argues that

> [t]he rage of someone terribly abused as a child is less resented than that of a person with a normal upbringing, and temperance in such a person is the more admired. But this should not be taken to indicate that his culpability somehow depends on the degree of control, direct or indirect, which he is thought to exercise over his anger. Rather, what gives us pause are doubts about whether this person's emotions are responsive to reasons at all, given his unusual upbringing.
>
> (Owens 2000: 120)

The problem with Owens' proposal is that we can imagine cases where a person's attitude was non-culpable – where the person lacked opportunities to engage in reasonable exercises of attitudinal management – and yet the attitude is fully responsive to reasons. If reasons-responsiveness was all that is relevant here, then the blameworthiness of this person would not be diminished by their lack of indirect control. But this is implausible. Instead, if someone lacks any reasonable opportunity to manage their irrational attitude, or if they have tried to avoid their irrational attitude by any means possible but failed, then our blaming responses toward this person should be moderated.

Take someone who tries to improve on their irrational angry outbursts. They try very hard, and they spend a lot of their free time in therapy, group sessions, meditation, and so on. Yet they end up having one of their

irrational outbursts again. As Owens (2000: 118) himself notes about a similar case, "one's reproaches would be tempered by a knowledge of [their] efforts at self-improvement". Yet this can hardly be explained by the fact that the attitude becomes less responsive to reasons in the process of self-improvement. If anything, the attitude becomes *more* responsive to reasons: if the efforts of the person were successful, then there will be more counter-factual scenarios in which the person would not have had one of their angry outbursts: they are responsive to a wider range of reasons against anger.

Thus, both reasons-responsiveness and indirect voluntary control are relevant for the blameworthiness of a person for their attitudes. Although we can distinguish between two faces of blameworthiness by pointing out that two distinct agential capacities ground blameworthiness, there is still just one concept of blameworthiness or responsibility involved. This is because the blaming-reactions involved in blameworthiness$_A$ are not categorically distinct from the blaming-reactions involved in blameworthiness$_H$: as I will argue in Chapter 7, both kinds of blameworthiness can give rise to forms of genuine moral blame, as well as to relationship modifications. Blameworthiness$_A$ can also give rise to *sui generis* blame, such as reducing epistemic trust (see Chapter 5). However, since epistemic failure can also warrant moral blame (as I argue in Chapter 7), epistemic blame and other forms of rational blame are located on a continuum, rather than being categorically distinct. This preserves the unity of our concept of responsibility while allowing for two concepts of normativity.[12]

6.7 Conclusion

This chapter has brought us from epistemology to a broader ethics of mind. We saw that the challenge for the normativity of epistemic rationality applies to all kinds of attitudes that allow for a distinction between right-kind and wrong-kind reasons, including also desires, intentions, and various emotions. However, we then saw that the strategy with which I defended the normativity of epistemic rationality in Chapter 5 also applies to the rationality of attitudes other than belief. If we allow for a broad concept of blameworthiness in terms of relationship impairments, irrationality can be blameworthy even in trivial cases as well as in cases of conflict between right-kind reasons and wrong-kind reasons. More precisely, holding irrational attitudes is blameworthy insofar as it is a manifestation of a general disposition to be irrational – that is, a vice.

Responsiveness to right-kind reasons is sufficient to ground a kind of mental responsibility. For there is a sense in which we can legitimately

blame each other for not responding correctly to right-kind reasons. However, mental responsibility also has another face: our historical responsibility that derives from our ability to indirectly manage our attitudes. Both abilities are distinct, insofar as they presuppose different capacities of responding to quite different kinds of reasons: responding to right-kind reasons by exercising reasons-responsiveness, and responding to wrong-kind reasons by exercising indirect voluntary control over our mind. Yet both abilities are relevant for evaluating the overall blame- or praiseworthiness of a person for holding an attitude. We are thus confronted with two *faces* of responsibility for attitudes and two *concepts* of mental normativity. Our initial problem dissolves as soon as we notice that only historical responsibility to the norms of prudence and morality requires indirect voluntary control, while direct answerability to rational requirements merely presupposes reasons-responsiveness.

At this point, we might still be skeptical about whether the kinds of relationship modifications that I have appealed to in order to make direct responsibility for rationality intelligible amount to genuine blaming responses – maybe epistemic blame and other forms of blame for irrational attitudes are not *genuine* blame. This could lead us to doubts about whether direct answerability for attitudes amounts to a *genuine* kind of responsibility. The next chapter will address this worry by arguing that irrationality can often give rise to genuine moral blame.

Notes

1 Note that, if certain forms of pragmatism were true, namely those that claim that practical reasons *directly* bear on what we ought to believe, desire, feel, or intend (what I labelled 'traditional pragmatism', see Schmidt 2022), then Indirect Voluntarism would run into trouble. For then we might be *directly* responsible for complying *with our wrong-kind reasons*: they would not merely imply indirect responsibility for *managing* our attitudes, but direct responsibility for *holding* attitudes. This is why it is essential for Indirect Voluntarism to argue that we lack direct voluntary control over our attitudes, thus blocking direct responsibility for responding to wrong-kind reasons for attitudes. See for instance the arguments against direct voluntary control over belief in Meylan (2013).
2 Note that this is not because the person is exempted. A depressed mood does not exempt you from rational requirements to be happy in the face of happy events (as depression would). Rather, the person seems blameless because they have no wrong-kind reason to make themselves happy.
3 Take one of the central questions of normative ethics: what are the right-making features of an action (e.g., its consequences, motives, or properties of the action itself)? Furthermore, in the debate about free will and determinism, we ask whether we are free with respect to and thus responsible for our actions

and their consequences in a way that is compatible with the causal structure of the universe. Thinking about our responsibility for and control over *attitudes* (instead of actions and their consequences) might well give us a clue as to how to solve this traditional problem (cf. Hieronymi ms; Wagner 2015). However, I am not here concerned with this issue.

4 I here exclude other actions from the consequences of an action. Consequences in my sense are *mere* consequences which are caused by previous actions, but which are not themselves voluntarily controlled. It is important that I include attitudes in the class of potential consequences of actions. I do this by saying that attitudes are not *performed* for reasons (although one might say that they are *held* or *formed* for reasons).

5 The claim is that *if* we are responsible for a consequence of our actions, then we are so in virtue of the fact that we are responsible for actions which caused it.

6 The same holds for *intentionally doing* A (Hieronymi 2014: 15), which implies having the relevant intention.

7 I here deviate from Randolph Clarke (2023), who restricts direct *accountability* to actions and so only allows for indirect accountability for attitudes (i.e., he argues for a version of Indirect Voluntarism, see Chapters 1 and 2). Yet Clarke allows that there is direct *answerability* for attitudes which, according to him, doesn't amount to accountability. I have also earlier committed to such a view in Schmidt (2020a). For reasons I give above and especially in Chapter 7, I now hold that there's no strict distinction between rational answerability and moral accountability (although there *is* a strict distinction between norms of rationality and moral normativity). Rather, we're rightly held to account for our answers to questions (i.e., for our reasons-responsive attitudes). I here avoid drawing the distinction between accountability and mere attributability – which are Gary Watson's (1996) 'two faces of responsibility' – precisely because it tempts us to draw strict distinctions where there are none. However, see Luvisotto (2022) for a recent defense of attributability as a kind of responsibility that's not accountability.

8 On the 'rational racist', see Basu (2019).

9 Meylan (2018) describes a similar case that is supposed to create a problem for reasons-responsivist views of doxastic responsibility, in which an indoctrinated person seems to be responsive to reasons but not responsible. My reply is that for us, the objective stance is indeed appropriate towards this person. However, this is not because she isn't responsible for her views, but because her views are so radically different from ours that we have trouble making sense of her apparent reasons and ways of justifying her beliefs. People within the community that indoctrinated her might still appropriately blame her for irrationality, given their shared system of beliefs.

10 Yet I think that also in the latter case, we would hold the person responsible for their attitude insofar as it is responsive to reasons after the implantation: we can be answerable for implanted attitudes.

11 To be responsible for our intentions, Owens argues, we need to have 'reflective control' over them, which he distinguishes from mere reasons-responsiveness. The view presented here, by contrast, takes it that reasons-responsiveness is sufficient for being directly answerable for all kinds of attitudes: as soon as one can intelligibly request a person's reasons, their attitude might turn out to be irrational, and thus potentially subject to blame.

12 I here deviate from my earlier view in Schmidt (2020a), where I have argued that we can solve the problem of mental responsibility by distinguishing two *concepts* of responsibility (see also Clarke 2023, and my brief discussion of his view in endnote 7). The reason for this deviation is the argument in Chapter 7 that I have developed afterwards, which shows us that the different blaming responses that earlier allowed me to distinguish two concepts of responsibility turn out to be on a continuum (see also Schmidt 2024b).

7 Moralizing Rationality

> It matters morally what we are for and what we are against, even if we do not have the power to do much for it or against it, and even if it was not by trying that we came to be for it or against it.
>
> Robert Marrihew Adams, "Involuntary Sins" (1985: 12)

Our investigation began with the question of how we can be responsible for our attitudes. It seemed intuitively puzzling how responsibility for attitudes can be anything but indirect. After all, attitudes are mental states, and the way we control being in states is by exercising indirect voluntary control over them. However, we saw that Indirect Voluntarism faces a serious problem: attitudes can be held (directly) for reasons. If the reasons for which we hold attitudes are normative reasons, then we are sometimes blameworthy merely for failing to comply with decisive normative reasons for attitudes. Indirect Voluntarists must therefore deny that we can hold attitudes for normative reasons, since allowing for normative reasons for attitudes would commit to direct blameworthiness for attitudes. A promising route for Indirect Voluntarists was to draw on recent doubts about the normativity of epistemic reasons, and to generalize these doubts to all right-kind reasons for attitudes. If right-kind reasons would turn out not to be normative reasons at all, then the fact that we can hold attitudes for right-kind reasons would no longer be an obstacle to Indirect Voluntarists: rationality could not possibly place any normative requirements on us to *be* rational; rather, we would only be required to *ensure* our rationality by indirect means whenever we are practically required to do so.

However, I have defended the normativity of right-kind reasons throughout the last two chapters. Failing to respond correctly to right-kind reasons – being irrational – can, all by itself, give rise to legitimate forms of blame. Drawing on the recent debate on epistemic blame, I have first argued that some forms of blame are *sui generis*: we can rightly modify our relationships with people who display vices due to their irrationality.

DOI: 10.4324/9781003382973-10

In this chapter, I will take a more radical step, arguing that irrationality can be a proper basis for genuine moral blame. I argue that this is revealed within our practice of apology and forgiveness, where we sometimes authentically apologize for moral harm even though the only norm which was violated was a rational requirement. Taken together, these forms of blaming each other for irrationality reveal the genuine normative significance of rational requirements. By regarding each other as rational or as irrational, we evaluate each other's character, thereby determining how we ought to relate to one another in our attitudes and actions.

I have mentioned that one might doubt whether the relationship modifications that I appealed to throughout the last two chapters amount to genuine blame. According to the views that doubt this, blame is essentially a *moral* notion, and we are never morally blameworthy merely for *being* in an irrational state of mind. We can at best be morally blameworthy for not managing our mental lives better. I will now argue against these views by arguing that irrationality can warrant moral blame. This will reinforce my case against Indirect Voluntarism. Even if one denies that there are non-moral kinds of blame, there is still a good argument to be made for the view that rational failure can give rise to genuine moral blame. This also reveals a hitherto unnoticed normative force of the requirements of rationality.

I will begin by recapping the main dialectic within the current discussion about responsibility for attitudes, before moving on to the main argument.

7.1 The Dialectic and the Strategy

We show reactive sentiments in response to our own and other people's attitudes. We sometimes *feel guilty* about not intending what we believe we ought to do. We might *resent* someone for wishing us harm. And we can *feel hurt* by what others think and feel about us. According to Peter Strawson (1962), such reactive sentiments reveal that we hold each other responsible. Since we show them in response to attitudes, it seems that we don't merely regard each other as responsible for actions and omissions, but also for attitudes.

However, philosophers have argued that we don't control our attitudes as we control our actions: we cannot choose what we believe, feel, desire, or intend.[1] Indirect Voluntarists therefore argue that responsibility for attitudes can at most be indirect. We can control beliefs by inquiry, emotions by going for a walk, and desires and intentions by deliberating about what is good and right. For the Indirect Voluntarist, the observation that we show reactive sentiments toward attitudes merely reveals that we regard each other as blameworthy for not properly managing our mental life. It allows us to conclude only that we can be indirectly responsible for

attitudes in virtue of being directly responsible for actions and omissions that caused the attitude.[2]

There has recently been opposition to voluntarist accounts, coming from epistemology and moral psychology. The common theme is that voluntarists misconceive of attitudes as if they were nothing but brute states that can be managed indirectly – like headaches that we can manage by taking a painkiller. Yet our attitudes, like our actions, are often direct responses to reasons, and we are evaluated as rational, or as irrational, in light of them. It therefore seems that responsibility for attitudes is as direct as responsibility for actions. *Rationalist* accounts, which explain responsibility for attitudes by appealing to their reasons-responsiveness, develop this idea.[3] However, voluntarists insist that *genuine* responsibility presupposes voluntary control, and that anything rationalists talk about is therefore not direct responsibility for an attitude: it is either merely rational evaluation of an attitude, and thus doesn't even amount to *responsibility*, or it is derived from responsibility for prior attitudinal self-management, and thus doesn't amount to *direct* responsibility for an attitude.[4]

This chapter proposes a way out of this stalemate by considering our practice of apology and forgiveness. I argue that this practice is sometimes fully intelligible when a person holds a 'non-culpable' attitude (NCA) – that is, an attitude that the person had no duty to avoid by prior actions or omissions. This reveals that we sometimes regard each other as morally blameworthy for NCAs. Voluntarists cannot reduce this blameworthiness to mere negative evaluation or to indirect blameworthiness: the latter strategy fails because the person had no duty to avoid the attitude; and the former strategy fails because our practice of apology and forgiveness makes sense only under the presumption of genuine moral blameworthiness. It follows that voluntarism is a false theory of how we regard each other as responsible. Voluntarists therefore shouldn't understand themselves as proposing an analysis of our practices of holding responsible, but a revision of these practices. This places the argumentative burden on voluntarists, and it changes the nature of the debate.

I first characterize NCAs and frame my discussion (Chapter 7.2). I then present my argument against Indirect Voluntarism (Chapter 7.3). Next, I argue that a rationalist account can make sense of blameworthiness for NCAs (Chapter 7.4). Finally, I conclude while relating the discussion back to the debate about the normativity of rationality (Chapter 7.5).

