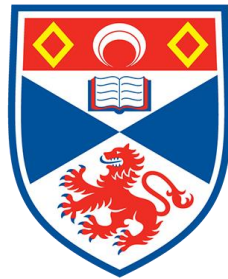# Is it OK to Make Mistakes? Appraisal and False Normative Belief.

Claire Anne Field

University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

June 2019

Candidate's declaration

I, Claire Anne Field, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 63,700 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in September 2015.

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

Date   Signature of candidate

Supervisor's declaration

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

Date   Signature of supervisor

Permission for publication

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

I, Claire Anne Field, confirm that my thesis does not contain any third-party material that requires copyright clearance.

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

Printed copy

No embargo on print copy.

Electronic copy

Embargo on all of electronic copy for a period of 1 year on the following ground(s):

- Publication would preclude future publication

Supporting statement for electronic embargo request

I would like to publish the material from my thesis in the next year.

Title and Abstract

- I agree to the title and abstract being published.

Date   Signature of candidate

Date   Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

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

Date   Signature of candidate

Contents

Acknowledgements

General Acknowledgements

This thesis represents an intellectual journey that began a long time ago, and I am grateful to the many others who have helped me along the way. Here, I acknowledge those to whom I am particularly indebted. Primarily, my primary supervisor Jessica Brown, for invaluable guidance and unending patience. Additionally, the other members of my supervisory team, Justin Snedegar and Brian Weatherson, as we all Clayton Littlejohn, who is responsible for supervising what became the beginnings of this project. Many other philosophers have also helped, and most of them can be found at the Philosophy Departments of St Andrews, Stirling, Michigan, and KCL. Finally, the journey would have been considerably more difficult and much less fun without the love and support of my friends and family. In particular: Marvin Backes, Lisa Bastian, Miguel Egler, Andy Field, Mike Field, Toni Field, Zoë Johnson-King, Ethan Landes, Stefano Lo Re, Sean Low, Quentin Pharr, Alistair Redgrift, Lewis Ross, Cat Saint-Croix, Clotilde Torregossa, Giulia Vivaldi, and Mike Willdig.

Publications

Abstract

Sometimes we make mistakes, even when we try to do our best. When those mistakes are about normative matters, such as what is required, this leads to a puzzle. This puzzle arises from the possibility of misleading evidence about what rationality requires. I argue that the best way to solve this puzzle is to distinguish between two kinds of evaluation: requirement and appraisal. The strategy I defend connects three distinct debates in epistemology, ethics, and normativity: the debate over how our theories of epistemic rationality should accommodate misleading evidence, the debate over the relationship between complying with requirements and deserving particular appraisals, and the debate over whether normative ignorance can excuse. Part 1 shows how three apparently plausible claims about epistemic rationality generate a puzzle when agents have misleading evidence about what rationality requires. Part 2 solves this puzzle by distinguishing between evaluations of requirement and appraisal and rejecting the idea that one is required to conform to the Enkratic Principle. I argue instead that complying with the Enkratic Principle provides defeasible evidence that the agent should be positively appraised. One of the consequences of this solution is that false normative beliefs can sometimes excuse agents from negative appraisal they would otherwise deserve for violating requirements. Part 3 defends the view that false normative belief can sometimes excuse against the rival views that false normative belief always excuses, and that false normative belief never excuses. I argue that false normative belief can sometimes excuse violations of requirements, when it is the case that the agent has done what it is reasonable to expect of her.

List of Cases

Introduction

Sometimes we make mistakes, even when we try to do our best. When those mistakes are about normative matters, such as what is required, this leads to a puzzle. This puzzle arises from the possibility of misleading evidence about what rationality requires. I argue that the best way to solve the puzzle is to distinguish between two kinds of evaluation: requirement and appraisal.

Part 1 shows how three apparently plausible claims about epistemic rationality generates a puzzle when agents have misleading evidence about what rationality requires. For example, suppose that rationality requires believing P. Then suppose that one has misleading evidence that in fact, rationality prohibits believing P. What does rationality require in this situation? It seems to require both believing P, since this is stipulated as a requirement of rationality; and refraining from believing P, since this is what is recommended by the evidence.

Part 2 argues that the best way to solve the puzzle is to reject the idea that one is required to conform to the Enkratic Principle. Instead, I suggest that whether one conforms to the Enkratic Principle provides defeasible evidence that agent should be positively appraised. This approach distinguishes between two kinds of agent evaluation: those that consider whether the agent has done what is required of her; and those that consider how the agent is to be appraised. This allows us to accommodate the intuitions that have contributed to the Enkratic Principle's popularity as a requirement of rationality, while avoiding the costs of existing solutions. One of the consequences of this solution is that false normative beliefs can sometimes excuse agents from negative appraisal they would otherwise deserve.

Part 3 develops further the solution argued for in Part 2, which suggested that false normative beliefs can sometimes excuse agents from negative appraisal they would otherwise deserve. This claim requires further defence, since whether or not false normative belief can excuse is controversial. Against existing views, Part 3 argues that it is neither the case that false normative belief always excuses, nor the case that it can never excuse. Rather, false normative belief can sometimes excuse, specifically when it is the case that the agent has done what it is reasonable to expect of her. This, in turn, is determined by the agent's capacities and roles, where this includes her psychological capacities. Two significant consequences of this account are that normative mistakes can sometimes excuse, and that what it is reasonable to expect of agents is sometimes more than we typically do expect.

Part 1

Chapter 1

Rational Mistakes About What Rationality Requires

Sometimes we make mistakes, even when we do our best. When those mistakes are about what epistemic rationality requires, this leads to a puzzle: "The Puzzle of Rational Requirement". Solving the Puzzle of Rational Requirement is the focus of this thesis. Part 1 of the thesis outlines the puzzle, and dismisses objections to the idea that there is a puzzle at all. Parts 2 and 3 defend a solution to the puzzle. This chapter focuses on outlining the puzzle, and dismissing some attempts to dissolve it.

Section 1 of this chapter sets out the puzzle's structure, distinguishing it from some importantly different puzzles that it may initially seem similar to. Section 2 identifies and discusses the motivations for three commitments necessary to motivate the puzzle: Externalism, Evidentialism, and the Enkratic Principle. Sections 3 and 4 dismiss two attempts to dissolve the puzzle by avoiding the problem it appears to present for our theorising about normative requirements of epistemic rationality. Section 3 dismisses Dilemmism – the strategy of embracing the apparent conflict as an epistemic dilemma. Section 4 dismisses indexing strategies that seek to separate the apparently conflicting requirements and associate them with different senses of ought, or different linguistic contexts. The following chapter deals with another attempt to avoid the puzzle: appeal to the Impossibility Thesis, which denies that mistakes about what rationality requires could ever be rational.

1. The Puzzle of Rational Requirement

The Puzzle of Rational Requirement has the following structure. Suppose that you are required to Φ, in virtue of some set of normative requirements. Then,

suppose that you have misleading evidence that in fact, you are prohibited from Φ-ing. Are you required to Φ? Or are you required to refrain from Φ-ing, as your evidence indicates? On the one hand, you seem to be required to Φ, since this is what the normative requirements in fact require. However, you also seem to be required to refrain from Φ-ing, since this is what your evidence indicates. These two answers are incompatible, thus presenting us with a puzzle. Here is an example involving the requirements of epistemic rationality:

> *Logic 101*. Suppose that rationality prohibits contradictory belief. Suppose also that it is your first day of university, and you are about to take your first philosophy class. You know nothing about philosophy, logic, or epistemic rationality, but you want to believe rationally. You believe that learning some logic will help you to do this. The class you sign up for is an introductory course in logic, and your instructor is an overzealous advocate of dialetheism[1]. He believes that rationality sometimes requires inconsistent belief, particularly in matters concerning truth[2], and he intends to set you on the right track by introducing you to the best arguments in favour of this position. In class, you study in depth all the best arguments for dialetheism, and you leave the class believing – on the basis of good but misleading evidence – that rationality sometimes requires contradictory belief[3]. As you walk out of class you see some graffiti that is a version of the Liar paradox ('The writing on this wall is false.').

---

[1] As developed in Priest (2005; 2006).

[2] See for example "it seems to me that anyone weighing up the state of play concerning [truth], ought rationally to be inconsistent." (Priest 2006: 125).

[3] For Priest's views on the specific question of what rationality requires dialetheists to believe, see Priest (1985).

Epistemic rationality, let us suppose, requires you to refrain from believing contradictions. However, you have misleading evidence indicating that in this case rationality requires you to believe a contradiction. This is an example of apparent intra-domain conflict between normative requirements. In this instance, the requirements of epistemic rationality seem to simultaneously require you to believe and *not* believe a contradiction. Assuming that epistemic rationality prohibits contradictory belief, two further putative requirements of rationality that contribute to making it the case that you are rationally required not to believe a contradiction in this case – an evidential requirement, and a requirement to be enkratic. We can express these as follows[4]:

> *Evidentialism*: S(e)p → O(S(Bp)). If S has evidence that supports
>
> P, then S ought to believe that P[5].

> *Enkratic Principle*: O (BOΦ → Φ). Rationally requires that agents
>
> do as they believe they ought to do[6].

These generate the requirement to believe a contradiction in the following way. By Evidentialism, you ought to believe what your evidence supports. In this case, your evidence supports the belief that rationality requires you to believe a contradiction – so, in this case, you ought to believe that you are rationally

---

[4] Throughout the thesis, unless specified otherwise, I use "required" and "ought" to express the normative concept.

[5] This is to be distinguished from the stronger bi-conditional claim defended in Conee and Feldman (2004), according to which S ought to believe P if *and only if* P is supported by her evidence ("Doxastic attitude D toward proposition p is epistemically justified for S at t if and only if having D toward p fits the evidence S has at t" (2004: 83). See also Clifford's claim that, "it is wrong always, everywhere, and for anyone to believe anything upon insufficient evidence" (Clifford (1877)). Note, however that Clifford is here referring to moral rather than epistemic wrongness.