7.2 Non-culpable Attitudes and Reactive Sentiments

I here introduce NCAs and discuss what blaming responses can be appropriate toward NCAs. I frame my discussion by arguing that although non-culpability should affect the intensity of reactive sentiments, it remains

open whether NCAs can impair relationships in such a way as to warrant blaming responses. The next subchapter then argues that NCAs are relationship impairing in the relevant way. I will work with the following definition:

> NCA: An attitude A of a person S is non-culpable *iff* S did not possess decisive practical (that is, prudential or moral) reasons to engage in practices of attitudinal self-management (that is, actions and omissions with foreseeable effects on attitudes) that probably would have led S to not holding A.

There are also non-culpable *absences* of attitudes. For instance, S's not believing that p is non-culpable when S lacks a true belief about p because S didn't inquire into p, but had sufficient reasons not to inquire into p, because S had more important things to do. Furthermore, *aspects* of an attitude can be non-culpable, even if the attitude is culpable. For instance, S might culpably fail to manage their anger, but then experience an uncontrollable increase in the anger's intensity due to further provocations. For simplicity, I focus on cases where people *hold* an attitude that is *fully* non-culpable, and that still has problematic consequences for themselves or others. In these cases, the attitude could probably not have been avoided by engaging in reasonable practices of attitudinal self-management. That is, there was no course of action S *ought* to have performed that would probably have led S to avoid the attitude. That makes the attitude non-culpable.

I employ the notion of 'ought', 'reason', and 'duty' (also 'justification', 'permission', and 'allowed') in their subjective or perspectivist sense. For I am interested in *possessed* reasons: violating objective 'oughts', as when you fail to leave a burning house because you have no clue about the fire, doesn't give rise to blameworthiness. By contrast, it is at least closer to a sufficient condition on blameworthiness that the reasons that explain why it was true that you *ought* to have done something that you failed to do were, in a sense, possessed by you when you violated the duty. This is the use of 'ought' that is central to discussions about responsibility and blame. Using the term in this perspectivist sense doesn't commit me to any view about which use of 'ought' is the 'central deliberative ought'.[5]

Note that most of your attitudes are non-culpable. It is seldom true that you had decisive reasons to avoid an attitude by engaging in self-management practices. Such practices are normally just not worth the effort. Most obviously, you had no duty to avoid most rational attitudes and attitudes with good consequences. Furthermore, beliefs that you formed reflectively by means of careful inquiry into an issue, but also beliefs that you just acquired spontaneously by moving around in the world, are

mostly non-culpable: you had decisive reasons to acquire reflective beliefs, and spontaneous beliefs were mostly unforeseeable because you normally don't know what you will encounter.[6] By contrast, suppose that you decide to call your ex, who recently broke up with you because you want to know whether they have a new romantic partner. It is reasonable for you to avoid the relevant belief and to avoid feeling the associated emotions (like sadness and anger) by not talking to your ex. In such cases, acquiring the attitudes is culpable, and you will be (at least prudentially) blameworthy. Similar considerations hold for emotions, desires, and intentions: while many are reflectively and responsibly acquired or come about unforeseeably and spontaneously in our everyday life and were thus acquired non-culpably, other attitudes should have been avoided by various strategies of self-management and are thus culpable.

Among NCAs are deeply ingrained implicit biases and problematic emotional patterns that we display due to our education and socialization. For instance, if an adolescent is amused by a sexist joke, this amusement might stem from an implicit bias, and the emotion and the laughing might well have come about so spontaneously that both were non-culpable, while the implicit bias was so ingrained that the young man had not yet had opportunities to reasonably get rid of it by exercising control over his own mind. The argument I develop here will allow us to say that he could be blameworthy for his bias and his amusement (given certain conditions) even though he had no duty to get rid of his bias up to now.

More generally, we can harm each other with behavioral responses that stem from our attitudinal dispositions that are deeply ingrained, and which are thus often non-culpable. Getting clear about which (if any) blaming-responses to adopt when we harm each other in these non-culpable ways will help us to see how we should relate to one another.

To see what it could mean to be blameworthy for a NCA, consider a case from Angela Smith (2005: 267–268). While Abigail adopted a racist ideology through growing up in a racist environment, Bert grew up in a tolerant family but then later reflectively endorsed racist attitudes. According to Smith,

> understanding of the circumstances in which a person's evaluative tendencies were formed may […] have a very important influence on the kind or degree of moral criticism we think it appropriate to make. We can appreciate how difficult it might be for Abigail to come to recognize the viciousness of her own evaluative judgments, given their early entrenchment in her psyche, and also how difficult it might be for her to modify these judgments once their viciousness is recognized. For this reason, we are likely to be less critical of Abigail than we are of Bert, who adopted his racist-intolerant commitments in a fully reflective way

> (after being exposed to the morally appropriate values of tolerance and inclusiveness).
>
> (268)

According to Smith's rationalist account, Abigail is blameworthy for holding racist attitudes because they reflect vicious evaluative judgments. Voluntarists baulk at the idea that a person who lacked reasonable opportunities to get rid of an attitude can be blameworthy for it. Even though Abigail's attitudes are morally bad, voluntarists regard it as unfair or inappropriate to blame her for attitudes that she acquired merely by being raised in a certain way. Smith's case can elicit the intuition that we cannot be responsible, and thus that we cannot be blameworthy, for NCAs. Note that to 'appreciate how difficult it might be for Abigail' to recognize that her attitudes are morally problematic is to acknowledge that Abigail, due to her distorted perspective, had no duty to engage in attitudinal self-management that would have led her to non-racist attitudes. Her racism is, up to a point in her life, non-culpable. Voluntarists will argue that although Abigail might *become* blameworthy for these attitudes when she gains access to information that casts doubt on her racist outlook, which will provide her with reasons that will then create an obligation for her to actively reconsider her view (by inquiry and deliberation), she is blameless right now.

Smith's case suggests that the degree to which a person had opportunities to engage in reasonable attitudinal self-management, that is, the degree of culpability of an attitude, is proportional to the degree to which it is appropriate to blame them. Even Smith acknowledges that we respond to Abigail and Bert *differently*. But if the emotional intensity of, say, our indignation diminishes with less reasonable opportunities to change, then it's natural to suppose that our indignation won't be appropriate when there were *no* reasonable opportunities for the person to change. Intuitively, it therefore seems that we are blameless for NCAs.

However, this becomes less convincing if we see that blame need not involve strong feelings of resentment, indignation, or guilt. As Smith (2013: 32) points out, some reactions deserve the label 'blame' without involving passionate components. In particular, blaming loved ones for moral failures commonly happens without indignation. Furthermore, we can modify a relationship by "dispassionately 'unfriending' someone on one's Facebook page, for example, or by simply refusing to trust anymore, and these too should qualify as blame" (2013: 32). Proponents of such an account of blame argue that blame's primary function is to mark impaired relationships (Scanlon 2008). Regarding one's relationship to someone as impaired because the person failed to show proper regard, *just is* the blame (Hieronymi 2004).

Employing this notion of blame, we could reply to the voluntarist that it is appropriate to treat one's relationship with Abigail as impaired due to her racist attitudes, and that she is therefore blameworthy, even if indignation might not be appropriate. Alternatively, one could agree that indignation appropriately diminishes when we realize that an attitude is less culpable than we thought. However, this observation doesn't imply that indignation becomes fully inappropriate in response to NCAs. Milder forms of resentment or indignation might still be appropriate.[7]

Could voluntarists explain the verdict that Abigail is blameworthy? It doesn't seem so. For if someone's reasons against an action are not decisive – if the action isn't forbidden – then the action is allowed. But it is incoherent to blame someone for doing something permissible (cf. Kiesewetter 2017: 29). Since Abigail lacked decisive reasons to avoid her racist attitudes, she was permitted to let them develop and is thus blameless for not managing her mental life better. Furthermore, even if we accept that one can be blameworthy for doing something that was allowed, note that Abigail was *unaware* that she was becoming racist, and avoiding becoming racist in her social environment is *difficult*. She therefore didn't even have *very strong reasons* to avoid her NCA. I will argue that she can still be blameworthy in such cases. Voluntarists cannot explain this, even if there *were* some blameworthy permissible acts.

Voluntarists must then argue that Abigail isn't blameworthy: the relationship modification that is appropriate due to Abigail's racist attitudes is not a kind of blame. I agree that we can modify relationships without blame, as for instance, when people drift apart when they live in different places. So the fact that we can appropriately modify our relationship toward Abigail does nothing, by itself, to show that she is blameworthy.

In reply, I argue that NCAs sometimes make a specific *kind* of relationship modification appropriate that counts as genuine blame. The literature on blame contains proposals about what turns a relationship modification into blame.[8] I will employ a sufficient condition that is compatible with most proposals: if a relationship modification makes the full practice of apology and forgiveness intelligible, then it's blame. I'll argue that, since a person's NCAs can make this practice intelligible, we can be blameworthy for NCAs. This poses a new problem for voluntarists.

7.3 A New Argument against Indirect Voluntarism

I will now argue that NCAs give rise to our full practice of apology and forgiveness and that they therefore can impair our relationships in the way required for moral blameworthiness. I first introduce the cases that will be counterexamples to voluntarism (Chapter 7.3.1). I then defend my argument in three steps (Chapters 7.3.2–7.3.4): first, it is appropriate for the

protagonists to apologize for a harm they have caused; second, the specific kind of apology makes the full practice of forgiveness intelligible, thereby revealing that we regard the protagonists as blameworthy; third, this blameworthiness is best understood as blameworthiness for their NCA.

7.3.1 Blameworthy Non-culpable Attitudes

This subchapter presents cases in which an NCA manifests in non-culpable behavior that causes harm to others. I will make the following assumptions in each case:

a The person (S) holds an attitude (A) that they acquired through a process in which there was nothing they should have done to avoid holding it, that is, A is non-culpable.
b S displays a behavior (B) that was (partially) caused by A, and there was nothing S should have done to avoid B, that is, B is also non-culpable.
c A is not pathological, but rather responsive to reasons to some degree.
d A is unjustified, that is, it is insufficiently supported by S's possessed reasons for A.

As I will argue, each of the following cases is a counterexample to the voluntarist's view that we are blameless for NCAs if we assume (a)–(d):

1 John (S) was raised with the sexist belief that men are supposed to lead (A). As a result, he often treats women unfairly in job interviews (B).
2 Sonja (S) is under stress this morning. She couldn't reasonably avoid the stressful situation. As a result, she has no time to think clearly about the hair she discovers in the sink. She forms the unjustified belief that they are from her roommate Sarah (A). She feels angry at Sarah, but she manages her behavior toward Sarah to the best of her abilities. Yet Sonja cannot manage to avoid behaving in an unfriendly way to Sarah and leaves the house with a suspicious look at Sarah that displays an element of contempt (B).
3 Tim (S) intends to become a successful entrepreneur in the oil and gas industry (A). He idolizes his father, who leads such a company. In his 20s, Tim becomes head of the company and leads it on with success (B). However, he gains evidence that his company is destroying our planet. Since his intention is constitutive of his social identity, it is very difficult for him to appreciate that evidence, and to revise his intention.
4 Carla and Jack (S) are visiting their daughter. Upon noticing the state of her garden, Carla and Jack desire to give her a lesson about proper

gardening (A). While they can reflectively avoid giving such a lesson, this desire poisons their social interaction (B). Their daughter notices the desire.[9]

5 Ramona (S) is a successful leader in higher management. Due to her stressful job, she has become irascible. On one particularly stressful day, she becomes angry at a subordinate (A) without any offense, and she behaves in a very unfriendly way (B).

The cases cover non-culpable beliefs, intentions, desires, and emotions. Notably, it is impossible to describe cases that involve merely one kind of attitude. For if an attitude is responsive to reasons (condition (c)), it enters into rational relations with other attitudes the person holds. Tim's intention will come with a belief that he ought to become a successful entrepreneur, and Carla's and Jack's desires will lead to certain emotions that influence their behavior. There is room in each case to argue that blameworthiness for one kind of attitude is more basic than blameworthiness for another. For instance, maybe blameworthiness for a particular emotion traces back to blameworthiness for a belief – which is plausible in (2). However, the position I sketch in Chapter 7.4 will imply that we can be directly responsible for any attitude that is sufficiently responsive to reasons.[10] This is compatible with there being hierarchies of responsibility for attitudes.

Direct responsibility and blameworthiness presuppose some degree of reasons-responsiveness. A pathological phobia of spiders doesn't indicate that the phobic is a coward, and therefore doesn't, by itself, make the phobic criticizable in any substantial sense (though they might be criticizable for not going to therapy). The protagonists (S) in (1)–(5) would be perceived differently if we learned that their attitudes aren't responsive to reasons *at all*: they would become more like obstacles to deal with than persons to relate to. But can NCAs be reasons-responsive?

Remember that evaluating an attitude as non-culpable is to look at the practical reasons the person had to manage this attitude, and then to judge that the person responded correctly to their reasons: their actions and omissions were permissible. That is, someone with a NCA complied with their duties of attitudinal self-management, at least with respect to that attitude. By contrast, when evaluating an attitude as reasons-responsive (that is, as non-pathological), we consider whether the attitude could *now* change in response to reasons. If rational discourse can have an influence on John's non-culpable sexist views in an appropriate subset of close possible worlds, then his non-culpably acquired attitudes are responsive to reasons to a sufficient degree to be rationally evaluable. As I will argue, his attitudes can then give rise to our full practice of apology and forgiveness, and thus to genuine blameworthiness.

I turn to condition (d) in Chapter 7.4, which I think follows as a precondition on direct blameworthiness from a broadly rationalist account of responsibility. For now, I elaborate on (a) and (b).

By assuming (a), we conceive of our cases in such a way that S lacked decisive reasons to engage in attitudinal self-management that could have avoided A. For instance, in (1) and (3), we assume that the sexist belief and the capitalistic intention were so deeply ingrained that the young protagonists could not be reasonably expected to get rid of them by long and careful reflection. In cases (2), (4), and (5), we assume that the attitudes that were the immediate cause of unfriendly behavior were either not foreseeable for S (say, Carla and Jack didn't know their daughter had a garden) or were tough to avoid for S (say, Sonja is sensitive about cleanliness, and Ramona needs things to be done her way). Here, we additionally suppose that S didn't yet have reasonable opportunities to get rid of the underlying disposition that caused them to form A. That is, we assume that A is truly non-culpable.

Concerning (b) – the non-culpability of behavior B – we need to assume either that B would be justified if A was justified, or else that B was too difficult to control by S, given A.

To illustrate the first assumption, we might suppose that if John's sexist belief in (1) was justified, then it would also be justified to treat women in the way he does in job interviews. Here the *original* moral fault seems to lie in his holding a sexist belief rather than in his behavior. This also holds for Tim's capitalistic intention in (3): leading his company to success would be justified if his intention was justified. Therefore, his behavior itself is non-culpable: any blameworthiness for it must originate in some attitude or deed before the behavior at issue occurred.

To illustrate the second assumption, we might suppose that in (4) Carla and Jack cannot reasonably avoid being distracted from the conversation by the neglected flowers outside the window, thereby revealing their judgmental desire, or we might conceive of (2) and (5) in such a way that Sonja or, respectively, Ramona cannot reasonably control the involuntary and automatic expressions of her anger, given the other duties she has right now, like getting ready for work, or, respectively, distributing tasks among her subordinates. Generally, the fact that it is difficult for a person to control a specific behavior can make it the case that they lack decisive reasons to avoid this behavior, thereby rendering it non-culpable.