[6] Or, at least that they intend to do as they believe they are rationally required to do. As Broome (1999) argues, the proper formulation of the Enkratic Principle for practical reasoning should take intentions, rather than actions, in the consequent; whether or not an agent succeeds in performing an action is not a matter of rationality, but her intentions are. I leave this complication aside here.

required to believe a contradiction. By the Enkratic Principle, you ought to do what you believe you ought to do. When the Enkratic Principle is thought of as a wide scope principle, as it is here, there are two ways to conform to it: by Φ-ing or by giving up the belief that you are rationally required to Φ. Here, you can conform with the Enkratic Principle either by refraining from believing contradictions, or by giving up the belief that you are rationally required to believe a contradiction. However, in this case your evidence supports the belief that you are rationally required to believe a contradiction, so giving up that belief would be to ignore the evidence and violate Evidentialism. So, in this instance, conforming to both Evidentialism and the Enkratic Principle means that you ought to believe the contradiction. However, your unfortunate epistemic situation – the fact that you have being exposed to arguments for a false philosophical view – does not change what rationality requires. If rationality really prohibits contradictory belief, then your unfortunate epistemic situation does not change this. This generates the puzzling result that when you have misleading evidence about what requires, then rationality appears to make conflicting demands of you: in this case requiring you to both believe and *not* believe the contradiction.

It is worth pointing out that generating the puzzle does not depend on the possibility of rationally doubting any particular principle of classical logic[7]. The following is also an example of the puzzle[8]:

> *Disagreement.* Suppose that rationality requires you to remain steadfast in your beliefs when epistemic peers disagree with you. Then suppose you write your PhD on the epistemology of disagreement. You exert significant effort thinking about and developing arguments that bear on the question of whether one should conciliate or remain steadfast in the face of disagreement. You reach the conclusion that rationality requires that we conciliate in the face of disagreement from epistemic peers[9]. Over dinner one evening, while discussing politics, you assert P (in which you have a credence of 0.8). Your partner (who is your epistemic peer on this matter) disagrees with you.

Let us suppose that you are wrong, and epistemic rationality requires you to remain steadfast in your beliefs. By Evidentialism you ought to believe what your misleading evidence supports – that rationality requires you to conciliate in the

---

[7] Some have argued that the irrationality of contradictions is too certain to be rationally doubted. For example, Putnam (1978) argues that his minimal principle of contradiction ('not every statement is true and false') cannot be rationally doubted because it presupposes the possibility of debate, thought, and explanation. Lewis declines to engage in a debate over the logical law of non-contradiction on the grounds that "the principles *not* in dispute are so very much less certain than non-contradiction itself that it matters little whether or not a successful defence of non-contradiction could be based on them" (2004: 176). The possibility of apparent intra-domain conflict between normative requirements does not turn on these issues.

[8] For further examples see Littlejohn (2015), who motivates the puzzle using the requirement to believe (or refrain from believing) lottery propositions to motivate the puzzle; Bradley, who motivates his version of the puzzle using rational requirements governing perception, testimony, and Lewis' Principal Principle (2019: 2); and Feldman who makes use of external world scepticism (2005).

[9] We can imagine that one of your reasons for thinking this is that you think that intellectual modesty is a more important feature of epistemic rationality than avoiding theories that are self-undermining (as Elga (2007) argues is a reason to reject conciliationism).

face of disagreement. By the Enkratic Principle you ought to either do what you believe you ought to do, or give up the belief that you ought to do it. So, in this instance, you ought to reduce your credence in P – to give up the belief that you ought to conciliate would be to violate Evidentialism, since this belief is supported by your evidence. However, the fact that you have being exposed to arguments for a false philosophical view does not change what rationality requires of you; it does not change the fact that you ought to remain steadfast (on the assumption that this is something rationality requires). So, in this instance, rationality appears to make conflicting demands of you: you are required to both reduce and *not* reduce your credence in P.

These examples of apparent in*tra*-domain conflict that motivate the puzzle are importantly different from a similar, and more benign, puzzle involving in*ter*-domain conflict between normative requirements. Here is an example involving the requirements of morality and epistemic rationality:

> *Lying.* Suppose that morality requires that you act so as to protect the lives of innocents, in all situations – even if this would sometimes require you to lie. Suppose also that you have recently taken a course on ethics led by a professor who defends a somewhat extreme Kantian view on the moral permissibility of lying. According to this view, lying is morally wrong in all cases, even if by lying you could save a life. In the ethics course, your professor introduces you to various arguments in favour of this view. You desire to do the morally right thing, and you know that morality is difficult, and sometimes at odds with our intuitions. You are a good student – you do the reading, you follow the arguments you are presented with in class, and under the guidance of the professor, you come to believe that lying is

morally impermissible in all cases. One day, an infamous axe

murderer comes to your door, in pursuit of your friend Ricardo.

In fact, Ricardo is hiding in your house. You know that you could

easily save his life by lying to the murderer[10].

Morality, we suppose, requires that you lie to protect Ricardo. However, you have misleading evidence indicating that morality prohibits lying. In this instance, morality requires you to lie, but rationality seems to require you not to lie (by Evidentialism and the Enkratic Principle). This is a case of int*er*-domain (rather than int*ra*-domain) conflict, and is thus importantly different from the Puzzle of Rational Requirement, which occurs entirely within the domain of epistemic rationality. We should find the Puzzle of Rational Requirement more worrying than the inter-domain conflict between the normative requirements involved in Lying. The observation that morality and rationality sometimes conflict is not, itself, surprising. Morality and rationality are distinct normative domains, and it is not obvious that we should expect them to be consistent. To resolve this inter-domain conflict, we need only decide which set of normative requirements to prioritise in this case – those of morality or rationality. While this is not always a straightforward question[11], we can answer it without revising our theories of the normative requirements of morality or rationality. However, the same cannot be said for the Puzzle of Rational Requirement motivated by cases such as Logic 101 and Disagreement, which involve apparent intra-domain conflict. Unlike int*er*-domain conflict, the appearance of int*ra*-domain conflict

---

[10] See Kant (1797). For an approach that defends the extreme interpretation of Kant mentioned here, on which Kant prohibits lying even in this case, see Schwarz (1970), Weinrib (2008). The tendency among Kantians is to attempt to save Kant from any commitment to prohibit lying in this case – that is, to make lying in circumstances such as these at least permissible, if not obligatory (see Korsgaard (1986); Schapiro (2006); Wood (2008, Ch. 14)). Thanks to Stefano Lo Re for making me aware of these points.

[11] See Broome (2013; 1999), as well as Harman (2015) for some useful discussions of this.

implies that our current theories of rationality have internal inconsistencies which demand resolution.

The Puzzle of Rational Requirement is also importantly different from puzzles involving misleading higher-order evidence[12]. The following is a typical example:

> *Medicine.* You are a medical doctor who diagnoses patients and prescribes appropriate treatment. After diagnosing a particular patient's condition and prescribing certain medications, you are informed by a nurse that you have been awake for 36 hours. You diagnose a patient with Disease D, forming the belief "patient has D". In fact, you are correct. However, you know that people are prone to making cognitive errors when sleep-deprived (perhaps you even know your own poor diagnostic track-record under such circumstances), so you also believe "I ought not believe that patient has D (because it is not supported by my evidence)". (see Christensen, (2010a: 186); Weatherson (2019: 130))[13].

In this case, you have good evidence for P, but misleading evidence about your current ability to assess the evidence for P effectively[14]. So, your evidence appears to support both of the following propositions[15]:

The patient has disease D.

My evidence does not support that the patient has disease D.

If these beliefs are supported by your total evidence, then Evidentialism says that you ought to believe both. However, this combination of beliefs is prohibited by

---

[12] This similar but distinct puzzle is discussed by (Christensen 2010a; Elga 2013; Horowitz 2014; Lasonen-Aarnio 2014; Sliwa & Horowitz 2015; Weatherson 2019; Worsnip 2018).

[13] For further discussion of examples of this kind, see (Christensen 2010b; Elga 2013; Lasonen-Aarnio 2014; Williamson 2011).

[14] See also, Christensen's Hypoxia case (2010), Lasonen-Aarnio's various cases (2010; 2014), as well as Williamson's unmarked clock case (2013).

[15] This is the view of what the total evidence supports according to (Christensen 2010; Lasonen-Aarnio 2014; Lasonen-Aarnio, forthcoming; Weatherson 2019).

the Enkratic Principle. That it is possible for one's total evidence to support such combinations is controversial[16]. In contrast, the Puzzle of Rational Requirement does not rely on this controversial thesis about evidential support. Instead, in addition to a commitment to Evidentialism and the Enkratic Principle, all that is needed to generate the puzzle is at least one requirement of rationality, other than the Evidential requirement, and some misleading evidence about what rationality requires. This makes possible situations in which the agent is rationally required to believe a false claim about what rationality requires (because this is what her evidence supports), and also rationally required (by the further requirement of rationality) to adopt a conflicting belief at the first order. For example, the belief that one is rationally required to believe a contradiction, but *no* first order belief in that contradiction. As it is set up here, this possibility is independent of any particular claims about what one's total evidence can support.

The Puzzle of Rational Requirement concerns apparent intra-domain conflict between normative requirements, of which Logic 101 and Disagreement are examples. This puzzle suggests internal inconsistency within our theory of rationality, and so it demands resolution. The following section outlines in more detail the exact commitments that the Puzzle of Rational Requirement depends on, and explains the motivations for accepting each one.

2. Three Commitments

The Puzzle of Rational Requirement depends on our endorsing the following three commitments:

---

[16] For some disagreement, see (Brown, 2018 (Ch. 5, 6); Horowitz, 2014; Sliwa & Horowitz, 2015). This also turns on the question of whether or not higher order evidence `screens off' first order evidence. For discussion of this, see (Feldman 2007; Fitelson 2012; Roche & Shogenji 2013; Roche 2018).

Externalism: What rationality requires is completely determined by the facts about what rationality requires ("the rational requirements")[17].

Evidentialism: S(e)p → O(S(Bp)). If S has evidence that supports P, then S is rationally required to believe that P.

Enkratic Principle: O (BOΦ → Φ). Rationally requires that agents do as they believe they ought to do.

Within the epistemic domain, the Enkratic Principle demands coherence between the agent's epistemic attitudes. Here, the Enkratic Principle is stated in a way that is relevant to the domain of epistemic rationality. Epistemic attitudes could include believing, refraining from believing, disbelieving, suspending, or having a particular credence in some proposition. In the formulisation above, 'Φ' could refer to any of these. This section motivates each of these commitments, and outlines in more detail how commitment to them generates the puzzle.

2.1 Externalism

First, Externalism. Externalism says that there is a fact of the matter about what rationality requires. In other words, there is a set of fixed rational requirements, setting out what it is that agents ought to do, if they are to count as rational[18]. Externalism is the claim that what rationality requires is objective, rather than

---

[17] One way to put this is that the requirements of rationality are `indefeasible' – that is, they remain binding in all possible situations. See (Bradley 2019; Ichikawa & Jarvis 2013) for examples of this way of thinking about requirements of rationality.

[18] By `rational', I mean fully rational. Full rationality is an absolute, binary notion that is to be distinguished from partial rationality, a graded notion. Throughout the discussion, I shall be concerned with the requirements of full rationality. However, this does not preclude the possibility of agents often being partially rational when they, for example, succeed in meeting some but not all of the requirements of rationality, or have some but not all of the attitudes specified by one of the requirements of rationality. What partial rationality requires is an important question that I will not pursue further (for some discussion of this, see Wedgwood 2017). Complete compliance with the requirements of rationality is required for full rationality, such that any failure to comply with a requirement of rationality is a failure to be fully rational.

subjective. That is, whether an agent is epistemically rational depends on the facts about what rationality requires and not on the agent's perspective regarding what rationality requires[19]. Externalism is to be contrasted with a perspectival view of rationality, according to which the content of the requirements of rationality can be affected by particular features of one's perspective, such as what one believes is rationally required, what one's evidence is, or how things appears to one[20]. According to Externalism, the requirements of rationality are universally binding on all agents[21]. So, for all agents, failure to conform to the

---

[19] One might think that if Evidentialism is a requirement of rationality, then there *is* a weak sense in which the requirements of rationality depend on your epistemic perspective: by Evidentialism, if E is part of your evidence, then you ought to believe what E supports. Since E is part of your epistemic perspective, your epistemic perspective means that you are rationally required to believe what E supports, and so in this sense, your epistemic perspective fixes what rationality requires of you in this case. This is correct. However, it is not the case that your epistemic perspective affects the content of the requirements of rationality – these remain fixed. One could, perhaps, insist that there is nevertheless a sense in which Evidentialism does imply that the content of the rational requirements is affected by one's perspective. Namely, that for every piece of one's evidence, E, one is subject to a specific requirement to believe what E supports. However, this kind of weak dependence on one's perspective is significantly different to the kinds of perspectival dependence that Externalism is to be contrasted with, and does not seriously threaten Externalism.

[20] Those who take what epistemic rationality requires to depend on the agent's perspective include H. Field (2009); Gibbons (2013); Kiesewetter (2011; 2013); Kolodny (2005); Kvanvig (2003); Lord (2018); Raz (2005); Way & Whiting (2017); Whiting (2014); Zimmerman (2008). For further discussion of whether what rationality requires is determined by the facts or by the agent's perspective, see Feldman (2005; 1988b); Littlejohn (2015; 2011; 2012); Titelbaum (2015b); Way & Whiting (2016; 2017).

[21] By 'agents' I mean agents of the kind that are properly subject to the requirements of rationality. Not all beings are necessarily such agents. Babies, for example, are not the kinds of beings that it makes sense to evaluate using the requirements of rationality, but they may be agents in the more limited sense of being possible authors of actions. Group agents, also, are not necessarily appropriate subjects of precisely the same requirements of rationality as individual agents. For example, unlike individual agents, groups may simultaneously hold many inconsistent beliefs without irrationality (see Condorcet (1995) for some discussion of paradoxes arising from applying traditional consistency requirements to beliefs). Of course, this also depends on how group belief is to be understood (for discussion of this, see Lackey (2016); Mathiesen (2006); Skipper & Steglich-Petersen (2019).

requirements of rationality is a failure to be rational, regardless of the reason why[22].

The following paragraphs provide some motivations for endorsing Externalism. Those who remain unconvinced by these motivations might take the puzzle that arises from commitment to Externalism as further reason to reject it. One reason to endorse Externalism is that having a fixed set of standards that can be used to evaluate an agent's epistemic activities independently of her particular epistemic situation is useful. At least, it is useful in so far as we are interested in the standards traditionally associated with epistemic rationality – logical consistency, probabilistic coherence, evidential support. Externalism allows us to measure epistemic states that approximate, but ultimately fall short of, these standards without needing to worry about whether or not the agent has the capacities to meet the standards[23]. As Christensen puts it, the standards of epistemic rationality, "need not grade on effort" (2004: 162): we need not consider the agent's capacities when evaluating whether her beliefs are rational[24].

Situation independent standards are useful in various arenas – in chess, we can specify the moves that a player ought to play in order to win, and these same moves are required regardless of whether the player is an expert or a mediocre chess player. Likewise, in ice skating, a figure skater of modest ability can be evaluated on an attempt to perform a triple axel by reference to the moves and

---

[22] This notion, which is part of Externalism, has also been expressed as the idea that rational requirements impose strict liability (Broome (2013); Littlejohn (2015)), or that they are 'indefeasible' (Ichikawa & Jarvis (2013: 33)). That is, agents are subject to them regardless of what they believe or intend.

[23] Which, as is well documented, it is not obvious that ordinary human agents have (Kahneman et al. 1982; Tversky & Kahneman 1993; Cohen 1981).

[24] Some have used these considerations to argue for accounts of epistemic rationality that require agents to have epistemically ideal states (Wedgwood 2017; Christensen 2004; Smithies 2015). However, the claim that requirements of rationality are external is to be distinguished from the claim that they require ideal states. This discussion is neutral on what the true requirements of rationality require, and so also neutral on whether rationality requires ideal epistemic states.

form that are required for a triple axel. The moves required for a triple axel are the same regardless of whether the skater is able, given her present situation and capacities, to execute it. For example, considerations such as that the skater is not currently wearing her skates, or that she is out of shape, or that she has a false belief about how many rotations a triple axel requires, do not affect what a triple axel requires. Rational belief can be thought of similarly. Just as for chess and figure skating, the requirements of fully rational belief are understood here as evaluative[25], and propositional, rather than doxastic[26]. That is, they state the attitudes that are required for an agent's belief to be evaluated as fully rational, and they are unaffected by considerations pertaining to the agent's abilities to now believe, in a way that is doxastically justified, what rationality requires. In so far as one accepts this claim about the standards of rationality, one is committed to Externalism.

However, one might object that the requirements of rationality are importantly different from chess or figure skating. Epistemic rationality might seem to differ from chess or figure skating in that it takes as its primary concern the agent's first person perspective[27], such that to believe rationally is to have the beliefs that make the most sense from your perspective, not to have the attitudes demanded by an external set of standards. This would make a commitment to Externalism inappropriate. However, there are good reasons to be suspicious of this perspectival view of epistemic rationality.

---

[25] See Steinberger (2019) for a helpful discussion of the distinctions between the evaluative, prescriptive, and appraising senses of norms. A commitment to Externalism, as I shall understand it, is a commitment to the claim that the requirements of rationality are a set of evaluative norms. This is compatible with there also being sets of norms that provide prescriptions, and appraisals pertaining to rationality, but these will not be the focus of the discussion.

[26] See Lord (2018: 9-10) for further discussion of the distinction between propositional (or 'ex post') and doxastic (or 'ex ante') rational requirements.

[27] As Kvanvig calls it, the "egocentric predicament" (Kvanvig 2014: 48).

First, perspectival views of what rationality requires need to define what counts as making sense according to the agent's perspective in such a way that both avoids the absurdity of any belief at all being potentially intelligibly from a perspective, and of involving constraints that are themselves fairly externalist. Any theory of what rationality requires faces pressure to avoid collapsing into triviality, such that any epistemic attitude or attitude combination is potentially rational. To avoid this, a perspectival view needs to introduce some constraints on which states can count as rational, in order to avoid the possibility that any attitude at all could count as rational. However, as soon as we begin introducing potential constraints we risk introducing requirements of rationality that are more external than perspectivism would like. For example, a possible constraint on a perspectival view of rationality could be expressed as follows:

> *Perspectivism*: If P appears to be supported by your epistemic
>
> reasons then rationality requires you to believe P[28].

This would avoid triviality by introducing constraints on what can count as appearing to be supported by the agent's perspective. However, what it is for P to be supported by your epistemic reasons, what it is for P to appear to be supported by your epistemic reasons[29], and what it is for something to be one of your epistemic reasons are all matters that do not depend *only* on one's epistemic perspective. Furthermore, on many understandings of what it is for a consideration to count as one's epistemic reason, that consideration must be a

---

[28] Huemer (2007; 2011) defends a weaker version of this claim, according to which if it seems to you that P, you have some defeasible justification to believe P. For further discussion of this phenomenal conservativist view see DePoe (2011); Littlejohn (2011); Hasan (2013); Lycan (2013).
[29] For a helpful discussion of how to understand one's apparent reasons, and how they bear on one's rationality, see Sylvan (2015).

fact, so, external to the agent's perspective[30]. Even if reasons are not facts, there will be a fact of the matter about which of the agent's mental items count as reasons – that is, which epistemic reasons are sufficiently accessible to count as reasons. So, a constraint such as the above, which is able to avoid triviality, may not turn out not to be quite as perspectivist as we might have wanted.