7.3.2 Apologies and Excuses

My argument starts by noting that S owes an apology to the harmed party for B. Assume that Carla and Jack realize that their desire poisoned the social interaction or that Ramona realizes how her irascibility together

with the stressful situation, led her to be unfriendly to her subordinates. They then should apologize. Providing merely an excuse would be inappropriate: explaining to the harmed party why they couldn't do otherwise would not repair the impaired relationship (cf. Hieronymi 2019). Only an apology can do this work. That is, an apology, rather than an excuse, is appropriate although A and B were non-culpable. This is so even if we assume that the harmed party knows about the non-culpability: they know how hard it is for S to change their views on a matter (in (1) and (3)), or to control their behavior (in the other cases).

So far, this is merely an observation about how we expect each other to interact: we want apologies for harms people caused, not just excuses. This holds even in cases where the person is blameless for the harm caused (cf. Sussman 2018: 788). The second, more controversial, step of my argument is that the *specific kind* of apology S owes to the harmed party implies that S is blameworthy. I now turn to this claim.

7.3.3 Apologies, Blameworthiness, and Forgiveness

We often apologize *conventionally* for behavior for which we aren't blameworthy. Conventional apologies range from harmless cases to severe ones. Among the harmless cases are apologizing for accidentally touching a stranger while sitting down on public transport and apologizing for hitting someone with a door after opening it with the normally expected amount of attentiveness and care. Here we are *causally* responsible for harm. But we didn't violate any norms that we were (normatively) expected to comply with. There are severe cases with the same structure. You might hit someone by opening the door with normal care, and yet the person might get hurt so badly that they end up in hospital. An honest apology is appropriate. Providing an excuse wouldn't satisfy our moral expectations: we demand that you apologize, rather than explain yourself, but we do not demand that you acknowledge any blameworthiness. Of course, you might become blameworthy for *failing* to apologize conventionally.

Thus, not all appropriate apologies imply blameworthiness. These conventional apologies differ from what I call, somewhat technically, *authentic* apologies: by apologizing authentically, one acknowledges blameworthiness for violation of a norm that one was rightly expected to comply with. In (1)–(5), S should apologize authentically for the harm they have caused, thereby acknowledging their blameworthiness. For now, I leave the nature of this norm violation aside – I return to it in Chapter 7.4. Instead, I argue that S indeed owes an authentic apology to the harmed party.

How can we determine whether the apology S owes is authentic? The initial problem is that one can apologize authentically in the same manner as one apologizes conventionally: both can be done very seriously, honestly,

etc. It is therefore often impossible to distinguish the apologies by looking at how they are physically conducted (which is why my label 'authentic' is somewhat technical). Instead, we need to look at the wider context of the practice in which authentic apologies are embedded. For only authentic apologies – that is, those that imply blameworthiness – open the space for the full practice of forgiveness.

According to a prominent understanding, to forgive is to let go of resentment while still acknowledging that the offender committed a genuine wrong for which they are blameworthy (cf. Hieronymi 2001). Forgiving is possible, normally, only if the offender realizes the wrong they have committed *as a wrong*, apologizes, and feels remorse about what they have done. The offender changes the reflective stance on their behavior. If the wrong is then indeed no longer a threat to your standing within the moral community, a continuation of resentment wouldn't be appropriate. Since this account explains paradigm cases of forgiveness, it will serve for my purposes.[11]

The claim is that the full practice of forgiveness is *pointless* when the subject is blameless. If all involved parties know that your touching was unintentional, or that you opened the door with normal care, then there is nothing to forgive, because you are not to blame. It then doesn't make sense for the harmed party to say 'I forgive you' and to really mean it. In the harmless cases, this sentence could be at best a joke. And in the more serious cases, it would indicate that the situation was misunderstood. There is something to forgive only if there is someone to blame. Furthermore, it doesn't make sense to *not* forgive you, or to *postpone* forgiveness. If no one was wronged, then there is no point in denying forgiveness because there is nothing to forgive in the first place. By saying that the *full* practice of forgiveness is intelligible, I mean that *each* of these moves – forgiving, denying, or postponing forgiveness – makes sense. Importantly, each move isn't always *justified*: sometimes you ought to forgive, but then denying forgiveness is still intelligible.

I proceed as follows. First, I show that the full practice of forgiveness is intelligible in (1)–(5). Second, I argue that this implies blameworthiness by replying to counterexamples. This will also clarify the distinction between authentic apologies and conventional apologies.

An important feature of my cases is that behavior B reveals a problematic aspect of the person's character or self. In (1) and (5), a person (the job applicant, the subordinate) was directly disrespected by S's participation in sexist or elitist structures. Insofar as S's attitudes aren't just cases of pathology, they reveal a feature of S's self on which the disrespected must take a stance. They could, for pragmatic reasons, view S as a victim of their social environment, thereby avoiding the "strains of involvement" (Strawson 1962). But this would only be a strategic move to psychologically deal with

the wrong. Alternatively, they could hold S accountable for their behavior, deny them forgiveness, and thereby continue to view the relationship as impaired. Doing so would amount to moral protest (cf. Smith 2013), and to emphasizing one's dignity and worth (cf. Hieronymi 2001), and therefore to blame. Similar responses can be appropriate to Tim in (3) if we imagine that he apologizes to the public after realizing how he deceived himself into thinking that business as usual wouldn't contribute to planetary destruction. The public, and especially those most affected by climate change, can intelligibly deny forgiveness.

In (2) and (4), we can imagine that the protagonists decide to reduce their involvement with S even though S apologized. The daughter in (4) might not invite her parents over to her place for a while. By doing so, she could intelligibly hold that her parents aren't just blameless victims of the conservatism of their generation. Instead, she might judge that their parents' desire to lecture her about the garden reflects a character trait that negatively affects the relationship: the parents do not treat their daughter as a peer who is in charge of her own life. Plausibly, the parents display the vice of judgmentalism. Gary Watson (2013) points out that this entails non-acceptance, rejection, and the distance of superiority.[12] Plausibly, Sonja's quick judgment in (2) manifests similar flaws in character. Sonja isn't merely a victim of her stress, as the voluntarist makes the case appear. Rather, her behavior reveals an aspect of her self. She thus needs to apologize, thereby acknowledging her blameworthiness. She then makes herself vulnerable by opening the room for Sarah to decide whether she will forgive her. Sarah's continuation of blame might manifest in her decision to move out (suppose Sonja is often non-culpably unfriendly). This decision could be a response to Sonja's self, rather than a mere strategic way of dealing with her.

In all these cases, the full practice of forgiveness is intelligible. The second step of my argument is that the intelligibility of this practice implies that the person is blameworthy. The recent literature suggests two counterexamples to this second step in my argument. First, one might appeal to cases in which a person is merely *causally* responsible for a harm, but in which making amends still makes sense. Take Williams' (1981) famous case of the lorry driver who non-culpably kills a child. The driver should feel 'agent-regret' and it seems intelligible that he apologizes to the child's parents, even in full knowledge of his blamelessness. This apology doesn't seem to be 'merely conventional', and the practice of forgiveness seems intelligible.

However, such cases don't make the *full* practice of forgiveness intelligible. As David Sussman (2018: 806–807) argues, only 'quasi-apology' and 'quasi-forgiveness' are appropriate. Such quasi-apologies "need not express any sort of change of heart, any resolution to act differently in

the future", and "although the driver should ask for forgiveness […], the parents should respond by telling him that none is needed" (2018: 806). The parents might feel overwhelmed after the accident, making it difficult to respond appropriately to the driver; doubts about whether the driver is blameless might plague the minds of all parties involved; and the driver should take up the parents' blame although he is blameless: it isn't the right moment to insist on excuses (2018: 800–801). However, if the situation was transparent, and if the parents are no longer overwhelmed, then they should (quasi-)forgive the driver: his apology does not make him vulnerable to the parents' judgment because he is blameless. It therefore wouldn't even be *intelligible* to deny him forgiveness. This is a crucial difference to the authentic apology that is appropriate in (1)–(5).[13]

Second, Julia Driver (2017) argues that we sometimes morally ought to violate relationship-based duties. In her example, Agamemnon has a moral duty (we assume) to sacrifice his daughter Iphigenia for the greater good: the survival of his whole army. He ought to violate a *pro tanto* relationship-based obligation toward his daughter. For Driver, this is a resolvable tragic dilemma of *blameless* wronging. Yet, Agamemnon should ask Iphigenia for forgiveness after attempting to sacrifice her. According to Driver, his obligation to his daughter is *not eliminated* by the moral duty to save the lives of his army – it is *merely outweighed*. She thinks that Agamemnon is blameless because he (a) does what he morally ought to do, and (b) is not of bad character, but acts on a reflective decision (or so we assume).

However, Driver's analysis of this case is problematic. Note first that, if Iphigenia agrees that Agamemnon is blameless, then how could she intelligibly *deny* him forgiveness? Their relationship might then be impaired just because it is difficult to get along with someone who attempted to kill you, even though his reasons were decisive. Again, only quasi-forgiveness would be intelligible. Driver would object that Agamemnon's violation of his *pro tanto* relationship-based obligation not to sacrifice Iphigenia makes the *full* practice of forgiveness intelligible. But it is puzzling why his *pro tanto* obligation has any such significance when it is outweighed.

The account I develop here offers a better explanation of why genuine forgiveness could still be intelligible even though Agamemnon does what he ought to: a good father doesn't even *think* about sacrificing his daughter for the greater good. The fact that he attempts to sacrifice her reveals that his attitudes fall short of the ideal of the parent–child relationship. While Agamemnon indeed performs all the actions he ought to perform, he violates norms that govern his attitudes: as a father, he should think and feel in a way that prevents him from sacrificing his daughter. That his attitudes fall short of this ideal makes him blameworthy, and at least his daughter is justified to blame him. That is, he holds blameworthy non-culpable attitudes that impair his relationship.

One might object that if Agamemnon ought to sacrifice Iphigenia, then he must also be permitted to have attitudes that allow him to go through with it. However, note that this would be the wrong kind of reason for attitudes (cf. Hieronymi 2006): the fact that he can only sacrifice his daughter if he doesn't love her is not a reason against loving her. It is at best a reason to get rid of his love. If he gets rid of it, he still fails to live up to the ideals of the parent–child relationship. Tragically, Agamemnon violates either a duty to act or a duty to think and feel. Both duties are all-things-considered duties in the sense that neither is outweighed by the other.

One might generally doubt that parents have an attitudinal obligation to love their children so much that they cannot sacrifice them for the greater good. If so, then Agamemnon is indeed blameless. But then only quasi-forgiveness is intelligible: Iphigenia couldn't intelligibly deny him forgiveness. The intuition that she *can* intelligibly deny him forgiveness, and that she furthermore has a *right* to do so, just indicates that parents *do* have such an attitudinal obligation: parents should love their children to such a degree that sacrificing them is not an option.

Thus, Driver's counterexample ends up in the following dilemma: either Agamemnon wrongs Iphigenia or he doesn't. If he doesn't, then he is indeed blameless; but then only quasi-forgiveness is intelligible. If he does, then the full practice of forgiveness is intelligible; but then he is blameworthy because his wronging Iphigenia consists in violating an *all-things-considered* attitudinal duty – rather than a mere *pro tanto* duty not to sacrifice her. More generally, whenever forgiveness seems to be intelligible even though the person seems to be blameless, it either turns out that only quasi-forgiveness is intelligible, or else an attitudinal relationship-based duty was violated, and therefore the person is in fact blameworthy.

I have argued that, in (1)–(5), S owes an authentic apology to the wronged party – that is, an apology that makes the full practice of forgiveness intelligible and thus implies blameworthiness. I have also defended the claim that the intelligibility of this practice implies blameworthiness: in any potential counterexample, either mere quasi-forgiveness is intelligible, or the person is, after all, blameworthy for violating an attitudinal norm. I thus conclude that S is blameworthy.

7.3.4 From Behavior to Attitudes

I now argue that if S is blameworthy, then S cannot merely be blameworthy for their non-culpable behavior B, but must also be blameworthy for their attitude A that caused B. The best explanation for S's blameworthiness for B is that S is *originally* or *directly* blameworthy for A. Remember that we have stipulated that there was nothing S should have done to avoid A, and that S already managed B as well as we can reasonably expect of S.

I have argued that this does nothing to annul S's blameworthiness: B still reveals an impaired relationship that calls for the full practice of apology and forgiveness. But does that show that S is blameworthy for *A*?

First, a clarifying remark: when S apologizes, they don't normally apologize *for the attitude*. Rather, they apologize for their behavior and the resulting harm.[14] This doesn't imply that S is only blameworthy for B but not for A. If S indeed managed B to the best of their abilities, then S is praiseworthy in one respect: S cares about not letting their mental state go on a rampage in the external world. But if S is praiseworthy for managing B, then S's blameworthiness cannot be explained merely by appealing to B. Furthermore, much of B concerns involuntary expressions which can, by themselves, no more provide a basis for blame than automatic reflexes (esp. in (2), (4), and (5)). Finally, the proposal that S is only blameworthy for B but not for A won't help the voluntarist. For they are committed to the view that S is blameless for anything non-culpable – whether attitude or behavior. So it isn't promising, and especially not for voluntarists, to locate the original or direct blameworthiness in the non-culpable behavior.

We are left with the task of looking for another original locus of S's blameworthiness. There are two reasons why S's attitudes are a plausible candidate. Firstly, attitudes are, like actions, governed by norms and reasons. If blameworthiness presupposes the violation of a norm or a failure to respond correctly to possessed reasons, then attitudes are, like voluntary actions, promising for locating direct blameworthiness. Secondly, vices manifest not only in actions and overt behavior, but also in attitudes. The irascible person tends to form unfavorable beliefs about others on the basis of which they become disproportionally angry. Such vices impair relationships, and we can therefore be blameworthy for displaying them. If attitudes can manifest vices, then our attitudes are plausible loci of original blameworthiness.

I have no conclusive argument that there is no other good explanation for S's blameworthiness. However, I have argued that S cannot be blameworthy only for B. At the very least, voluntarists cannot retreat to this claim. Attitudes, which are part of the space of reasons and which can be manifestations of vicious character, seem the most plausible original locus of S's blameworthiness. I conclude that we can be blameworthy for NCAs, and that therefore voluntarism is false.