Second, endorsing perspectivism would mean holding an `error theory' of traditional rational requirements (see Lasonen-Aarnio (2014: 332)). If what rationality requires depended directly on the agent's perspective, this would mean that a statement of what rationality required of an individual would be impossible to make in advance of considering the agent's precise situation. This would, surprisingly, make cases of misleading evidence about what rationality requires counterexamples to particular requirements of rationality. For example, Logic 101 would be a counter-example to the putative requirement not to believe contradictions. On a perspectivist view, Logic 101 would show that while we might have thought that agents were always required to refrain from believing contradictions, this is false, because there can be perspectives that support contradictory belief. This would constitute a significant revision to how we ordinarily think of rational requirements, and so we would need very good reason to think it necessary.

Third, endorsing perspectivism about requirements of rationality would prevent requirements of rationality from playing the role of guiding agents towards more rational attitudes[31]. We would not be able to, for example, advise agents to avoid

---

[30] For defenders of this view, see (Alvarez 2010; Dancy 2000; Parfit 2011; Raz 1975; Skorupski 2010). Some accept this view while demanding an additional epistemic condition, for example Dancy's claim that for a fact to provide one with a reason, it must pass through an 'epistemic filter'. That is, it must be knowable (see Dancy 2000: 57-59).

[31] Despite this, some have used the thought that norms should be guiding to argue *for* perspectivism (see, for example, Gibbons (2013)). For some reasons to think this fails, see Way and Whiting (2017).

contradictory belief, since there would be possible perspectives on which contradictory belief would be rational. If we thought that the agent's perspective is the arbiter of what it is rational for her to believe, then the only true requirement of rationality would be a very long and complicated Uber-Rule, specifying exactly what agents should believe given different situations[32], or perhaps only the enigmatic requirement `do what the situation requires'. Neither of these would be of much use in guiding agents towards believing more rationally.

Externalism avoids the difficulties mentioned here. It implies that the requirements of rationality are binding for all agents and in all situations. So, if rationality requires that you refrain from believing contradictions, then you ought to refrain from believing contradictions – even if you have misleading evidence that indicates otherwise.

2.2 Evidentialism

Second, Evidentialism. Evidentialism says that rationality requires that we believe what our evidence supports. As stated here, Evidentialism is deliberately uncontroversial[33]. Various worries have been raised against principles that, at first glance, might seem equivalent to Evidentialism. This section addresses the most important of these, and explains how Evidentialism, as it is stated here, avoids them.

Firstly, it might be thought that Evidentialism is too strong as a necessary condition of epistemic rationality, because no ordinary human agent could be

---

[32] See Bradley (2019) for a defence of Uber-Rules in epistemic rationality, as well as Dancy (2004) for a particularist account of ethical requirements. For criticism of the viability of Uber-Rules in epistemology, see (Lasonen-Aarnio 2014; Littlejohn 2015).

[33] For example, Williamson (2002: 164) and Kelly (2006) both call it a "platitude", and it is widely accepted by various epistemologists who have otherwise divergent views (see, amongst many others, Feldman and Conee (1985); Fantl and McGrath (2009); Greco (2014); Huemer (2011a); Kelly (2002); Smithies (2012).

expected to believe everything that her evidence supports – this would be too demanding[34]. The force of this worry depends on our understanding of both what it is for E to be part of one's evidence, and what it is for E to support P. What it is for E to be part of one's evidence is typically understood in a way that restricts what can count as one's evidence to items that are, in some sense, epistemically available to the agent. For example, the agent's beliefs, her beliefs and her experiences[35], what she is in a position to know[36], or what she knows[37]. However, these restrictions alone do not avoid the problem. If we employ a standard notion of evidential support from decision theory, according to which E supports P when it makes P more likely (Williamson (2002, Ch. 10)), then even with these restrictions Evidentialism would imply that agents are required to believe all the logical consequences of their evidence. All of the logical consequences of P are entailed by P, so they are all afforded the strongest possible support by our evidence. If this is correct, then Evidentialism requires us to have an infinite number of beliefs. However, that this is the correct view of evidential support is disputable. Conee and Feldman deny that we ever have evidence for an infinite number of beliefs, on the grounds that E must do more than make P likely, it must also make P evident in some way[38]. However, we might worry that restricting the notions of evidence and evidential support cannot completely avoid the problem of demandingness, because even with the restricted notions, the agent will not always be in a position to discern either what is contained in

---

[34] See Kornblith (1983: 46); Thagard (1982: 34)).

[35] See Conee and Feldman (2004).

[36] As Skorupski puts it, only facts within our `epistemic field' (2010: 42).

[37] See Williamson (2002).

[38] As they put it, "the evidence that people have under ordinary circumstances never makes it evident, concerning every one of an infinite number of logical consequences of that evidence, that it is a consequence" (2004: 87).

that restricted set, or what that restricted evidence set supports[39]. If one cannot discern what one's evidence is or what it supports, then it will always seem too demanding to expect the agent to believe what her evidence supports.

Another respect in which Evidentialism can seem too demanding is in its judgments of agents who have evidence for P, but have not yet come to believe P. For example, suppose a detective is working through a case, and has not yet reached a conclusion about who committed the crime. In fact, her evidence supports the conclusion that Stephanie committed the crime, but she has not realized this yet. We might think it inappropriate to say that the detective has failed to meet the requirements of rationality, it is just that she has not finished thinking yet.

Fortunately, these worries that Evidentialism is too demanding can be met by employing a distinction between evaluating (1) which attitudes are required by rationality, and (2) whether the agent employs good epistemic conduct in forming the attitudes she forms. Employing this distinction allows us to acknowledge that while it is true that believing that Stephanie is the thief is the attitude that epistemic rationality requires of the detective, we need not evaluate her failure to *now* believe this negatively. Chapter 3 further motivates, outlines, and argues for this distinction.

Secondly, we might worry that Evidentialism leaves out some important further requirements on rational belief. I will mention two putative further requirements. The first is the requirement that agents gather evidence in a responsible way[40]. Consider Kornblith's case:

> *Physicist.* Jones is a headstrong young physicist, eager to hear the
>
> praise of his colleagues. After Jones reads a paper, a senior

---

[39] This is what Williamson' anti-luminosity arguments show (2002, Ch. 4). I discuss these and their implications in more detail in Chapter 3.
[40] For example, see Clarke (1984; Kornblith (1983: 34-5)).

colleague presents an objection. Expecting praise and unable to tolerate criticism, Jones pays no attention to the objection; while the criticism is devastating, it fails to make any impact on Jones' beliefs because Jones has not even heard it (1983: 36).

Jones' conduct clearly seems epistemically irresponsible, and we might take this to imply that it is also epistemically irrational. Since Jones does not hear the objection, it does not count as part of his evidence on the majority of understandings of what counts as one's evidence: he does not believe, experience, or know the objection[41]. So, on these understandings, Jones does not violate Evidentialism. We might think that this indicates a problem with Evidentialism. However, three points are worth noting as a way of dismissing this objection. First, as it is used here, Evidentialism is not intended to capture a complete description of what epistemic rationality requires – it is compatible with there being further requirements governing responsible evidence gathering. Second, Kornblith describes this case using hypological terms – for him, it is a case of "epistemically culpable ignorance" (1983: 36). That responsible evidence gathering practices fall under the domain of epistemic praise and blame, rather than epistemic requirement is part of the view defended here, and is discussed further in Chapter 3. Third, the cases of misleading evidence about what rationality requires that motivate the puzzle are all cases in which the agents have gathered their evidence in an epistemically responsible way, and they could not be reasonably expected to gather more evidence. So, even if we were to amend Evidentialism so that it required responsible evidence gathering, such a requirement would be met by the agents in the cases at issue.

---

[41] However, we may think that it is knowable, and that this makes it part of his evidence (see Skorupski (2010)).

Thirdly, one might think that the possibility of pragmatic encroachment threatens the plausibility of a requirement such as Evidentialism. For example, some have thought that P's being supported by my evidence will not always be sufficient to justify belief in P, because there are also pragmatic constraints on justification (Fantl & McGrath (2002)). If this is right, then there will be cases in which P is supported by my evidence, but I would not be justified in believing P, because the pragmatic stakes are too high. For example, consider the following pair of cases:

> *Train Case 1.* You're at the station, preparing to take the train to see friends. You ask a fellow passenger, "Does this train stop in Foxboro?" It doesn't matter much to you whether it does or not. He answers, that it does. Nothing about him seems particularly untrustworthy.
>
> *Train Case 2.* You absolutely need to be in Foxboro, the sooner the better. Your career depends on it. You overhear a conversation like that in Train Case 1 concerning the train that just rolled into the station and leaves in 15 minutes. (See Fantl and McGrath (2002: 67-8)).

In both cases, you have the same evidence, however the pragmatic considerations are different. If epistemic justification is affected by pragmatic considerations, then having the same evidence would not mean that you have the same epistemic justification. However, it is not clear that we should think that epistemic justification is affected by pragmatic considerations. Arguments for pragmatic encroachment from cases like the two Train Cases typically rely on an assumption that the agent is only epistemically justified if she would be right to rely on this justification in acting – in deciding to take the train or not. However, as Brown argues, if this is right, then even very strong epistemic justification will

not be enough for the agent to rely on if the stakes are very high – certainty might be required in some cases, for example (Brown (2008: §6-7)). This shows not that evidential support is insufficient to justify belief in P, but rather that in some contexts epistemic justification is insufficient to justify action[42].

In summary, Evidentialism represents an important and uncontroversial element of epistemic rationality. Furthermore, the evidentialist principle defended here is even more uncontroversial than others that have been defended, and is not subject to the objections that are typically raised against those principles.

2.3. *The Enkratic Principle*

Third, the Enkratic Principle. The Enkratic Principle requires coherence between first and higher order attitudes. It is generally thought to be a requirement of rationality in both the epistemic and practical domains, but here we will only be concerned with the epistemic version of the enkratic principle, which concerns coherence between first and higher order epistemic attitudes. Read wide scope, the epistemic Enkratic Principle[43] says that one ought not believe that one ought to form an epistemic attitude (either believing P or refraining from believing P), without also having that attitude[44]. Reading `O' as `rationally required', we can state the principle as follows:

> *Enkratic Principle (positive):* O (BOBp   Bp)

---

[42] One might nevertheless be tempted by the idea that there is pragmatic encroachment on what we are justified in believing, on the grounds that some beliefs carry significant moral and practical consequences, for example racist or sexist generalisations (see Basu and Schroeder (2019); Bolinger (forthcoming)). To accommodate these worries, we need not jettison Evidentialism. Instead, we can note that the notion of 'support' involved in Evidentialism implies *adequate* support, and what counts as 'adequate support' can vary between contexts.

[43] Henceforth, by "the Enkratic Principle" I shall mean the epistemic Enkratic Principle (unless otherwise indicated).

[44] The wide scope reading, with the 'O' ranging over the whole conditional, has been the most popular formalisation of the principle. For further discussion of whether rational requirements should be read wide or narrow scope, see Broome (1999); Kolodny (2005); Titelbaum (2015a).

Enkratic Principle (negative): O (BO ~Bp ⊃ ~Bp)

If the Enkratic Principle is a requirement of epistemic rationality, then there are two options for complying with it. For the positive version, one may either believe P, or refrain from believing that one is rationally required to believe P. In other words, the Enkratic Principle prohibits the combination of attitudes that includes the belief that one is required to believe P, but not the belief P.

However, when one has misleading evidence about what rationality requires, this leads to the apparently conflicting requirements that are the focus of the discussion. By Evidentialism, agents who have misleading evidence sufficient to support a false belief about what rationality requires are required to believe what their misleading evidence supports. So, they are required to have a false belief about what rationality requires. Suppose this false belief is, "rationality requires believing P". By the Enkratic Principle they are required to either believe P, or give up the belief that they ought to believe P. However, because their false higher order beliefs are supported by their evidence, the option of giving up the higher order belief is not available to them in this case. This leaves them with only one option: forming the attitudes required by their false beliefs about what rationality requires – coming to believe P. But, since their evidence is misleading, this would conflict with what the true requirements of rationality require. In Chapter 3 I argue that we should solve the puzzle by rejecting the idea that the Enkratic Principle is a requirement of rationality. Nevertheless, it is worth noting that this strategy is not entirely without theoretical cost.

One attractive motivation for thinking that the Enkratic Principle is a requirement of rationality is that the belief combinations ruled out by the Enkratic Principle have an air of Moorean absurdity to them[45]. An agent who

---

[45] See Smithies (2012) on this point.

makes an assertion of the form "P, but I ought not believe that P" would seem to be open to the charge of irrationality, and intuitively the same might seem to apply to agents who hold beliefs of this form[46]. Nevertheless, Chapter 3 argues that giving up the Enkratic Principle as a requirement of rationality is the best way to solve the puzzle. The rest of this chapter, and the next, lays the groundwork for that argument by dismissing some attempts to dissolve the puzzle and deny that it is in need of a solution. Thus these two chapters combined establish that the Puzzle of Rational Requirement is in need of a solution.

3. Dilemmism

One attempt to dissolve the puzzle is to embrace the apparent conflict, and claim that epistemic rationality sometimes generates dilemmas. On this view – Dilemmism – both of the apparently conflicting requirements that agents in the puzzle are, apparently, subject to, express genuine requirements of rationality[47]. Dilemmism encourages us to embrace this as a feature of epistemic rationality, rather than a puzzle in need of a solution. This section argues, however, that it is theoretically preferable that epistemic rationality does not generate dilemmas. If this is right, then we should prefer strategies that avoid dilemmas, and see views that embrace the apparent conflict of the puzzle as, at best, a last resort.

One reason to rule out the possibility of dilemmas of epistemic rationality is that permitting them leaves us with very little to say about the cases at hand – cases of misleading evidence about what rationality requires. This is unfortunate, at least in so far as we expect epistemic rationality to generate judgments for all agents. While we might restrict the applicability of the requirements of rationality to some extent – for example, perhaps they do not apply to children or animals – any restriction should certainly include most adult humans with normal

---

[46] See Williamson on assertion-belief parallels (2002).
[47] See Alexander (2013); Christensen (2010); Hughes (2019).

reasoning capacities, even if they happen to have misleading evidence about what rationality requires. While some have suggested that such cases could be cases in which there is no rationally permissible attitude available to the agent, it is worth resisting this idea[48]. A significant part of the purpose of epistemic rationality would seem to be to offer normative judgments for what agents ought to believe. This purpose is best fulfilled if epistemic rationality is able to offer judgments that are applicable to all agents, regardless of their epistemic situation.

Another problem with Dilemmism is that it is incompatible with Standard Deontic Logic. Given Standard Deontic Logic, embracing dilemmas leads to contradictions and, in turn, triviality – entailing that everything is required. Suppose that rationality prohibits believing P. According to Dilemmism, agents with misleading evidence that rationality requires believing P are subject to both of the following requirements:

(a) O Bp

(b) O ¬Bp

They are subject to (a) because their misleading evidence indicates that they are required to believe P, and they are subject to (b) because the true requirements of rationality hold that they are prohibited from believing P. Although by itself, these requirements do not contradict each other – the negation of (a) is not (b), but rather ~O Bp – we can easily generate a contradiction from (a) and (b) given Standard Deontic Logic's principle of consistency[49]:

Principle of Consistency: Op   ¬O¬p

Given the following three further principles of Standard Deontic Logic[50], dilemmic requirements lead to triviality[51]:

Ought Entailment: Necessarily ((p → q) → (Op → Oq))

Agglomeration: Op & Oq → O (p & q)

Ex Falso Quodlibet: (p & ¬p) → q

When agents are subject to both of (a) and (b), these three principles lead to triviality in the following way[52]:

1. OBp (assump.)

2. O ¬ Bp (assump.)

3. O (Bp & ¬ Bp) (by Agglomeration)

4. O (Bp & ¬ Bp) → Oq (by Ought Entailment and Ex Falso Quodlibet)

5. Oq (from 3 and 4)

This generates a trivial theory of requirement, on which anything and everything can be required: for any q we can get the result that one ought to q. Triviality is worth avoiding, so we should steer clear of Dilemmism, at least while there are other live options. The following section discusses another set of attempts to dissolve the puzzle – views that use indexing strategies to separate apparently conflicting requirements.

---

[50] Others have pointed out that standard deontic logic has many imperfections (see Pigozzi et al. (2007); Sayre-McCord (1986)). Nevertheless, I will not dispute it here.

[51] See Horty (2003) for this argument.

[52] The two assumptions are the two requirements that Dilemmism holds agents in the puzzle to be subject to. By Agglomeration, these combine to give (3). Since this is a contradiction, Ought Entailment and Ex Falso Quodlibet together mean that Oq follows, for any q.

4. Indexing Strategies

Indexing strategies attempt to dissolve the puzzle by arguing that its apparently conflicting requirements of rationality can be indexed in some way that avoids conflict. This section argues, firstly, that these attempts to dissolve the puzzle are unsuccessful, and, secondly, identifies some crucial differences between these strategies and the solution to the puzzle that is defended in Chapter 3. Although my solution might be thought of as an indexing strategy, because it separates conflicting requirements by arguing that they pertain to distinct epistemic evaluations (requirement and appraisal), this section argues that my solution is importantly different from the indexing strategies discussed here.

Indexing strategies seek to separate items that appear to conflict so that they do not conflict. An example of this strategy is Lewis' proposal for understanding apparently inconsistent belief sets. He suggests that inconsistent propositions can be 'quarantined' to separate belief sets, thus limiting their potential problematic effects (1982: 435). This general strategy has been used to resolve apparent conflicts in the requirements of rationality in various ways. One such way begins from the claim that the ought of epistemic rationality admits of both a subjective and an objective sense, that is sometimes used in a way that is ambiguous between these two senses[53]. Asking what agents with misleading evidence about what rationality requires are 'rationally required' to believe might be thought to be an example of this ambiguous usage. If this is right, then we can dissolve the apparent conflict by distinguishing between an objective and a subjective sense of 'rationally required'. We can understand agents in the puzzle as objectively 'rationally required' to believe what the true requirements of rationality require, and subjectively 'rationally required' to believe what is required by the putative

---

[53] See, for example, Alston (1985); Feldman (1988a); Gibbard (2005); Gibbons (2013); Goldman (1986); Kvanvig (1984); Pollock (1979); Schroeder (2009); Unger (1986)).

requirements that their misleading evidence indicates they are subject to. On this view, in Logic 101 you are objectively 'rationally required' to avoid believing contradictions, and you are subjectively 'rationally required' to believe the contradiction that you see on the wall. Similarly, in Disagreement, you are objectively 'rationally required' to remain steadfast in your 0.8 credence that P, and you are subjectively 'rationally required' to conciliate in response to peer disagreement. According to this strategy, there is no further question about what you are rationally required to believe, and therefore there is no puzzle in need of solution. Instead, when you have misleading evidence that rationality requires believing P, there is a sense in which you ought to believe P, and a sense in which you ought not believe P.