7.4 A Rationalist Alternative

When considering blameworthiness for NCAs, we experience an ambiguity. On the one hand, the person holding a NCA had no duty to avoid it, and thus, they seem to be blameless. On the other hand, the NCA might still be responsive to reasons, and it might cause harm to others. As I have

argued, this can give rise to our full practice of apology and forgiveness, which in turn implies that we regard each other as blameworthy for NCAs. This subchapter proposes that we can capture this ambiguity by adopting a rationalist account that acknowledges that both reasons for actions as well as reasons for attitudes shape our responses toward people who hold unjustified NCAs. Given my previous argument, this allows us to see that answerability, or, more precisely, failures of reasons-responsiveness, can appropriately give rise to moral blame. Voluntarists therefore cannot reply that rational evaluation doesn't give rise to 'genuine' responsibility and blame (see Chapter 7.1).

Return to Smith's case (Chapter 7.2). Abigail's racist attitudes are, earlier in her life, non-culpable, and later, when she has more access to relevant information that casts doubt on her racist outlook, still less culpable than those held by Bert, who adopted them reflectively after growing up in a tolerant family. As granted earlier to the voluntarist, the non-culpability, and thus the lack of voluntary control Abigail had over her mind, affects the emotional intensity of our blaming responses. I suggested that this is because the difficulty of exercising control over her mind (due to lack of information, but also because her attitudes are recalcitrant) makes it the case that she lacks decisive reasons to engage in attempts at attitudinal self-management.

We can see now how a rationalist account can make sense of the intuitive ambiguity we experience when it comes to NCAs. First, a rationalist can argue that our reactive sentiments are rightly affected by Abigail's lack of opportunities to exercise voluntary control: they are affected by whether she responded correctly to her practical reasons for attitudinal self-management. Since we assume that Abigail's attitudes are non-culpable, we assume that she didn't commit any such fault in self-management when she responded to her practical reasons with her actions and omissions. We are less critical of her than of Bert because, insofar as it comes to how Abigail conducted her actions and omissions in shaping her mental life, she is faultless.

Furthermore, a rationalist will argue that Abigail is still *answerable* for her racist attitudes: it is intelligible to request her reasons for them, rather than just her reasons for self-management. When asking 'why do you believe that p?' or 'why are you angry about p?' we often do not merely ask for explanation, but for justification: we ask for the reasons the person took to support believing p, or the reasons the person took to support being angry about p. A good reason for belief is evidence for p, and a good reason for being angry is an offense. If the person is unable to provide a good answer to our request for justification, this can give rise to blame. Therefore, if we suppose that Abigail's reasons do not support a racist attitude, a rationalist can locate a genuine shortcoming in Abigail's

reasons-responsiveness: she fails to acknowledge reasons against holding racist attitudes. This failure, however, isn't a failure in exercising voluntary control.[15]

The rationalist assumes that, if Abigail is indeed blameworthy, then her reasons favor not holding racist attitudes: her attitudes are unjustified. This seems plausible. For suppose that Abigail's perspective was so distorted by her community that all her possessed reasons *favor* racist beliefs. Abigail *was* in this situation until a certain age. It was only when she gained more information in light of which we would rightly expect her to revise her beliefs that she becomes blameworthy. For up to this point, Abigail might have responded correctly to all her reasons and yet, due to her unfortunate circumstances, ended up with racist attitudes. However, if a person responds correctly to *all* their reasons for attitudes and actions, then they are blameless.

Remember that the voluntarist made a similar claim: that Abigail becomes blameworthy once she has relevant information accessible that would then make it the case that she ought to engage with this information actively and to consider it until her beliefs are revised in light of it (see Chapter 7.2). The difference to the rationalist's proposal is that only the voluntarist requires that Abigail has reasonable opportunities to revise her attitude. To see this difference, suppose that, after Abigail reads information that should rationally dislodge her racist stance, she irrationally fails to revise her beliefs. She would have to invest more time in engaging with the information. However, suppose that she has no such duty to engage further with the information – say, because the library is on fire and she should leave. The voluntarist would claim that she is blameless, for she fulfilled all her duties to manage her attitudes. The rationalist, by contrast, can argue that she is blameworthy because her racist attitudes are now unjustified in light of her new information.[16]

My argument against Indirect Voluntarism from Chapter 7.3 concluded that Abigail's NCAs might still warrant moral blame. When she gains awareness of her failure to acknowledge her reasons against her racist attitudes, she owes the people she harmed an authentic apology, thereby making herself vulnerable to their decisions about whether they forgive her. Again, this practice wouldn't make any sense while knowing that Abigail's attitudes were a result of her responding correctly to *all* her reasons – for both her actions and her attitudes. However, if the NCAs themselves are unjustified by her possessed reasons, then by denying forgiveness and continuing resentment, victims of her racism might appropriately blame her, thereby expressing moral protest and emphasizing their own dignity. Abigail's racist attitudes can then reflect various vices, thereby revealing features of her character or self, thereby warranting relationship modifications that count as moral blame.[17]

If this is correct, then Indirect Voluntarists can no longer insist that rationally evaluating an attitude by noting that it is insufficiently supported by reasons cannot give rise to 'genuine' blame. Rather, my argument reveals that failing to respond correctly to one's reasons for attitudes can give rise to reactive responses that are very similar to the responses that are appropriate when people fail to respond correctly to their reasons for actions. For these reactive responses to be appropriate, it is often sufficient that an attitude is non-pathological, is unjustified by the subject's reasons, and causes harm to another person. It isn't necessary that one violated any duty of attitudinal self-management. The voluntarist therefore cannot account for our blameworthiness for NCAs.

In reply, the voluntarist might propose that our practice of apology and forgiveness needs a fundamental revision (see Nussbaum 2016). Nothing I have said excludes this possibility. However, if I am right, then our actual practice commits us to the view that we are sometimes blameworthy for NCAs, and thus to the falsity of Indirect Voluntarism. The argumentative burden is on the voluntarist to argue that we should revise our actual practice.

7.5 Conclusion: Moral Blameworthiness for Irrationality

The present investigation is not concerned with potential reasons to revise our overall practice of holding each other responsible. Rather, I am concerned with understanding our current practice and spelling out our implicit commitments that we share by being involved in this practice. If the argument of this chapter is sound, then responsibility for rationality is not a *sui generis* form of responsibility that differs fundamentally from our moral responsibility. Rather, failures of reasons-responsiveness – that is, of rationality – can give rise to moral blame.

Importantly, the failures at issue need not involve commitment to *moral* reasons for or against attitudes. For instance, in case (2) where Ramona forms an irrational belief about the hair in the sink, and as a result of this cannot reasonably manage her behavior toward her roommate, the belief is irrational in an *epistemic* sense. The epistemically irrational belief is the source of Ramona's *moral* blameworthiness: the fact that it is irrational explains why the practice of apology and forgiveness is intelligible. It follows that mere epistemic irrationality can give rise to moral blame.[18] This shouldn't surprise us. We often become angry or feel indignant about people holding epistemically irrational beliefs – namely, when they hold *morally harmful* beliefs, like climate change denial. We do not then stop and ask whether these people had reasonable opportunities to avoid being epistemically irrational. Even if they didn't have such opportunities to manage their beliefs better, certain blaming responses can still be appropriate. For sure, if we

realize that they lacked such opportunities, then this should temper our blaming attitudes, and sometimes even affect the kind of blame that is appropriate. But that doesn't imply that the blame at issue is not a form of *moral* blame.

What we could conclude from this is that kinds of blame must be individuated by the kind of harm they cause (moral or epistemic harm), and not merely by the kind of mistake in which they are grounded.[19] In Chapters 4 and 5, I understood epistemic blame as the form of blame that is appropriate in response to mere epistemic norm violations. No wonder then that some forms of epistemic blame also amount to moral blame: epistemic failure can cause moral harm. Such harms warrant different reactions than more harmless cases of epistemic norm violations. We might therefore want to resist calling moral blaming reactions involving anger or indignation 'epistemic'. However, this is merely a terminological issue. For these moral blaming reactions are appropriate in response to violations of epistemic norms. It is up to us to determine how best to carve up the conceptual landscape of kinds of blame.

Importantly, epistemic norms are the norms of epistemic rationality. So what my argument shows is that being epistemically irrational, as well as displaying other forms of attitudinal irrationality, can give rise to moral blame, rather than to 'mere rational criticizability'. Indeed, it shows that the dichotomy between 'mere rational criticism' and 'genuine blame' is a false one. For rational failures seem to have a greater normative significance than even the most recent proponents of the normativity of rationality assume. Being irrational is not just a pure cognitive flaw that can give rise to soft forms of 'personal criticizability', where it remains open whether this deserves the label 'blame' (see Kiesewetter 2017: chapter 2). Instead, being irrational can be a severe character flaw that can even deserve moral blame.

I will stay largely neutral with regard to the forms moral blame for irrationality can take – ranging from weaker to stronger forms of passionate emotions, maybe even resentment and indignation, but at least forms of rebuke and anger, as well as feelings of disappointment, and including also moral forms of relationship modification (rather than just epistemic ones). No matter what stance we take here, we should allow that this blame is as *moral* as blame could be. For otherwise, apology and forgiveness would not be intelligible.

Thus, even if we were to deny that there is a distinctively epistemic kind of responsibility and blame, or corresponding other distinctive forms of responsibility and blame for (ir)rationality, we should allow that we are directly responsible for our attitudes because we are sometimes *directly morally blameworthy* for holding them. This is implied by our actual practice. Of course, one could argue that this practice needs a fundamental

revision. Yet again, the burden of argument here rests with the opponent of mental morality.[20]

Note, however, that my view isn't affected by the common objection to mental morality that moral norms don't govern our attitudes. To defend moral blameworthiness for attitudes, it's sufficient to argue that they're governed by the norms of rationality, and that these norms can downstream cause moral harm. I am skeptical about whether wrong-kind or practical reasons (including moral ones) are genuine reasons *for attitudes*, rather than just reasons for causing or bringing about attitudes.[21] So, in a sense, my view provides an *error-theory* about mental morality: that we're morally blameworthy for our attitudes can be explained by the norms of rationality; it need not commit us to moral norms for attitudes.

Notes

1 See Hieronymi (2006, 2009a). See Kavka (1983) and Owens (2000: 81–82) on why we cannot choose intentions. I put this issue aside. If we can choose some attitudes, then some reactive sentiments are appropriate for this reason.
2 See Meylan (2013, 2017) and Peels (2017) as well as Price (1954) on doxastic responsibility, Oakley (1992) on emotional responsibility, and Jacobs (2001) on responsibility for character. See also Rosen's (2004) view that blameworthiness always originates in akratic action and Fischer and Tognazzini's (2009) view on tracing back all responsibility to voluntary action.
3 Cf. Hieronymi (2006, 2008, 2014), McCormick (2015), McHugh (2013a, 2013b; 2017), Osborne (2021), Owens (2000, 2017a: intro.), Portmore (2019), Roberts (2015), Smith (2005), and White (2019). Adams (1985) is the *locus classicus* of the opposition to Indirect Voluntarism. He did not appeal to the idea of reasons-responsiveness at the time.
4 For both responses, see Peels (2017: 46–48, 159–160). For the latter response, see Fischer and Tognazzini (2009).
5 Cf. Kiesewetter (2017: chapter 8) for a perspectivist account.
6 Some aspects of spontaneous beliefs are reasonably foreseeable: it is foreseeable *that you will acquire beliefs about the environment* that you are planning to walk through, but you don't know which *concrete* beliefs you will acquire.
7 Wallace argues that "it would indeed be strange to suppose that one might blame another person without feeling an attitude of indignation or resentment toward the person, or that one might blame oneself without feeling guilt" (1994: 75). The second reply to the voluntarist in this paragraph does justice to this point. That is, I am not committed to the view that we can blame without feeling reactive sentiments, though I sympathize with this idea due to the arguments brought forward by Smith, Scanlon, and others. Hieronymi (2014, 2019) claims that there is a reactive attitude, 'resentment+', that is only appropriate towards someone for holding an attitude M if this person had a reasonable opportunity to avoid M, but she denies that ordinary resentment presupposes reasonable opportunity.
8 For instance, that one's attitudes fall short of the normative ideal of the relationship (Scanlon 2008), or that it is an expression of moral protest (Smith 2013).

9 This case is motivated by a case in Tognazzini's (2020) discussion of blameworthiness for judgmental thoughts.
10 Plausibly, one might argue that in (4), there would be nothing problematic if the parents merely *believed* that the garden isn't properly tended – that is, without desiring to lecture. This indicates that the desire is the problem, rather than the belief. Furthermore, Ramona's anger in (5) need not be based on any full-fledged belief.
11 Fricker's (2019) gifted forgiveness isn't captured by the account.
12 Cf. Tognazzini (2020) for a more detailed discussion of Watson's take on judgmentalism.
13 One might take issue with calling the driver's apology 'conventional'. I sympathize with this. Calling all non-authentic apologies 'conventional' is technical. There are important differences *within* the camp of non-authentic apologies.
14 However, maybe we sometimes *do* apologize for 'believing badly' about someone (cf. Basu 2019), and plausibly we apologize for having been angry, where we need not be precise about the object of the apology.
15 Especially Hieronymi (2006, 2014: 22–24) and Roberts (2015) have worked out these two dimensions of holding each other responsible for our attitudes by distinguishing between reasons for attitudinal self-management and genuine reasons for the attitudes themselves, and by connecting them to different exercises of agency (see Hieronymi 2009b). See Chapters 6.3, 6.4 and 6.6 for more on the two agential capacities that are relevant here.
16 I say 'has the resources' because the rationalist might formulate further necessary conditions on blameworthiness.
17 Am I denying that blaming racists who, due to their distorted perspective, responded correctly to all their reasons is legitimate? Not necessarily, for the sense of 'appropriateness' here is narrower than 'all things considered' legitimacy (Calhoun 1989; Coates 2020). Furthermore, one can protest systemic oppression without blaming individuals.
18 See also Ayars (2021). I am here not ruling out that moral blameworthiness for epistemic failure requires moral encroachment on epistemic rationality (see Dandelet 2023 for a recent discussion). However, if so, then the moral considerations raise the evidential threshold for epistemic rationality or count themselves as epistemic reasons against belief (see Schroeder 2021 for the latter view). Importantly, they need not be *moral* reasons bearing on a distinctive *moral* normativity of belief, as pragmatists about reasons for belief would claim (see Chapter 1.3 for discussion).
19 Thanks to Cameron Boult for pointing out this implication of my view to me.
20 See Sher (2021) for such an opponent's account, to which I cannot do full justice here.
21 I didn't defend my skepticism here. But see Schmidt (2022) for relevant arguments and considerations. See also Chapter 3.2 of this book for a summary of the dialectic about the debate on practical reasons for belief.

8 Conclusion

This chapter first summarizes the results of the present investigation. I begin with my results on the problem of mental responsibility (Chapter 8.1) and then turn to the results on the normativity of rationality (Chapter 8.2). I then present research avenues for further exploration of the ethics of mind (Chapter 8.3).