Some have sought to further flesh out this strategy by arguing that `ought' is context-dependent, such that utterances and propositions involving the word `ought' require reference to a context in order to be meaningful (Björnsson & Finlay 2010; Pittard & Worsnip 2017; Worsnip, forthcoming). Contextualism about ought says that to ask what the agent `ought' to do is to ask a question that does not make sense until we specify the context of the ought in question. So, all that can be said of the puzzle cases is that there is one context in which the agent ought to have the attitudes demanded by the true requirements of rationality, and another in which she ought not. However, it is not obvious that this disambiguation helps. One difficulty is that what the objective sense of the ought of epistemic rationality demands depends on our first order theory of rationality. One of the commitments of the first order theory of rationality that generates the puzzle is that the Enkratic Principle is a requirement of rationality. If this is right, then the puzzle would arise even within the 'objective sense' of rational requirement. The puzzle arises from three commitments of 'objective' epistemic rationality (Externalism, Evidentialism, and the Enkratic Principle), so

distinguishing between objective and subjective senses of rationality does not help.

However, there are other ways to disambiguate senses of rational requirement. Worsnip (2018) distinguishes between evidential and coherence-based senses of `rationality requires'. The evidential sense of rationality requires you to have the doxastic attitudes supported by your evidence, while the coherence based sense requires you to have coherent mental attitudes. Similarly, Williamson (2017) argues that epistemic rationality is ambiguous between a `content-orientated' sense, according to which it is rational to believe p if and only if one's evidence supports p; and a `disposition-orientated' sense, according to which it is rational to believe p if and only if `in the same circumstances with the same evidence someone disposed to conform their beliefs to what their evidence supports would believe p'[54]. Disambiguating between these two senses of rationality means that when we ask what rationality requires of a particular agent, the answer depends on which sense of rationality we mean.

However, it is not clear that the disambiguations offered by Worsnip and Williamson can help us with the puzzle. This is because it is not clear that the two apparently conflicting requirements of the puzzle can be easily understood either as requirements of 'evidence' and 'coherence'[55], or as 'content-orientated' and 'disposition-orientated' senses of rationality. In the puzzle, the agent's evidence supports a false belief about what rationality requires. So, the evidential or content-orientated sense of rationality would require her to have this false belief. On this sense of what rationality requires, she is required to believe that

---

[54] In other work, Williamson identifies this sense with epistemic blamelessness (forthcoming). I argue against his view on this in Chapter 4.

[55] This represents an important difference between this puzzle, and puzzles arising from situations in which evidence is misleading about itself, which are more plausibly understood as apparent conflicts between requirements of evidence and coherence (see Christensen (2010a); Lasonen-Aarnio (2014); Weatherson (2019)).

rationality requires an attitude that it, in fact, prohibits. Meanwhile, the coherence sense of rationality would seem to require the agent to conform to the Enkratic Principle, and so form the first order belief recommended by her false belief about what rationality requires (and prohibited by the true requirements of rationality)[56]. For the disambiguation to be helpful in solving the puzzle, the evidential or content-orientated sense of rationality would need to line up with the true requirements of rationality in every case. However, as the puzzle is set up, the requirement to believe what one's evidence supports is not the only requirement of rationality. So, for the disambiguation to be helpful, we would need an argument that one's evidence could never support attitudes that are prohibited by rationality. This would ensure that evidential requirements did, in every case, line up with the true requirements of rationality[57]. However, to assume this would be too quick. It would be to deny that cases such as Logic 101 and Disagreement are possible[58]. So, one serious problem with indexing strategies for dissolving the puzzle is that there is no appropriate, non-question-begging distinction that we can appeal to in order to usefully distinguish the two senses or contexts of rational requirement.

Another problem with indexing strategies is that they leave unanswered the question of which of the two senses, or contexts, of rationality should govern the agent's response in any particular instance. Indexing strategies do not answer

---

[56] We might also think that Williamson's disposition-orientated sense of rationality would also require conformity with the Enkratic Principle, because someone disposed to conform her beliefs to what her evidence supports would adopt the beliefs that she believes to be rationally required. Lasonen-Aarnio makes some suggestions in this direction, noting that higher-order beliefs are "conspicuous reasons" that agents would be negatively evaluable for failing to respond to (see Lasonen-Aarnio (forthcoming)). However, establishing this would require more work. More will be said about Williamson's dispositional view of epistemic rationality and blamelessness in Chapter 4 (§2).

[57] Williamson, for example, defends this view (2002; 2017). However, it requires commitment to the idiosyncratic view that P is evidence iff P is knowledge.

[58] The following chapter discusses and dismisses more extensive arguments that such cases are impossible.

this question, because they claim that there is never a univocal answer to what rationality requires of a particular agent. This lack of an answer would seem to be a cost worth avoiding if possible, and we should thus prefer solutions that are able to answer this question[59]. Nevertheless, views that appeal to indexing strategies hold that this lack of answer is not a problem; it is rather the vain hope of always being able to find a univocal answer to the question of what agents are rationally required to believe that is the problem. However, this requires some argument. Worsnip's (2018) argument for this is that Evidentialism and the Enkratic Principle express distinct sets of normative requirements. On his view, the Enkratic Principle expresses what 'rationality' requires of us, since rationality (as he sees it) is to be associated with enkratic coherence. Evidentialism expresses what "evidence-responsiveness" requires of us (2018: 39), and this is to be thought of as entirely distinct from the demands of coherence. He compares the putative distinction between the demands of evidence and coherence to the distinction between the demands of morality and prudence. Although they both bear on actions, morality and prudence are not demands of the same kind. On Worsnip's view, Evidentialism and the Enkratic Principle are similar: while they both bear on beliefs, they do not make demands of the same kind.

The problem with this argument is that it is not as clear as Worsnip needs it to be that the demands of evidence and coherence are distinct in the way that morality and prudence are. The puzzle is an instance of apparent int*ra*-domain conflict, but Worsnip's solution targets only int*er*-domain conflict. The significance of the puzzle is that it concerns requirements within the same normative domain: unlike morality and prudence, Evidentalism and the Enkratic Principle seem to make the same kind of demands of us – demands of epistemic rationality. Of

---

[59] Kvanvig also makes this point in discussing the 'egocentric predicament' (2014: 48).

course, adjudicating this dispute requires some way to individuate normative domains. One way to individuate normative domains is by reference to the kind of force that their requirements have. The requirements of morality have a moral force – one is required to comply with them for moral reasons, such that failing to do so would be morally wrong, or morally blameworthy. Prudential requirements do not have this force. One should comply with the requirements of prudence for prudential reasons, such as that to fail to do so would be against one's best interest. However, Evidentialism and the Enkratic Principle both seem to have epistemic force. For both, it seems that one should comply with them for epistemic reasons, such as that failing to do so would be epistemically irrational[60]. Indexing strategies, such as Worsnip's, do not give us reason to think otherwise.

In the remaining chapters I defend the view that we should solve the puzzle by distinguishing between requirements and evaluations of appraisal. While this strategy also involves distinguishing two notions that together generate contribute the puzzle, it is different from the indexing strategies discussed in this section. Importantly, the two notions I appeal to – requirement and appraisal – are not different senses of the same notion (such as 'rationality', or 'ought'), but rather two distinct kinds of epistemic evaluation that are simultaneously at work in our evaluations of the rationality of an agent's beliefs, giving the appearance of intra-domain conflict. In contrast to the indexing strategies discussed here, the view defended in Chapter 3 is able to univocally state what the agent must do in order to be evaluated as having done what rationality requires. This allows it to satisfactorily solve the puzzle.

---

[60] This sidesteps a vast literature on whether we have reason to be rational. There is much more to be said about whether and why one should comply with the demands of epistemic rationality, but this is not the focus of the discussion (see Broome (1999; 2013); Kolodny (2005)).

5. Summary

This chapter has presented the Puzzle of Rational Requirement with which this thesis is concerned. The puzzle involves apparent intra-domain conflict arising from the possibility of rational mistakes about what rationality requires. This chapter has shown how the puzzle arises and distinguished it from similar but importantly different puzzles. It has identified and discussed the motivations for the three claims that one must be committed to in order for the puzzle to arise: Externalism, Evidentialism, and the Enkratic Principle. Finally, it has dismissed two attempts to avoid the problem that apparent intra-domain conflict appears to present for our theorising: Dilemmism, and indexing strategies.

Before arguing for my preferred solution to the puzzle in Chapter 3, it is necessary to discuss and reject one further strategy for dismissing the puzzle – The Impossibility Thesis. This is the focus of the following chapter. The Impossibility Thesis implies an asymmetrical account of mistakes: mistakes about what rationality requires are different from other kinds of mistakes, such that to make them is itself a failure of rationality. On this view, agents can *never* be rational to believe falsely about what rationality requires. If this were true, then the apparent intra-domain conflict between requirements of rationality that generates the puzzle would never arise, because agents would never be rational to hold a false belief about what rationality requires. The following chapter considers and rejects arguments for this view.

Chapter 2

No Way Out: Rejecting the Impossibility Thesis

This chapter considers and rejects arguments for the Impossibility Thesis. The Impossibility Thesis says that rational false beliefs about what rationality requires are impossible. If the Impossibility Thesis were true, then this would offer a way to deny that the Puzzle of Rational Requirement could arise. This would dissolve the puzzle, thus removing the need for a solution. However, as this chapter argues, the Impossibility Thesis is false.

Section 1 states the Impossibility Thesis, and shows how it would dissolve the puzzle, if it were true. The remaining sections present and argue against three strategies for defending the Impossibility Thesis. Section 2 concerns a defence of the Impossibility Thesis as an instance of a more general claim about the impossibility of rational false belief of any kind. Section 3 concerns an appeal to a claim that mistakes about what rationality requires constitute a kind of incompetence. Section 4 concerns an appeal to our putative justificatory assets for truths about what rationality requires. Section 5 briefly dismisses an argument for the Impossibility Thesis from the Enkratic Principle, noting that this argument falsely assumes that giving up the Enkratic Principle is a less plausible strategy than accepting the Impossibility Thesis. The following chapter argues that this assumption is false: giving up the Enkratic Principle is in fact the best strategy for resolving the puzzle.

1. The Impossibility Thesis

As outlined in the previous chapter, when agents have rational false beliefs about what rationality requires, then rationality seems to issue conflicting

requirements, generating a puzzle. This puzzle depends on commitment to the following three claims:

> *Externalism:* What rationality requires is completely determined by the facts about what rationality requires ("the rational requirements")[61].

> *Evidentialism*: S(e)p → O(S(Bp)). If S has evidence that supports P, then S is rationally required to believe that P.

> *Enkratic Principle*: O (BOΦ → Φ). Rationally requires that agents do as they believe they ought to do.

One way to avoid this apparent conflict would be to deny the possibility of rational mistakes about rationality. This would be to endorse the Impossibility Thesis.

> *Impossibility Thesis*[62]: Rational mistakes about what rationality requires are impossible.

If the Impossibility Thesis were true, then commitment to Externalism, Evidentialism, and the Enkratic Principle together would not generate the Puzzle of Rational Requirement outlined in the previous chapter. However, as this chapter argues, we should reject the Impossibility Thesis; it is implausible and there is no good argument for it. The following sections present and reject three strategies for defending the Impossibility Thesis – via a more general claim about

---

[61] One way to put this is that the requirements of rationality are `indefeasible' – that is, they remain binding in all possible situations. See Bradley (2019) and Ichikawa and Jarvis (2013) for examples of this way of thinking about requirements of rationality.

[62] The Impossibility Thesis has various defenders. For example, "mistaken beliefs about what rationality requires of you are themselves irrational beliefs." (Littlejohn (2015: 270)); "mismatches between what seems true and what is true constitute rational failings in the logical domain but not the empirical domain." (Smithies (2015: 2778)); "mistakes *about* rationality are mistakes *of* rationality" (Titelbaum (2015: 253)). Additionally, parallel claims in the spirit of the Impossibility Thesis have been made in the moral domain (see Harman (2011)).

the impossibility of false beliefs (§2), by appeal to an argument from the nature of rational competence (§3), by appeal to a set of putative justificatory assets for the truth about what rationality requires (§4), and by appeal to the Enkratic Principle (§5).

2. No Rational False Beliefs

One way to defend the Impossibility Thesis is via a defence of the more general claim that there are no rational false beliefs at all[63]. On this view, rational mistakes about what rationality requires are impossible, because rational mistakes about any topic are impossible. However, this general claim about the impossibility of rational false belief faces the difficulty of explaining away the apparent rationality of various cases of false belief that we are usually inclined to evaluate in some positive sense: victims of Gettier cases, who deduce justified true beliefs from apparently justified false beliefs; agents experiencing perceptual illusions; brains in vats, and recipients of false testimony. Those who deny the possibility of rational false belief have attempted to explain away these cases of apparently rational false belief by appeal to some further dimension of epistemic evaluation, for example, excusability (Littlejohn (forthcoming); Sutton (2005; 2007); Williamson (forthcoming)). One problem with this is that it risks running together an important normative distinction between genuinely rational or justified belief and excusable false belief that fails to meet the conditions of rational of justified belief (see Brown (2018); Cohen and Comesaña (2013); Kelp (2016)). If rational or justified false belief is impossible, then all positively evaluable false belief needs to be captured using the notion of excusability. This

---

[63] Those who deny the possibility of rational and/or justified false belief of any kind include Sutton (2007); Williamson (2017)).

means running together cases that would seem to demand very different evaluations.

There are important differences between apparently rational false belief, and cases in which we might want to excuse agents for poor epistemic conduct. For example, consider victims of cults might who are manipulated into reasoning in a way that will not lead to epistemic goods, for example placing trust in unreliable authority figures. Although they form beliefs in a way that has little epistemic merit, their vulnerable situation means that they might be excused for doing so. In contrast, brains in vats are using methods that would lead to epistemic goods if they were not envatted, and various other cases of false belief can be attributed to bad luck rather than bad methods (for example, receiving false testimony from usually reliable source). The notion of rationality or justification – as distinct from epistemic excusability – allows us to mark this difference, but only if rational or justified false beliefs are possible. We can say that brains in vats are rational because they employ generally epistemically good methods of reasoning that in this case do not result in knowledge, while the victims of cults are not rational since they employ epistemically bad methods that would never result in knowledge, although we might think that their unfortunate situation excuses them[64]. This distinction is not available if rational false beliefs are impossible, and that is unfortunate. This gives us one reason to reject the general claim that rational false beliefs are impossible, and with it the defence of the Impossibility Thesis that rests on it. The following sections examine and reject some further strategies for defending the Impossibility Thesis, beginning with the Incompetence Argument.

---

[64] Much more will be said in later chapters about the conditions under which agents can be excused for epistemic failings.

3. The Incompetence Argument[65]

Littlejohn (2015) offers another way to argue for the Impossibility Thesis, namely by appeal to the claim that to believe falsely about what rationality requires is to manifest incompetence with respect to rationality in a way that precludes those beliefs from being rational. Littlejohn argues that false beliefs about what rationality requires are irrational because they exhibit incompetence in dealing with one's (epistemic) reasons[66]. We can reconstruct his argument, which I will call the Incompetence Argument, in the following way:

Incompetence Argument

> P1. Rationality is competence in handling the reasons that apply to you.
>
> P2. Competence in handling reasons that apply to you requires understanding what is required when reasons apply to you.
>
> P3. Having false beliefs about rational requirements involves failing to understand what is rationally required of you in your particular situation.
>
> P4. If you fail to understand what is rationally required of you in your particular situation, then you manifest an incompetence with respect to rationality.

---

[65] The material in this section is a revised version of the arguments in C. Field (2019a).

[66] As he puts it, "rationality requires an understanding of what's required when reasons apply to you … [the Impossibility Thesis] isn't true because we all happen to have evidence for the right list of rational requirements; rather, it's true because the grounds for saying that someone's attitudes are irrational is that those attitudes reveal a kind of incompetence with respect to handling reasons and their demands. As it happens, mistaken beliefs about what rationality requires will manifest that kind of incompetence" (Littlejohn (2015: 14)). Harman, following Arpaly (2002), makes a similar argument in the moral case. She sees false belief about what is morally required as itself a moral failure, because it is a failure to care adequately about what is morally important (2011: 459).

Conclusion: False beliefs about what rationality requires are not rational.

If this argument is correct, then mistakes about what rationality requires differ from mistakes about other topics because false beliefs about rationality manifest an incompetence with respect to rationality, and this incompetence means that those beliefs cannot be rational. The problem with the argument is that there is no way of interpreting 'fail to understand', as it appears in Premises 3 and 4, in a way that makes the fourth premise likely to be true. There are two ways of reading 'fail to understand', as 'lack true beliefs about', and as 'hold mistaken beliefs about'. As the following paragraphs outline, both of these lead to problems that should lead us to reject Premise 4 of the argument. First, the reading of 'fail to understand' as 'lack true beliefs about':

Lack Reading: to fail to understand P is to lack true beliefs about P.

On the Lack reading, Premise 4 says that anyone lacking true beliefs about what rationality requires is manifesting rational incompetence. However, for many candidate rational requirements it seems to be the case that agents can fulfil the requirements without holding any beliefs at all about what the requirements are. It is possible to fulfil the non-contradiction requirement, for example, by refraining from believing contradictions, and it is possible to refrain from believing contradictions while also taking no view about what rationality requires. We can see this by considering how, on the Lack reading, the Incompetence Argument evaluates particular agents. For example, consider the following agents:

Innocent Agents: Lack some (and perhaps all) true beliefs about what rationality requires. They take no view about what is

required by rationality. Despite this, their beliefs are completely

in line with what is in fact required by rationality.

According to the Lack reading, innocent agents fail to understand what is rationally required of them, and so would count as incompetent with respect to rationality. There are at least two reasons to resist this result. Firstly, it is far too demanding. Given the widespread disagreement over what the rational requirements are – whether, for example, one is rationally required to believe lottery propositions, conciliate in the face of peer disagreement, or believe the logical consequences of one's beliefs – it seems that no-one, except for perhaps a few enlightened epistemologists, will count as rationally competent. Secondly, on the Lack reading, Premise 4 introduces further requirements on rationality, over and above the true requirements that Externalism commits us to. In order to count as competent with respect to rationality, agents must not only obey the true requirements, but they must also have second order beliefs about what they are rationally required to believe. This is an awkward addition, reminiscent of Carroll's tortoise[67], meaning that it is not enough for agents to obey the requirements at the first order, they are also required to have second order beliefs about what is required at the first order. We might wonder how far up this demanding requirement goes. Must agents also have correct beliefs about their second order beliefs – that is, must they hold the correct beliefs about what rational competence requires (that it requires correct belief about what rationality requires at the first order)? If so, this seems to generate a needlessly complex picture of the requirements of rationality. This awkward result might push one towards the second reading of 'fail to understand' – 'hold mistaken beliefs about':

---

[67] See Carroll (1895).

> *Mistake Reading*: To fail to understand P is to hold mistaken beliefs about P.

This allows innocent agents to count as rational, but accuses those who explicitly believe falsehoods about what rationality requires of incompetence with respect to rationality. This is an improvement on the Lack reading since it says the right thing about innocent agents – they count as rational in virtue of believing in line with the requirements at the first order and holding no mistaken beliefs at the second order. However, it says the wrong thing about other kinds of agents. For example, consider the following agents:

> *Misguided Akratic Agents:* Hold false second order beliefs about rationality but fail to believe in line with these beliefs at the first order, and so believe in accordance with rationality.