8.1 Solving the Problem of Mental Responsibility

Responsibility for attitudes seemed initially puzzling because attitudes are not actions but states. How can we be responsible for just *being* in a state? Intuitively, we are at most responsible for causing a state or for failing to avoid it. However, we also saw that attitudes are responsive to reasons. In this respect, they are quite unlike brute states, such as headaches. For sure, we can also be responsible for having a headache – say, if we forgot to bring our painkillers and now cannot take one to get rid of our headache. However, our responsibility for reasons-responsive states, such as beliefs, desires, emotions, and intentions, seemed to be more *direct* than that. For we often respond to our right-kind reasons for attitudes *directly*, that is, by forming or maintaining the attitude on the basis of these reasons without performing any actions of bringing the attitude about or of maintaining it. Furthermore, others can demand a justification for why we hold an attitude, thereby asking for right-kind reasons for this attitude, and we seem to be criticizable as irrational if it turns out that our attitudes are not properly based on right-kind reasons, or if the right-kind reasons we base our attitudes on are insufficient to justify the attitude. It thus seems that we hold each other directly accountable merely for *being* in certain reasons-responsive states.

One reaction is to deny that we are ever directly responsible for our attitudes. This is the claim of Indirect Voluntarism, which attempts to reduce all responsibility for attitudes to responsibility for prior actions and omissions. Indirect Voluntarists are committed to the view that rational criticizability

DOI: 10.4324/9781003382973-11

does not amount to genuine blame: we are never blameworthy for mere rational failure. We saw that recent epistemological discussions about the normativity of epistemic reasons provide food for this idea. When we hold trivial attitudes that are insufficiently based on right-kind reasons, or when our right-kind reasons conflict with our wrong-kind reasons, then it seems that the right-kind reasons do not provide genuine normative support. Instead, it seems that they derive their normativity entirely from wrong-kind reasons. This brought out a neglected challenge for the normativity of rationality. For we saw that the challenge could be generalized from epistemic reasons to all other kinds of right-kind reasons. Therefore, it can seem that we have good arguments to explain away the intuition that we are ever directly blameworthy merely for holding irrational attitudes, and Indirect Voluntarism can again seem like a plausible view.

However, we also saw that there is a plausible reply to this challenge for the normativity of rationality. First, we considered the possibility of a distinctively epistemic kind of blame – that is, a blame for mere failures of epistemic rationality that does not amount to genuine moral blame. We saw that the recent epistemological literature presents us with some convincing accounts according to which epistemic blame need not amount to strong passionate emotions, like resentment or indignation, but can rather come in the form of, for instance, modifying one's epistemic relationship by reducing one's epistemic trust. Such relationship modifications too can amount to blame. For they are appropriate in response to epistemic vices, and thus do not reduce to merely treating someone as an unreliable truth-tracker. If someone's beliefs are responsive to epistemic reasons to a sufficient degree, then it is appropriate to engage with and relate to the other person as a responsible epistemic agent, rather than as someone who is exempted from epistemic responsibility.

Our practice of holding each other epistemically responsible, and specifically our practice of blaming each other for mere epistemic failures, both in trivial cases and in cases of epistemic-practical conflicts, reveals that epistemic reasons have genuine normative significance: we are answerable for how we respond to our epistemic reasons. Furthermore, we saw that this can be generalized to other forms of rationality: we are also responsible for the way we respond to our right-kind reasons for desires, intentions, and emotions. Failing to respond correctly to these reasons manifests a vice, a defect in character, for which we are blameworthy. Irrationality, that is, not being attuned to features of our environment that make certain attitudes fitting, gives others, as well as ourselves, reasons to relate to us differently than if we were rational within our mental architecture. The normativity of rationality and of right-kind reasons is revealed in our practice of evaluating each other for our character.

This is the first part of my response to Indirect Voluntarism. The second part is an argument to the effect that rational failure can make genuine moral blaming responses appropriate. This argument is meant to address the understandable worry that the distinctively epistemic kind of blame and other forms of non-passionate blame that I appealed to do not count as genuine blaming responses. Anyone who presents this worry has two tasks.

First, they must engage with recent views of epistemic blame and argue that the object of these views does not deserve the label 'blame'. They must even argue that it does not deserve the label 'personal criticism' either. For according to the anti-normativist or nihilist about epistemic reasons, epistemic criticism could at best be directed at a subpersonal system for malfunctioning, but not at the person, who isn't directly responsible for such malfunctioning. If we are criticizable for rational failures at a personal level, then this would already establish the normativity of epistemic and other right-kind reasons. Second, they must object to my argument from apology and forgiveness, according to which our practice of apology and forgiveness is sometimes fully intelligible when mere rational failure causes moral harm. For, as I have argued, this practice presupposes direct moral blameworthiness for irrationality.

I thus conclude that we are directly responsible for having rational or irrational attitudes, and that, since Indirect Voluntarism is committed to a denial of any direct responsibility for attitudes, Indirect Voluntarism fails. What solves our initial philosophical puzzle is close attention to our actual practices of holding each other responsible. As it turns out, these practices imply genuine responsibility for manifesting our character, our virtues and vices, in our attitudes. We are responsible not only for what we do but also for the way we are (Adams 1985). More precisely, we're responsible for the way our character manifests in our reasons-responsive actions and attitudes. Our responsibility for being in mental states is the proper foundation for an *evaluative* kind of normativity that gives other people reasons for relating to us in ways that involve holding us directly responsible for our attitudes. This is how responsibility for rationality becomes intelligible.

8.2 Defending the Normativity of Rationality

Next, let me return to my argument for the normativity of rationality that I presented in Chapter 3 and summarize the new motivation for the normativity of rationality provided by our overall discussion. Central to this argument is the restriction of uses of 'rational' and 'irrational' to cases that are candidates for being directly responsible for rationality or

irrationality. This excludes cases of incoherence that are severely pathological, like severe delusions or phobias. I thus restricted myself to those cases of (ir)rationality in which the attitudes in virtue of which we are called 'rational' or 'irrational' are still responsive to reasons to some degree. There were two reasons for this restriction. First, the intuition that we can be criticizable for irrationality – an intuition that motivates much of the debate about the normativity of rationality – only gets its grip on us if we focus on reasons-responsive attitudes. Second, a *requirement* to be rational can only be intelligible when our attitudes are responsive to reasons, or else we are exempted. Severely pathological attitudes aren't subject to genuine requirements: we cannot coherently expect the pathologically irrational person to be rational. Someone who suffers severe pathology must be dealt with rather than blamed or criticized for their pathological irrationality.[1]

Employing this use of '(ir)rational', I have argued that we fail to respond correctly to right-kind reasons whenever we are irrational: irrationality is a serious failure that merits personal criticism or blame absent excuse. The presented argument works with a plausible connection between the capacity that grounds our direct answerability for attitudes—reasons-responsiveness—and blameworthiness for violating rational requirements. Roughly, your blameworthiness for violating rational requirements is grounded in your failures to respond correctly to right-kind reasons. Since any case in which you seem to be *blameless* for irrationality is either a case where you have an *excuse* or turns out to be a case of *harmless incoherence* rather than irrationality, being irrational guarantees that you fail to respond correctly to right-kind reasons. This is the *argument from responsibility*.

Finally, my defense of the normativity of right-kind reasons for attitudes amounts to an extensive reply to the worry that rationality as reasons-responsiveness might not be normative because right-kind reasons lack normativity. Against this, I have argued throughout Chapters 5–7 that right-kind reasons are genuinely normative reasons because various kinds of blaming responses are appropriate if we fail to respond correctly to these reasons without excuse. These blaming responses reveal the normativity we attach to right-kind reasons, and thus they reveal the significance that these considerations have for our lives. In particular, being irrational can have implications for our interpersonal relationships: our failures to respond correctly to epistemic reasons provide others with reasons to, say, reduce their epistemic trust, or change the way they treat us as fellow inquirers. Other failures of rationality can have similar implications for what attitudes it is appropriate to have toward us, and thus for how we should relate to one another. Such failures can even give rise to moral blame. This reveals the genuine normative force of rational requirements.

8.3 Research in the Ethics of Mind

This concludes our investigation into responsibility for attitudes and the rationality of attitudes. We saw that we are sometimes directly responsible for being rational or irrational and that this responsibility can sometimes even give rise to appropriate moral blame. The present investigation thus provides a foundation for an ethics of mind – including ethics of belief – that is concerned with (evaluative) normative requirements to *be* in certain mental states (and to *believe* certain things) rather than merely with prudential and moral requirements to manage one's mental states (including requirements to manage one's beliefs). Additionally, the investigation provides new motivation for the normativity of rationality in the responsibility-implying sense of rationality (putting aside the dispute between subjectivists and objectivists about reasons). It also provides a *pluralist diagnosis* of the nature of the disagreement between coherentism and reasons-responsivism about the nature of rationality: various philosophical disagreements can motivate this dispute, and we would do well to keep them apart (see Chapter 3.7).

However, many questions remain unanswered. They will motivate further inquiry into the ethics of mind. I wish to close this book by giving some idea of what some of these inquiries could be, without any claim to providing a complete list in this rich and open-ended field. Rather, my aim is to give an idea of how the present inquiry could help us discussing these further topics. I start with the more theoretical issues and then turn to the more applied discussions.

8.3.1 Non-evidential Reasons in Epistemology

This book has attempted to stay neutral about the exact content of epistemic norms. Although Chapter 5 has argued that epistemic norms do not require any backup by pragmatic reasons to be genuinely significant and that we should retain a distinctively epistemic kind of normativity that cannot be weighed or compared with a distinct practical normativity of belief, there is plenty of room for pragmatic or non-evidential reasons to become relevant in epistemology. First, pragmatic reasons are relevant for whether and how to perform epistemic actions, like inquiry or deliberation, that is, the norms of the 'zetetic' (Friedman 2020). Furthermore, they might bear indirectly on the epistemic rationality of belief by being relevant to the rationality of inquiry if the zetetic norms influence the epistemic rationality of belief. They also might bear directly on the epistemic rationality of *not believing* something (Schroeder 2021), or the rationality of *suspending judgment* (Lord 2020), thereby determining the evidential threshold for epistemically rational belief. Finally, even if we wish to resist

categorizing pragmatic influences on belief as genuine normative reasons for belief, we might still want to accept a version of pragmatic or moral encroachment that grants that these considerations can raise the evidential threshold for epistemic rationality without being themselves reasons for or against belief (or even for suspension of judgment). Which of these views about epistemic normativity is more plausible has been left open by the present inquiry, and further research is required to clarify these issues. However, pragmatic reasons against belief or for suspension will meet the worry that these reasons cannot properly motivate beliefs or function as a basis for deliberating about what to believe (Shah 2006; Hieronymi 2006). This issue should continue to be taken seriously (see also Schmidt 2022).

8.3.2 Practical Reasons for Belief

Next, there could be distinctively *practical* reasons for belief – that is, pragmatic reasons for belief that do not bear on epistemic rationality but that still determine what we ought *practically* to believe (Feldman 2000). If this view was right, then practical reasons could not just directly motivate actions of bringing about or maintaining beliefs, but they could also directly motivate beliefs. Some pragmatists argue that indirect motivation is sufficient for genuine practical reasons for belief (Rinard 2019a). I have argued against this view elsewhere (Schmidt 2022). What I have not discussed here is whether we can believe directly for practical reasons, say, by believing at will, or by exercising other forms of direct control over belief, such as 'guidance control' (McCormick 2015). However, any view that defends the possibility of a practical rationality of belief must explain the significance of this rationality over and above the practical rationality of belief management. Why is it not sufficient for our theoretical purposes to grant that practical reasons are relevant to how we *deal* with our beliefs? Why do we need a practical reason to apply *directly* to belief? This is a new problem in addition to the question of motivation or guidance mentioned before.[2]

8.3.3 Implications of the Present View on These Issues

Thus, the question of how pragmatic considerations are relevant both to epistemic rationality and (potentially) practical rationality of belief is still an open debate on which I remained largely neutral here. Nevertheless, my argument has some implications for this debate. First, practical rationality cannot be all there is to belief or belief-management. Rather, there is a distinctively epistemic kind of rationality – one that can sometimes have a similar normative significance as moral duties in cases where epistemic irrationality causes moral harm. To explain this conclusion, we

might need some version of pragmatic or moral encroachment on epistemic rationality – an issue I did not discuss. Or we might not: after all, even if moral considerations affect the evidential threshold for epistemically rational belief, we're still in the sphere of *epistemic* evaluation. So it is difficult to see how moral encroachment on *epistemic* rationality could explain the apparent *moral* or *interpersonal* significance of epistemic norm violations, at least by itself. For that task, we might rather need a *metaepistemological* account of the significance of epistemic normativity for our lives (see Chapter 8.3.4). In any case, we should acknowledge that beliefs are governed by a distinctive and interpersonally significant kind of epistemic normativity – one that in turn might not be restricted just to belief, but also govern other states, like suspension of judgment, or activities, like inquiry. There is plenty of epistemological theorizing to be done on these issues.

Next, I have provided motivation for the idea that these debates *generalize* to other attitudes. If I am right in this, then there is a distinctive rationality of desires, intentions, the different kinds of emotions, as well as other mental states. In all these cases, rationality has genuine normative significance. Here we can ask questions like: Are all right-kind reasons object-given, or are there state-given reasons of the right-kind – i.e., state-given reasons that bear on the distinctive rationality of the relevant attitude? Can wrong-kind reasons for an attitude bear on the *practical* rationality of an attitude, or only on the rationality of bringing about this attitude? Is there pragmatic encroachment on the rationality of desires, emotions, or intentions? These questions have not been explored in as much detail as the corresponding questions on epistemic rationality, although the parallel epistemological discussions should help us to get a better grip on them. A further task would also be to see whether there are important differences between the rationality of belief and other attitudes. I have focused on the analogies here rather than on potential differences. So, there is still much work to do for extending normative epistemological theory to the ethics of mind.

8.3.4 The External Authority of Rationality

These are questions *internal* to the different kinds of attitudinal rationality – questions about which reasons determine what attitudes we should have and how they do so. Next to these questions, there are also *external* questions about the overall *justification* or the *authority* of a specific kind of rationality. In epistemology, there is a long debate about what justifies our adherence to epistemic norms. Traditionally, epistemologists have appealed to our desire for truth (Kornblith 1993; Papineau 2013), or to the value of epistemic goods, like true belief (McHugh 2013b), knowledge (Williamson 2000), or understanding (Kvanvig 2003). However, sometimes we do not

desire truth – as when we do not wish to know the latest celebrity gossip; and sometimes it is advisable to avoid epistemic goods – as when ignoring uncomfortable truths gives us peace of mind. Therefore, traditional accounts have difficulty explaining why epistemic norms always apply to us. This worry is avoided by constitutivist accounts, which claim that we can only be believers or knowers if we accept truth as the categorical aim of our epistemic endeavors (cf. Wedgwood 2002; Shah and Velleman 2005; Littlejohn 2012; Whiting 2012; Nolfi 2015). However, these accounts invite us to ask *why we should care* about being believers and knowers in the first place.

Due to these problems, many epistemologists have raised doubts about the authority of epistemic norms, some of which have been addressed in the present inquiry. A certain kind of instrumentalists and radical pragmatists argue that epistemic norms are best understood on the model of instrumental practical norms, like game rules, that become normatively relevant only when we have reason to 'play the game of belief' (see Chapter 4, and my arguments against these views in Chapters 5–7). By contrast, others have recently attempted to preserve a distinctively epistemic dimension of normativity by grounding it in our epistemic sociality (Goldberg 2019; Chrisman 2020, 2022; Dyke 2021; Wei 2022; Boult 2024b; Fleisher forthcoming; Hannon and Woodard forthcoming) or in the practical value of being subject to epistemic norms (Owens 2017b), or in human well-being (Reisner forthcoming).

The present inquiry was in line with these latter approaches that aim at *understanding* epistemic normativity, instead of explaining it away by reducing it to instrumental practical normativity. I have argued that epistemic normativity is revealed in our practice of epistemic blame. I have suggested in Chapter 5.5 that the authority of epistemic rationality might ultimately rest in the practical justification of our practice of epistemic blame: it is *useful* – in the broadest sense of the term – to blame each other for distinctively epistemic failures and hold one epistemically accountable in various ways. If this is on the right track, then we should stick with our practices of epistemic accountability and maintain epistemically healthy relationships instead of attempting to revise our practice. Again, the present inquiry did not discuss the issue of revising our practices. Rather, it was concerned with working out our implicit commitments within the epistemic practice and the practice of rational evaluation more generally. However, if the question of the authority of epistemic norms ultimately depends somehow on the utility of our distinctively epistemic practice, then pragmatism could be right at least about this *external* question.

Again, this epistemological debate must be generalized to the external question about the authority of all rational requirements (see Schmidt forthcoming b). Even if I was right that our actual practices of holding each

other responsible for (ir)rationality commit us to the normativity of rationality, our actual practices are not self-justifying. Is it all-things-considered better to have a practice in which we hold each other responsible for rational failure than not to have such a practice? This question might deserve different answers, depending on whether we talk about, say, the practice of blaming each other for emotions and the practice of blaming each other for beliefs, or intentions. As I see it, there is at least *final value* in this practice insofar as it helps us to relate to each other properly as responsible agents (see also Boult 2024b). This value of our epistemic relationships is, I think, a value that we should care about, even in societies that pervert this value by punishing rationality, or by making it costly to be rational.[3] However, there might be ways of revising our practices that help us to relate better to each other. If my view is correct, then the external question about the ultimate grounds for the authority of rational requirements seems to come down to a question about the justification of our responsibility practices. This connects the debate on rationality with discussions about the value of blame (e.g., Nussbaum 2016). Again, the ethics of mind here opens interesting new avenues for philosophical research.

8.3.5 The Ethics of Mind and Relationship-Based Duties

In Chapter 7.3, I have suggested that some duties we have in virtue of standing in relationships with people close to us come down to requirements to have certain attitudes – like a requirement to love your child or to be sad at a friend's misfortune. Here, the ethics of mind might open an interesting new perspective on relationship-based duties. First, we might solve some puzzles about such duties. For instance, how can it be that we wrong someone by doing what we *morally* ought to do? This could be explained by conflicts between rational requirements and moral duties: we might be rationally required to love a person close to us even though we morally ought to, say, sacrifice them (as in Driver's case of Agamemnon and Iphigenia discussed in Chapter 6.3). Second, we might explore whether this idea can be extended: maybe all rational requirements derive from our relationships in a broad sense of the term 'relationship'? For instance, epistemic requirements might derive from the value of *epistemic relations*, which are constituted by the attitudes in virtue of which we hold each other epistemically responsible (Boult 2024b). It seems worthwhile to explore this idea and to connect it with recent insights from social epistemology (see Chapter 8.3.4). And it might extend to other kinds of rationality as well – maybe all rationality derives its authority from the value of relationships, so that all rational requirements would turn out to be relational duties. I explore the prospects of such a view in a more recent project.

8.3.6 Responsibility to Ourselves

Another issue that I have merely touched upon is how we hold *ourselves* responsible for rational failure. We sometimes feel guilty for holding certain attitudes. Does that mean that we stand in some relationship with ourselves? If so, do we have special duties toward ourselves that involve certain attitudes? These questions point to more general debates on how we are related to ourselves, and how this relationship mirrors the relationship we have toward others. This connects the ethics of mind with debates about self-knowledge and debates in the philosophy of mind or philosophy of action about our relationship to ourselves, like self-control. It also connects the ethics of mind with debates about *duties to ourselves*: is there some sense in which we owe it to ourselves to be rational, in a similar way as we might owe it to ourselves to look out for our own well-being, at least in certain cases?

8.3.7 Applied Epistemology: Vices, Oppression, and Epistemic Pollution

Many questions arise in applied epistemology and applied ethics for which the ethics of mind is relevant. While I cannot comment on these issues here in too much detail, I wish to at least point out some fruitful directions for connecting these issues to the present inquiry.

First, the present inquiry has provided some foundation for holding us responsible for our vices, thus giving us some idea of how to solve the 'responsibility problem' for vice epistemology (Battaly 2019): manifesting vices implies being irrational, and rational failure is a basis for various kinds of blaming responses, even if our irrational attitudes were not under our direct or indirect voluntary control. However, the present account of responsibility for rationality also implies that fully rational attitudes are blameless and thus do not manifest genuine vices.

There is thus a limit to responsibility and blame for oppressive ideologies in cases where people in oppressive systems – whether oppressors or oppressed – are rational in holding these attitudes. As a result, many cases of indoctrination (as discussed by Tobi 2022) might turn out to involve largely rational beliefs and other rational attitudes (though they're objectively false or unfitting) and so imply the blamelessness of some perpetrators of injustice. This might hold especially for members of marginalized groups, who often have strong practical reasons to take up the ideology of the oppressors in order to thrive in a world dominated by them, thus ending up with many rational but false beliefs that are spread in an epistemically hostile environment (Meylan and Schmidt 2023). This should encourage us to look more on the *systemic* dynamics of oppression, in

particular when the perpetrators were indoctrinated by a hegemonic system and are themselves members of marginalized groups.

Finally, new technology offers plenty of ways of influencing our minds – both our own minds as well as the minds of others. On a daily basis, our attitudes are influenced by algorithms that are epistemically opaque – that is, where the workings of the algorithms could not even be in principle transparent to experts. Furthermore, companies, political parties, fake experts and their organizations, as well as other interest groups manipulate our minds, by making certain information unavailable to us, discrediting the reliability of expert testimony, mimicking expertise, and more generally creating a polluted epistemic environment (Levy 2022). Such developments might ultimately undermine our doxastic autonomy by creating conditions of epistemic domination or epistemic anarchy (Chrisman 2022: 139–143, ms).

Here the ethics of mind finds its most urgent application in the modern world: by helping us to see through epistemic pollution and enabling healthy epistemic relationships, which in turn is a precondition for tackling global problems, like climate change (Schmidt forthcoming a). My suggestion is *not* that we can solve these issues simply by trying to make each other more rational. Quite the opposite. If we take seriously the possibility that most agents are *already* sufficiently rational but simply *misled* in epistemically hostile environments (Nguyen 2023), then the proposed account implies that our epistemic relationships toward agents with false and morally repugnant beliefs might after all *not* be impaired. This should give us some hope that rational discourse is still possible despite political polarization, and shift the discussions from a focus on individual believers to the trustworthiness of groups and institutions.

Questions for the ethics of mind here include whom we should trust and why and how one can permissibly manage the minds of rational agents and groups without infringing their autonomy. Distinguishing between direct answerability – what attitudes we should hold toward whom and why – and indirect accountability – how we should manage our own minds and the minds of others – is a helpful starting point for framing these issues properly. However, this application is beyond the scope of the present book. Thus, while the present inquiry has not addressed these applied questions, it has provided the foundations for a systematic treatment of applied questions in the ethics of mind.

Maybe most importantly, this book has emphasized the relevance of reasons and rationality for *interpersonal* questions regarding our mind: since we are directly responsible for our attitudes, and since failing to be rational in our attitudes can cause severe moral harm and warrant moral blaming responses, we may hold agents morally accountable who base their attitudes on insufficient reasons. Irrationality is not just some private

mistake in our own mind that doesn't concern others. Instead, rationality has its own distinctive significance for our lives with each other by fostering valuable kinds of interpersonal relationships.

Notes

1 Importantly, a person suffering severe pathology might be responsible for managing their mental life (see Chapter 6 for discussion of the two faces of mental responsibility). However, the kind of indirect responsibility that is grounded in this ability to exercise indirect voluntary control over one's mind is not the answerability that grounds our responsibility for being rational or irrational. This is direct responsibility.
2 See Chapter 3.2 on the dialectical situation as I see it. See also my argument that there is no point in weighing epistemic and practical reasons in Chapter 5.5, as well as Schmidt (forthcoming c).
3 Social epistemic instrumentalist accounts of epistemic normativity imply that epistemic norms lose their authority when compliance with them is overall harmful to a community (see Scott 2023 for this worry). Appealing to a special kind of value of certain kinds of epistemic relationships might allow us to avoid this worry.

References

Adams, Robert M. (1985). Involuntary Sins, *Philosophical Review* 94, 3–31.

Adler, Jonathan (2002). *Belief's Own Ethics*, Cambridge: MIT Press.

Alston, William P. (1988). The Deontological Conception of Epistemic Justification, *Philosophical Perspectives* 2, 257–299.

Alvarez, Maria (2010). *Kinds of Reasons. An Essay in the Philosophy of Action*, New York: Oxford University Press.

Alvarez, Maria (2017). Reasons for Action: Justification, Motivation, Explanation, in *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), ed. by Edward N. Zalta, <https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl>.

Alvarez, Maria (2018). Reasons for Action, Acting for Reasons, and Rationality, *Synthese* 195, 3293–3310.

Anscombe, Gertude Elizabeth Margaret (1957). *Intention*, Cambridge, MA: Harvard University Press.

Arpaly, Nomy (2023). Practical Reasons to Believe, Epistemic Reasons to Act, and the Baffled Action Theorist, *Philosophical Issues* 33, 22–32.

Arpaly, Nomy, and Timothy Schroeder (2014). *In Praise of Desire*, New York: Oxford University Press.

Ayars, Alisabeth A. (2021). Blaming for Unreasonableness: Accountability without Ill Will, *Journal of Ethics and Social Philosophy* 19, 56–79.

Ballarini, Cristina (2022). Epistemic Blame and the New Evil Demon Problem, *Philosophical Studies* 179, 2475–2505.

Basu, Rima (2019). The Wrongs of Racist Beliefs, *Philosophical Studies* 176, 2497–2515.

Battaly, Heather (2019). Vice Epistemology Has a Responsibility Problem, *Philosophical Issues* 129, 24–36.

Berker, Selim (2018). A Combinatorial Argument against Practical Reasons for Belief, *Analytic Philosophy* 59, 427–470.

Bondy, Patrick (2018). *Epistemic Rationality and Epistemic Normativity*, New York: Routledge.

Bonjour, Laurence (1999). The Dialectic of Foundationalism and Coherentism, in *The Blackwell Guide to Epistemology*, ed. by John Greco and Ernest Sosa, Malden, MA: Blackwell, 117–144.

Boult, Cameron (2017). Epistemic Normativity and the Justification-Excuse Distinction, *Synthese* 194, 4065–4081.
Boult, Cameron (2021). There Is a Distinctively Epistemic Kind of Blame, *Philosophy and Phenomenological Research* 103(3), 518–534.
Boult, Cameron (2023). The Significance of Epistemic Blame, *Erkenntnis* 88, 807–828.
Boult, Cameron (2024). Degrees of Epistemic Criticizability, *The Philosophical Quarterly* 74(2), 431–452.
Boult, Cameron (2024a). Epistemic Blame as Relationship Modification: Reply to Smartt, *Philosophical Studies* 181, 387–396.
Boult, Cameorn (2024b). Epistemic Relationships and the Source of Epistemic Normativity, *Philosophical Issues* 34.
Boult, Cameron (2024c). *The Puzzle of Epistemic Blame. On the Nature and Norms of Epistemic Relationships*, New York: Oxford University Press.
Boyle, Matthew (2011). 'Making up Your Mind' and the Activity of Reason, *Philosopher's Imprint* 11, 1–24.
Broome, John (2005). Have We Reason to Do as Rationality Requires? – A Comment on Raz, *Journal of Ethics and Social Philosophy* 1, 1–8.
Broome, John (2007). Does Rationality Consist in Responding Correctly to Reasons? *Journal of Moral Philosophy* 4, 349–374.
Broome, John (2013). *Rationality through Reasoning*, New York: Oxford University Press.
Broome, John (2016). A Linguistic Turn in the Philosophy of Normativity? *Analytic Philosophy* 57, 1–14.
Broome, John (2020). Rationality versus Normativity, *Australasian Philosophical Review* 4, 293–311.
Brössel, Peter, Eder, A. Anna-Maria, and Franz Huber (2013). Evidential Support and Instrumental Rationality, *Philosophy and Phenomenological Research* 87, 279–300.
Brown, Jessica (2020). What Is Epistemic Blame? *Noûs* 54, 389–407.
Calhoun, Cheshire (1989). Responsibility and Reproach, *Ethics* 99, 389–406.
Callahan, Laura F. (2023). Why Are Epistemic Reasons Normative? *Episteme* 1–16. Online first, doi: 10.1017/epi.2023.13.
Chrisman, Matthew (2008). Ought to Believe, *Journal of Philosophy* 105, 346–370.
Chrisman, Matthew (2020). Believing as We Ought and the Democratic Route to Knowledge, in *The Ethics of Belief and Beyond. Understanding Mental Normativity*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge, 47–70.
Chrisman, Matthew (2022). *Belief, Agency, and Knowledge*, New York: Oxford University Press.
Chrisman, Matthew (ms). Freedom of Belief, Political Rather than Metaphysical, unpublished manuscript.
Christensen, David (2021). The Ineliminability of Epistemic Rationality, *Philosophy and Phenomenological Research* 103, 501–517.
Clarke, Randolph (2023). The Source of Responsibility, *Ethics* 133, 163–188.
Clifford, William K. (1877). The Ethics of Belief, in *The Ethics of Belief. And Other Essays*, New York: Prometheus Books, 1999, 70–96.

Coates, D. Justin (2020). The Ethics of Blame: A Primer, in *The Ethics of Belief and Beyond. Understanding Mental Normativity*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge, 192–214.

Coates, D. Justin, and Neal Tognazzini (2013). *Blame. Its Nature and Norms*, New York: Oxford University Press.

Côté-Bouchard, Charles (2017). Is Epistemic Normativity Value-Based? *Dialogue* 56, 407–430.

Côté-Bouchard, Charles, and Clayton Littlejohn (2018). Knowledge, Reasons, and Errors about Error Theory, in *Metaepistemology*, ed. by C. Kyriacou and R. McKenna, Cham: Palgrave Macmillan, 147–171.

Cowie, Christopher (2014). In Defence of Instrumentalism about Epistemic Normativity, *Synthese* 191, 4003–4017.

Cowie, Christopher (2019). *Morality and Epistemic Judgment*, New York: Oxford University Press.

Crawford, Lindsay (2020). Relationships and Reasons for Belief, in *The Ethics of Belief and Beyond. Understanding Mental Normativity*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge, 87–108.

Dandelet, Sophie (2023). Doxastic Wronging and Evidentialism, *Australasian Journal of Philosophy* 101, 82–95.

Daoust, Marc-Kevin (forthcoming). Coherence, First-Personal Deliberation, and Crossword Puzzles, *Philosophical Topics*.

Davidson, Donald (1963). Actions, Reasons, and Causes, *Journal of Philosophy* 60, 685–700.

Dennett, Daniel, and Gregg D. Caruso (2021). *Just Deserts. Debating Free Will*, Cambridge: Polity Press.

Descartes, René (1649). *The Passions of the Soul*, in *The Philosophical Writings of Descartes: Volume I*, transl. by Cottingham et al., Cambridge: Cambridge University Press, 1985, 325–404.

D'Oro, Guiseppina, and Constantine Sandis (eds.) (2013). *Reasons and Causes: Causalism and Non-causalism in the Philosophy of Action*, London: Palgrave-Macmillan.

Dogramaci, Sinan (2012). Reverse Engineering Epistemic Evaluations, *Philosophy and Phenomenological Research* 84, 413–530.

Dotson, Kristie (2014). Conceptualizing Epistemic Oppression, *Social Epistemology* 28, 115–138.

Driver, Julia (2017). Wronging, Blame, and Forgiveness, *Oxford Studies in Agency and Responsibility 4*, 206–219.

Dyke, Michelle (2021). Could Our Epistemic Reasons Be Collective Practical Reasons? *Noûs* 55, 842–862.

Ernst, Gerhard (2008). *Die Objektivität der Moral*, Paderborn: Mentis.

Ernst, Gerhard (2020). Two Kinds of Rationality, in *The Ethics of Belief and Beyond. Understanding Mental Normativity*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge, 176–190.

Feldman, Richard (2000). The Ethics of Belief, *Philosophy and Phenomenological Research* 60, 667–695.

Feldman, Richard (2007). Reasonable Religious Disagreements, in *Philosophers without God: Meditations on Atheism and the Secular Life*, ed. by Louise Antony, New York: Oxford University Press, 194–214.

Field, Claire (forthcoming). Embracing Incoherence, in *Epistemic Dilemmas*, ed. by Nick Hughes, New York: Oxford University Press.
Fischer, John, and Mark Ravizza (1998). *Responsibility and Control. A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.
Fischer, John, and Neal Tognazzini (2009). The Truth about Tracing, *Noûs* 43, 531–556.
Fleisher, Will (forthcoming). Epistemic Practices: A Unified Account of Epistemic and Zetetic Normativity, *Noûs*, online first, doi: 10.1111/nous.12514.
Flores, Carolina, and Elise Woodard (2023). Epistemic Norms on Evidence-Gathering, *Philosophical Studies* 180: 2547–2571.
Frankfurt, Harry G. (1969). Alternate Possibilities and Moral Responsibility, *Journal of Philosophy* 66, 829–839.
Fricker, Miranda (2019). Forgiveness – An Ordered Pluralism, *Australasian Philosophical Review* 3, 241–260.
Friedman, Jane (2020). The Epistemic and the Zetetic, *Philosophical Review* 129, 501–536.
Gaultier, Benoit (2020). Responsibility for Doxastic Strength Grounds Responsibility for Belief, in *The Ethics of Belief and Beyond. Understanding Mental Normativity*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge, 71–85.
Gerken, Mikkel (2011). Warrant and Action, *Synthese* 178, 529–547.
Gertken, Jan, and Benjamin Kiesewetter (2017). The Right and the Wrong Kind of Reasons, *Philosophy Compass* 12, e12412, 1–14.
Ginet, Carl (2001). Deciding to Believe, in *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. by Matthias Steup, New York: Oxford University Press, 63–75.
Goldberg, Sanford C. (2019). *To the Best of Our Knowledge. Social Expectations and Epistemic Normativity*, New York: Oxford University Press.
Graham, Peter A. (2014). A Sketch of a Theory of Moral Blameworthiness, *Philosophy and Phenomenological Research* 88, 388–409.
Greco, John (2019). Justifications and Excuses in Epistemology, *Noûs* 55, 517–537.
Grimm, Stephen (2009). Epistemic Normativity, in *Epistemic Value*, ed. by Adrian Haddock, Alan Millar, and Duncan Pritchard, New York: Oxford University Press, 243–264.
Hannon, Michael, and Elise Woodard (Forthcoming). The Construction of Epistemic Normativity, *Philosophical Issues* 35.
Harman, Gilbert (1986). *Change in View: Principles of Reasoning*, Cambridge: MIT Press.
Hazzlet, Allan (2013). *A Luxury of the Understanding: On the Value of True Belief*, New York: Oxford University Press.
Heil, John (1992). Believing Reasonably, *Noûs* 26, 47–62.
Hieronymi, Pamela (2001). Articulating an Uncompromising Forgiveness, *Philosophy and Phenomenological Research* 62, 529–555.
Hieronymi, Pamela (2004). The Force and Fairness of Blame, *Philosophical Perspectives* 18, 115–148.
Hieronymi, Pamela (2006). Controlling Attitudes, *Pacific Philosophical Quarterly* 87, 45–74.

Hieronymi, Pamela (2008). Responsibility for Believing, *Synthese* 161, 357–373.
Hieronymi, Pamela (2009a). Believing at Will, *Canadian Journal of Philosophy, Supplementary Volume* 35, 149–187.
Hieronymi, Pamela (2009b). Two Kinds of Agency, in *Mental Actions*, ed. by L. O'Brian and M. Soteriou, New York: Oxford University Press, 138–162.
Hieronymi, Pamela (2014). Reflection and Responsibility, *Philosophy and Public Affairs* 42, 3–41.
Hieronymi, Pamela (2019). I'll Bet You Think This Blame Is about You, in *Oxford Studies in Agency and Responsibility V: Themes from the Philosophy of Gary Watson*, ed. by D. Justin Coates and Neal Tognazzini, New York: Oxford University Press, 60–87.
Hieronymi, Pamela (ms). *Minds That Matter*, unpublished manuscript (spring 2019 version).
Hirvelä, Jaakko (2023). The Structure of Moral Encroachment, *Philosophical Studies* 180, 1793–1812.
Howard, Christopher (2020). Weighing Epistemic and Practical Reasons for Belief, *Philosophical Studies* 177, 2227–2243.
Hubacher Haerle, Pablo (2023). Is OCD Epistemically Irrational? *Philosophy, Psychiatry, and Psychology* 30, 133–146.
Hyman, John (2015). *Action, Knowledge, and Will*, New York: Oxford University Press.
Jacobs, Jonathan (2001). *Choosing Character*, Ithaca, NY: Cornell University Press.
James, William (1896). The Will to Believe, in his *The Will to Believe. And Other Essays in Popular Philosophy, and Human Immortality*, Dover, DE: Dover Publishing, 1956, 1–31.
Kauppinen, Antti (2016). Character and Blame in Hume and Beyond, in *Perspectives on Character*, ed. by Iskra Fileva, New York: Oxford University Press, 46–62.
Kauppinen, Antti (2018). Epistemic Norms and Epistemic Accountability, *Philosopher's Imprint* 18, 1–16.
Kauppinen, Antti (2021). Rationality as the Rule of Reason, *Noûs* 55, 539–559.
Kauppinen, Antti (2023). The Epistemic versus the Practical, *Oxford Studies in Metaethics* 18, 137–162.
Kavka, Gregory S. (1983). The Toxin Puzzle, *Analysis* 43, 33–36.
Kelly, Thomas (2002). The Rationality of Belief and Some Other Propositional Attitudes, *Philosophical Studies* 110, 163–196.
Kelly, Thomas (2003). Epistemic Rationality as Instrumental Rationality: A Critique, *Philosophy and Phenomenological Research* 66, 612–640.
Kelly, Thomas, and Stewart Cohen (2024). There Are No Practical Reasons for Belief, In *Contemporary Debate in Epistemology* (3rd edition), ed. by M. Steup et al., Hoboken, NJ: Wiley-Blackwell, 81–92.
Kiesewetter, Benjamin (2017). *The Normativity of Rationality*, Oxford: Oxford University Press.
Kiesewetter, Benjamin (2020). Rationality as Reasons-Responsiveness, *Australasian Philosophical Review* 4, 332–342.

Kiesewetter, Benjamin (2021). Are Epistemic Reasons Normative? *Noûs* 56, 670–695.
Kiesewetter, Benjamin (2024). Epistemic Normativity without Epistemic Teleology, *Philosophical Issues* 34.
Kolodny, Niko (2005). Why Be Rational? *Mind* 114, 509–563.
Korcz, Keith Allen (2019). The Epistemic Basing Relation, in *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), ed. by Edward N. Zalta, <https://plato.stanford.edu/archives/fall2019/entries/basing-epistemic>.
Kornblith, Hilary (1993). Epistemic Normativity, *Synthese* 94, 357–376.
Korsgaard, Christine (1996). *The Sources of Normativity*, Cambridge: Cambridge University Press.
Kvanvig, Jonathan (2003). *The Value of Knowledge and the Pursuit of Understanding*, Cambridge: Cambridge University Press.
Lasonen-Aarnio, Maria (forthcoming). *Feasible Dispositions: A Normative Framework*, New York: Oxford University Press.
Leary, Stephanie (2017). In Defense of Practical Reasons for Belief, *Australasian Journal of Philosophy* 95, 529–542.
Levy, Neil (2007). Doxastic Responsibility, *Synthese* 155, 127–155.
Levy, Neil (2022). *Bad Beliefs. Why They Happen to Good People*, New York: Oxford University Press.
Lindner, Martina (2020). Implications of the Debate on Doxastic Voluntarism for Clifford's Ethics of Belief, in *The Ethics of Belief and Beyond*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge, 23–46.
Littlejohn, Clayton (2012). *Justification and the Truth-Connection*, Cambridge: Cambridge University Press.
Littlejohn, Clayton (forthcoming). A Plea for Epistemic Excuses, in *The New Evil Demon Problem*, ed. by Julien Dorsch and Fabian Dutant, New York: Oxford University Press.
Lord, Errol (2018). *The Importance of Being Rational*, New York: Oxford University Press.
Lord, Errol (2020). Suspension of Judgment, Rationality's Competition, and the Reach of the Epistemic, in *The Ethics of Belief and Beyond*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge, 126–145.
Luvisotto, Giulia (2022). Two Faces of Responsibility for Beliefs, *Canadian Journal of Philosophy* 52, 761–776.
Madison, Brent J. C. (2018). On Justifications and Excuses, *Synthese* 195, 4551–4562.
Magalotti, Tricia (Forthcoming). Emotions and the Phenomenal Grasping of Epistemic Blameworthiness, *Philosophical Issues* 34.
Maguire, Barry, and Jack Woods (2020). The Game of Belief, *Philosophical Review* 129, 211–249.
Mantel, Susanne (2019). Do Epistemic Reasons Bear on the Ought Simpliciter? *Philosophical Issues* 29, 214–227.
Marušić, Berislav (2015). *Evidence and Agency: Norms of Belief for Promising and Resolving*, New York: Oxford University Press.
McCormick, Miriam S. (2015). *Believing against the Evidence. Agency and the Ethics of Belief*, London: Routledge.

McCormick, Miriam S. (2017). Rational Hope, *Philosophical Explorations* 20, 127–141.
McCormick, Miriam S. (2018). Responding to Skepticism about Doxastic Agency, *Erkenntnis* 83, 627–645.
McCormick, Miriam S. (2020). Believing Badly: Doxastic Duties Are Not Epistemic Duties, in *Epistemic Duties: New Arguments, New Angles*, ed. by Kevin McKain and Scott Stapleford, New York: Routledge, 29–43.
McDowell, John H. (1994). *Mind and World*, Cambridge, MA: Harvard University Press.
McHugh, Conor (2013a). Epistemic Responsibility and Doxastic Agency, *Philosophical Issues* 23, 132–156.
McHugh, Conor (2013b). Normativism and Doxastic Deliberation, *Analytic Philosophy* 54, 447–465.
McHugh, Conor (2014). Exercising Doxastic Freedom, *Philosophy and Phenomenological Research* 88, 1–37.
McHugh, Conor (2017). Attitudinal Control, *Synthese* 194, 2745–2762.
Meiland, Jack W. (1980). What Ought We to Believe? Or the Ethics of Belief Revisited, *American Philosophical Quarterly* 17, 15–24.
Menges, Leonard (2017). The Emotion Account of Blame, *Philosophical Studies* 174, 257–273.
Meylan, Anne (2013). *Foundations of an Ethics of Belief*, Frankfurt: Ontos.
Meylan, Anne (2017). The Consequential Conception of Doxastic Responsibility, *Theoria* 83, 4–28.
Meylan, Anne (2018). The Reasons-Responsiveness Account of Doxastic Responsibility and the Basing Relation, *Erkenntnis* 84, 877–893.
Meylan, Anne (2021). Doxastic Divergence and the Problem of Comparability. Pragmatism Defended Further, *Philosophy and Phenomenological Research* 103, 199–216.
Meylan, Anne, and Sebastian Schmidt (2023). Refusing the COVID-19 Vaccine: What's Wrong with This? *Philosophical Psychology* 36, 1102–1124. doi: 10.1080/09515089.2023.2181151.
Mitova, Veli (2017). *Believable Evidence*, Cambridge: Cambridge University Press.
Mitova, Veli (ed.) (2019). *The Factive Turn in Epistemology*, Cambridge: Cambridge University Press.
Montmarquet, James (1993). *Epistemic Virtue and Doxastic Responsibility*, Lanham, MD: Rowman and Littlefield.
Moran, Richard (2001). *Authority and Estrangement. An Essay on Self-Knowledge*, Princeton, NJ: Princeton University Press.
Nguyen, C. Thi (2023). Hostile Epistemology, *Social Philosophy Today* 39, 9–32.
Nolfi, Kate (2015). How to Be a Normativist about the Nature of Belief, *Pacific Philosophical Quarterly* 96, 181–204.
Nussbaum, Martha (2016). *Anger and Forgiveness*, New York: Oxford University Press.
Oakley, Justin (1992). *Morality and the Emotions*, London: Routledge.
Olson, Jonas (2011). Error Theory and Reasons for Belief, in *Reasons for Belief*, ed. by A. Reisner and A. Steglich-Petersen, Cambridge: Cambridge University Press, 75–93.

Olson, Jonas (2014). *Moral Error Theory*, New York: Oxford University Press.
Olson, Jonas (2018). Moral and Epistemic Error Theory: The Parity Premise Reconsidered, in *Metaepistemology*, ed. by Conor McHugh, Jonathan Way, and David Whiting, New York: Oxford University Press, 107–121.
Osborne, Robert C. (2021). A Social Solution to the Puzzle of Doxastic Responsibility, *Synthese* 198, 9335–9356.
Owens, David J. (2000). *Reason without Freedom. The Problem of Epistemic Normativity*, London: Routledge.
Owens, David J. (2003). Does Belief Have an Aim?, reprinted in his *Normativity and Control*, Oxford: Oxford University Press, 51–68.
Owens, David J. (2017a). *Normativity and Control*, Oxford: Oxford University Press.
Owens, David J. (2017b). Value and Epistemic Normativity, in his *Normativity and Control*, Oxford: Oxford University Press, 87–109.
Paakkunainen, Hille (2018). Doubts about 'Genuinely Normative' Epistemic Reasons, in *Metaepistemology*, ed. by Conor McHugh, Jonathan Way, and David Whiting, New York: Oxford University Press, 122–140.
Papineau, David (2013). There Are No Norms of Belief, in *The Aim of Belief*, ed. by Timothy Chan, New York: Oxford University Press, 64–79.
Parfit, Derek (2001). Rationality and Reasons, in *Exploring Practical Philosophy*, ed. by Dan Egonsson, Jonas Josefsson, and Toni Rønnow-Rasmussen, Burlington, VT: Ashgate, 17–39.
Parfit, Derek (2011). *On What Matters. Volume I*, New York: Oxford University Press.
Pascal, Blaise (1670). *Pensées*, Seattle, WA: Pacific Publishing Studios, 2011.
Peels, Rik (2017). *Responsible Belief*, New York: Oxford University Press.
Pink, Thomas (2009). Reason, Voluntariness and Moral Responsibility, in *Mental Actions*, ed. by L. O'Brian and M. Soteriou, Oxford: Oxford University Press, 95–120.
Piovarchy, Adam (2021). What Do We Want from a Theory of Epistemic Blame? *Australasian Journal of Philosophy* 99, 791–805.
Podgorski, Abelard (2017). Rational Delay, *Philosophers' Imprint* 17, 1–19.
Pojman, Louis (1993). Believing, Willing, and the Ethics of Belief, in: *The Theory of Knowledge*, ed. by Louis Pojman, Belmont, CA: Wadsworth, 525–543.
Portmore, Douglas (2019). Control, Attitudes, and Accountability, in *Oxford Studies in Agency and Responsibility Volume 6*, ed. by David Shoemaker, New York: Oxford University Press.
Price, Henry H. (1954). Belief and Will, *Proceedings of the Aristotelian Society, Supplementary Volumes* 28, 1–26.
Rabinowicz, Wlodek, and Tomi Rønnow-Rasmussen (2004). The Strike of the Demon: On Fitting Pro-Attitudes and Value, *Ethics* 114, 391–423.
Raz, Joseph (2011). *From Normativity to Responsibility*, New York: Oxford University Press.
Reisner, Andrew (2008). Weighing Pragmatic and Evidential Reasons for Belief, *Philosophical Studies* 138, 17–27.
Reisner, Andrew (2009). The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem, *Philosophical Studies* 145, 257–272.

Reisner, Andrew (2018). Pragmatic Reasons for Belief, in *The Oxford Handbook of Reasons and Normativity*, ed. by Daniel Star, New York: Oxford, 705–728.
Reisner, Andrew (forthcoming). *A New Theory of Reasons for Belief: Pragmatic Foundations and Pluralistic Reasons*, New York: Oxford University Press.
Rinard, Susanna (2015). Against the New Evidentialists, *Philosophical Issues* 25, 208–223.
Rinard, Susanna (2017). No Exception for Belief, *Philosophy and Phenomenological Research* 94, 121–143.
Rinard, Susanna (2019a). Equal Treatment for Belief, *Philosophical Studies* 176, 1923–1950.
Rinard, Susanna (2019b). Believing for Practical Reasons, *Noûs* 53, 763–784.
Rinard, Susanna (2022). Eliminating Epistemic Rationality, *Philosophy and Phenomenological Research* 104, 3–18.
Roberts, Tom (2015). Emotional Regulation and Responsibility, *Ethical Theory and Moral Practice* 18, 487–500.
Rosen, Gideon (2004). Skepticism about Moral Responsibility, *Philosophical Perspectives* 18, 295–313.
Ryan, Shane (2018). Epistemic Environmentalism, *Journal of Philosophical Research* 43, 97–112.
Scanlon, Thomas M. (1998). *What We Owe to Each Other*, Cambridge, MA: Harvard University Press.
Scanlon, Thomas M. (2008). *Moral Dimensions. Permissibility, Meaning, Blame*, Cambridge, MA: Belknap Press.
Schmidt, Eva (2023). Facts about Incoherence as Non-Evidential Epistemic Reasons, *Asian Journal of Philosophy* 2, 1–22.
Schmidt, Eva (2024). Replies to Critics, Asian Journal of Philosophy 3 (1), 1–16.
Schmidt, Sebastian (2016). Können wir uns entscheiden etwas zu glauben? Zur Möglichkeit und Unmöglichkeit eines doxastischen Willens, *Grazer Philosophische Studien* 93, 571–582.
Schmidt, Sebastian (2017). Why We Should Promote Irrationality, *Grazer Philosophische Studien* 94, 605–615.
Schmidt, Sebastian (2020a). Responsibility for Attitudes, Object-Given Reasons, and Blame, in *The Ethics of Belief and Beyond. Understanding Mental Normativity*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge: 149–175.
Schmidt, Sebastian (2020b). Rationality and Responsibility, *Australasian Philosophical Review* 4, 379–385.
Schmidt, Sebastian (2020c). Introduction: Towards an Ethics of Mind, in *The Ethics of Belief and Beyond. Understanding Mental Normativity*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge: 1–20.
Schmidt, Sebastian (2022). On Believing Indirectly for Practical Reasons, *Philosophical Studies* 179, 1795–1819.
Schmidt, Sebastian (2023). Incoherence and the Balance of Evidential Reasons, *Asian Journal of Philosophy* 2, 1–10.
Schmidt, Sebastian (2024a). Epistemic Blame and the Normativity of Evidence, *Erkenntnis* 89, 1–24.
Schmidt, Sebastian (2024b). Blameworthiness for Non-Culpable Attitudes, *Australasian Journal of Philosophy* 102, 48–64.

Schmidt, Sebastian (forthcoming a). The Ethics of Belief in a Burning World, *Australasian Philosophical Review*.
Schmidt, Sebastian (forthcoming b). Should We Respond Correctly to Our Reasons? *Episteme*.
Schmidt, Sebastian (forthcoming c). Doxastic Dilemmas and Epistemic Blame, *Philosophical Issues* 34.
Schroeder, Mark (2007). *Slaves of the Passions*, New York: Oxford University Press.
Schroeder, Mark (2009). Means-End Coherence, Stringency, and Subjective Reasons, *Philosophical Studies* 143, 223–248.
Schroeder, Mark (2012). The Ubiquity of State-Given Reasons, *Ethics* 122, 457–488.
Schroeder, Mark (2021). *Reasons First*, New York: Oxford University Press.
Scott, Jordan (2023). Should Epistemic Instrumentalists Be More Social? *Synthese* 201, 1–20.
Sellars, Wilfried (1956). Empiricism and the Philosophy of Mind, in *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, ed. by Herbert Fiegl and Michael Scriven, Minneapolis: University of Minnesota Press, 253–329.
Shah, Nishi (2003). How Truth Governs Belief, *Philosophical Review* 112, 447–482.
Shah, Nishi (2006). A New Argument for Evidentialism, *Philosophical Quarterly* 56, 481–498.
Shah, Nishi, and David Velleman (2005). Doxastic Deliberation, *Philosophical Review* 114, 497–534.
Sher, George (2002). Blameworthy Action and Character, *Philosophy and Phenomenological Research* 64, 381–392.
Sher, George (2006). *In Praise of Blame*, New York: Oxford University Press.
Sher, George (2009). *Who Knew? Responsibility without Awareness*, New York: Oxford University Press.
Sher, George (2021). *A Wild West of the Mind*, New York: Oxford University Press.
Siegel, Susanna (2017). *The Rationality of Perception*, New York: Oxford University Press.
Skorupski, John (2010). *The Domain of Reasons*, New York: Oxford University Press.
Simion, Mona (2024). Resistance to Evidence and the Duty to Believe, *Philosophy and Phenomenological Research* 108, 203–216.
Smartt, Tim (2023). Skepticism about Epistemic Blame, *Philosophical Studies* 180, 1813–1828.
Smith, Angela M. (2005). Responsibility for Attitudes: Activity and Passivity in Mental Life, *Ethics* 115, 236–271.
Smith, Angela M. (2013). Moral Blame as Moral Protest, in *Blame*, ed. by D. Justin Coates and Neal A. Tognazzini, New York: Oxford University Press, 27–48.
Smith, Angela M. (2015a). Attitudes, Tracing, and Control, *Journal of Applied Philosophy* 32, 115–132.

Smith, Angela M. (2015b). Responsibility as Answerability, *Inquiry* 58, 99–126.
Soteriou, Matthew (2020). Determining the Future, in *The Ethics of Belief and Beyond. Understanding Mental Normativity*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge, 234–255.
Steglich-Petersen, Asbjørn (2009). Weighing the Aim of Belief, *Philosophical Studies* 145, 395–405.
Steglich-Petersen, Asbjørn (2011). How to Be a Teleologist about Epistemic Reasons, in *Reasons for Belief*, ed. by Andrew Reisner and Asbjørn Steglich-Petersen, Cambridge: Cambridge University Press, 13–33.
Steglich-Petersen, Asbjørn (2018). Epistemic Instrumentalism, Permission, and Reasons for Belief, in *Normativity: Epistemic and Practical*, ed. by Conor McHugh, Jonathan Way, and David Whiting, New York: Oxford University Press, 260–280.
Steglich-Petersen, Asbjørn, and Mattias Skipper (2019). An Instrumentalist Account of How to Weigh Epistemic and Practical Reasons for Belief, *Mind* 129, 1071–1094.
Steglich-Petersen, Asbjørn, and Mattias Skipper (2020). Instrumental Reasons for Belief: Elliptical Talk and Elusive Properties, in *The Ethics of Belief and Beyond*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge, 109–125.
Stich, Stephen (1990). *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*, Cambridge: MIT Press.
Strawson, Galen (2003). Mental Ballistics or the Involuntariness of Spontaneity, *Proceedings of the Aristotelian Society* 103, 227–256.
Strawson, Peter F. (1962). Freedom and Resentment, *Proceedings of the British Academy* 48, 1–25.
Stroud, Sarah (2006). Epistemic Partiality in Friendship, *Ethics* 116, 498–524.
Sussman, David (2018). Is Agent-Regret Rational? *Ethics* 128, 788–808.
Tobi, Abraham (2022). Appreciative Silencing in Communicative Exchange, *Episteme*, 1–15. Online first, doi: 10.1017/epi.2022.24.
Tognazzini, Neal (2020). Silence and Salience. On Being Judgmental, in *The Ethics of Belief and Beyond*, ed. by Sebastian Schmidt and Gerhard Ernst, New York: Routledge, 256–269.
Toole, Briana (2019). From Standpoint Epistemology to Epistemic Oppression, *Hypatia* 34, 598–618.
Vahid, Hamid (2022). No Practical Reasons for Belief: The Epistemic Significance of Practical Considerations, *Synthese* 200, 1–18.
Vargas, Manuel (2013). *Building Better Beings: A Theory of Moral Responsibility*, New York: Oxford University Press.
Wagner, Verena (2015). On the Analogy of Free Will and Free Belief, *Synthese* 194, 2785–2810.
Wallace, R. Jay (1994). *Responsibility and the Moral Sentiments*, Cambridge, MA: Harvard University Press.
Wallace, R. Jay (2011). Dispassionate Opprobrium: On Blame and the Reactive Sentiments, in *Reason and Recognition: Essays on the Philosophy of T. M. Scanlon*, ed. by R. Jay Wallace, Rahul Kumar, and Samuel Freeman, New York: Oxford University Press, 348–372.

Watson, Gary (1996). Two Faces of Responsibility, *Philosophical Topics* 24, 227–248.
Watson, Gary (2013). Standing in Judgment, in *Blame*, ed. by D. Justin Coates and Neal A. Tognazzini, New York: Oxford University Press, 282–301.
Way, Jonathan (2009). Two Accounts of the Normativity of Rationality, *Journal of Ethics and Social Philosophy* 4, 1–9.
Way, Jonathan (2010). Defending the Wide-Scope Approach to Instrumental Reason, *Philosophical Studies* 147, 213–233.
Way, Jonathan (2016). Two Arguments for Evidentialism, *Philosophical Quarterly* 66, 805–818.
Wedgwood, Ralph (2002). The Aim of Belief, *Philosophical Perspectives* 36, 267–297.
Wedgwood, Ralph (2017). *The Value of Rationality*, New York: Oxford University Press.
Wedgwood, Ralph (2023). *Rationality and Belief*, New York: Oxford University Press.
Wei, Xintong (2022). A Practice-Based Account of the Truth Norm of Belief, *Episteme*, 1–22. Online first, doi: 10.1017/epi.2022.26.
White, Roger (2005). Epistemic Permissiveness, *Philosophical Perspectives* 19, 445–459.
White, Stephen J. (2019). Against Voluntarism about Doxastic Responsibility, *Journal of Philosophical Research* 44, 33–51.
Whiting, Daniel (2010). Should I Believe the Truth? *Dialectica* 64, 213–224.
Whiting, Daniel (2012). Does Belief Aim (Only) at the Truth? *Pacific Philosophical Quarterly* 93, 279–300.
Whiting, Daniel (2013). Truth: The Aim and Norm of Belief, *Teorema* 32, 121–136.
Williams, Bernard (1965). Ethical Consistency, in his *Problems of the Self*, Cambridge: Cambridge University Press, 1973, 166–186.
Williams, Bernard (1981). Moral Luck, in his *Moral Luck*, Cambridge: Cambridge University Press, 20–39.
Williams, Bernard (2002). *Truth and Truthfulness: An Essay in Genealogy*, Princeton, NJ: Princeton University Press.
Williamson, Timothy (2000). *Knowledge and Its Limits*, Oxford: Oxford University Press.
Williamson, Timothy (forthcoming). Justifications, Excuses, and Sceptical Scenarios, in *The New Evil Demon*, ed. by F. Dorsch and J. Dutant, New York: Oxford University Press.
Wolf, Susan (1994). *Freedom within Reason*, New York: Oxford University Press.
Wolf, Susan (2011). Blame, Italian Style, in *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, ed. by R. Jay Wallace et al., New York: Oxford University Press, 332–347.
Worsnip, Alex (2021). *Fitting Things Together. Coherence and the Demands of Structural Rationality*, New York: Oxford University Press.

Index

Note: Page numbers followed by "n" refer to end notes.