The Mistake reading of 'fail to understand' means that if the Incompetence Argument is correct, then Misguided Akratic agents are manifesting incompetence with respect to rationality, in virtue of their holding mistaken beliefs about rationality. However, it is not immediately clear why misguided akratic agents should be treated differently to innocent agents. Like innocent agents, misguided akratic agents obey the requirements of rationality at the first order, it is only their higher order beliefs that are problematic. The Mistake reading means that innocent and misguided akratic agents would be treated differently, if the Incompetence Argument is correct. Whereas innocent agents are forgiven for their lack of true beliefs, misguided akratic agents are not, even though both agents obey the requirements of rationality at the first order. Some explanation is needed for this difference in treatment. One possible explanation is that the mismatch between higher and first order beliefs in cases of akrasia exhibits an obvious kind of irrationality, and is justification enough for difference in treatment. However, this response introduces something of a complication

into what would otherwise be a simple and elegant theory of rational requirements of rationality: contrary to what we might have thought, epistemic rationality requires more than complying with the requirements of rationality. The Mistake reading assesses the misguided akratic agent as irrational, despite her complying with all the requirements at the first order. This implies that first order compliance with the requirements is not enough. To be rational, agents must also comply with the further requirement to not have false beliefs about what the requirements require. This means that rationality involves both first and higher order requirements. Since it is possible to fulfil the first order requirements while holding false beliefs at the higher order, it is not clear why this extra requirement to avoid false beliefs is necessary. Further support for the implausibility of the idea that rationally competent agents must avoid beliefs about what is rationally required of them can be drawn from the literature on skill. It is well documented that skilled agents – that is, agents competent at various tasks – do not always have true beliefs, in fact they often have false beliefs, about what they are required to do in order to perform the task successfully[68].

In summary, the Incompetence Argument is caught between a rock and a hard place, and should be rejected. On the one hand, taking the Lack reading of 'fail to understand' and requiring agents to have beliefs, true ones, about what rationality requires of them seems too strong. The Incompetence Argument needs to make an exception for innocent agents to avoid absurdity. On the other hand, taking the Mistake reading in order to make this exception invites the question as to why misguided akratic agents should not also be granted an exception – on the grounds that their first order beliefs obey the requirements of rationality. Neither of these options is attractive, and so we should reject the

---

[68] For some support for this see Reed et al. (2010), Brownstein (2014), and Montero (2016).

Incompetence Argument as a strategy for supporting the Impossibility Thesis. The following section presents and rejects an attempt to defend the Impossibility Thesis by appeal to a putative set of justificatory assets for the truth about what rationality requires.

4. Justificatory Assets

Another way to defend the Impossibility Thesis is by appeal to the claim that, in fact, all agents in all possible situations are justified in believing the truth about what rationality requires:

> *Justificatory Assets[69]:* All agents, in all possible situations have justification for true propositions about what rationality requires.

If Justificatory Assets were true, then this would offer a way to defend the Impossibility Thesis. However, this defence relies on the truth of the following three further claims:

> *Access:* Agents always have access to justification for true beliefs about what rationality requires.
>
> *Indefeasibility:* Our justification for the truth about what rationality requires is indefeasible.
>
> *Sufficiency:* Our justification for the truth about what rationality requires is sufficient to rule out the possibility of rational false belief about what rationality requires.

As I will argue, each of these claims is contentious, and the defence of the Impossibility Thesis from Justificatory Assets requires all three – the failure of any of the three would be decisive against the strategy. The rest of this section discusses each in turn, arguing that each should be rejected.

---

[69] Defenders of this claim include Ichikawa and Jarvis (2013); Smithies (2012, 2015); Titelbaum (2015b).

According to Access, agents always have access to justification for the truth about what rationality requires.

> *Access:* Agents always have access to justification for true beliefs about what rationality requires.

Access seems prima facie implausible in light of apparent cases of rational false belief about what rationality requires. Consider again Logic 101 (Ch. 1, p. 12). In this case, you acquire misleading evidence for the false claim that rationality sometimes requires contradictory belief through expert testimony and by thinking about convincing arguments. Importantly, in Logic 101 you are not in a position to gather any evidence that would defeat the false view, because you are a beginner student who currently lacks the capacity to marshal the complex arguments that would defeat the dialetheist view of what rationality requires. If Access were true then, despite appearances, Logic 101 would be a case in which you have access to justification for the true belief that rationality prohibits inconsistent belief.

Access is at least prima facie implausible. As the rest of this subsection outlines, it is also not well motivated. There are two principal ways that Access might be motivated: as a consequence of Bayesian epistemology, or by appeal to the putative unrestrictedness of a priori reasoning.

First, Access might be seen as a consequence of Bayesian commitments to logical omniscience. Logical omniscience says that for all agents in all situations, their evidence – whatever it is – supports all logical truths to credence 1. On the assumption that there is an intimate connection between logical truths and requirements of rationality, this might offer a way of establishing that we have access to justification for truths about what rationality requires, since our

evidence always supports them to a maximal degree. However, this possibility is not warranted; there is no straightforward relationship between the truths of logic and truths about what rationality requires. For example, while (P & (P → Q)) → Q) is a logical truth, it is not a demand of rationality that ((Bp & B (P → Q)) → Bq. There will be some cases in which what rationality demand is that one stop believing either P or (P→ Q)[70]. This suggests that we should not assume that the kind of access Bayesianism putatively gives us to justification for the truths of logic necessarily translates into justification to believe truths about what rationality requires.

Furthermore, the view that we are justified to a maximal degree in all logical truths is often, if not usually, taken to be a bug rather than a feature of Bayesian epistemology. For example, logical omniscience seems straightforwardly false if we are considering ordinary, non-ideal agents, and it implies some counterintuitive consequences for rational support – any evidence set whatsoever supports credence 1 in all logical truths, and this includes misleading evidence sets which directly contradict these logical truths, as in Logic 101. Taking Bayesian commitments as a guide to justification for ordinary, non-ideal agents introduces the significant challenge of working out how logical omniscience applies to ordinary agents, and how to make intelligible the claim that we have access to these justificatory assets despite their often being opaque to us. Unless this challenge is met, a commitment to Bayesianism is not enough, on its own, to establish Access[71].

---

[70] See Besson (2010) and Harman (1986) on this point. Articulating the sense in which logic is normative for belief has typically taken the form of a search for the correct 'bridge principles' between truths of logic and normative epistemic principles. Macfarlane (2004) offers 36 different ones to choose from, each with their own theoretical costs and benefits.

[71] One approach is to emphasise a distinction between formal systems associated with rationality and good epistemic practices (Cohen (1981); Harman (1986); Russell (2017)), another, very different, approach is offered by Stalnaker (1991). Some have nevertheless attempted the project

A second way to motivate Access is by appeal to the putative unrestrictedness of a priori reasoning. We might think that when justification is a priori, agents need only reflect on the relevant propositions in order to believe them with justification. This means that so long as agents have the ability to reflect, they have access to propositional justification for the truth about what rationality requires. This means that unlike for empirical reasoning, agents' particular situations cannot introduce restrictions on what they can become justified in believing a priori. While there are physical limitations on, for example, where I can transport my body – and thus the perceptual equipment I need to gather empirical evidence – there are no such limitations on what I can think about. So, there is a sense in which all rational agents could come to believe the truth about what rationality requires because these truths are accessible by a priori reasoning alone[72].

However, we might question the putative unrestrictedness of a priori reasoning, and worry that the sense in which agents in some situations 'could' come to believe, with justification, the truth about what rationality requires is too weak to establish the Access claim. Although there is perhaps a sense in which agents in epistemic situations such as Logic 101 *could* come to believe, with justification, the truth about what rationality requires – if they were to reflect on the relevant justifying propositions in the right way – this would nevertheless be very difficult. It is worth comparing their situation with respect to the truth about what rationality requires with analogous empirical justification. Suppose that there is a very unusual large blue rock on top of the Eiger. The Eiger is a difficult

---

of defending this unpopular logical omniscience requirement (see Smithies (2015), Wedgwood (2017)).

[72] Wedgwood (2017), for example, motivates Access in this way. Although he does concede that this is restricted in a such a way that respects human limitations, even with this restriction this implies that we have access to significantly more a priori justification than empirical justification.

climb, but not impossible if one is a skilled mountaineer. Suppose that you are not a particularly skilled mountaineer. Although there is a sense in which you could climb the Eiger, and become justified in believing that there is a blue rock at its summit, it is not clear that we would consider you to have access to justification for the truth about rocks at the summit of the Eiger. In order to make use of this justification you would need to undertake some relatively demanding and specific further tasks – either becoming a much better mountaineer, or undertaking some research about the summit of the Eiger. If we do not think that this situation gives agents access to empirical justification for truths about blue rocks, then we should not think that the similarly structured situations involving a priori reasoning, such as Logic 101, give agents access to truths about what rationality requires.

One way to address this worry would be to deny that what we have epistemic access to is constrained by our limitations. We might take this to hold for either or both of a priori and empirical justification, perhaps following Christensen[73] and Smithies[74] in understanding our justification to be defined by what the ideally rational agent would be justified in believing. This would mean that we could consider ourselves to have access to justification for the truth about what rationality requires so long as the ideal agent is. Of course, this is contentious and would require further argument. More importantly for this chapter, even if this connection between justification and the epistemic ideal could be used to establish the truth of Access, in order to establish the Impossibility Thesis it would also need to be the case that no further justification could defeat this ideal

---

[73] For example, ''not all evaluation need be circumscribed by the abilities of the evaluated. In epistemology, as in various other arenas, we need not grade on effort'' (2004: 162).

[74] For example, "the propositions that one has justification to believe are just those propositions that one would believe if one were to be idealized in relevant respects", and "the limits on one's doxastic capacities do not constrain the epistemic ideal, but only the extent to which one is capable of approximating towards the ideal" (2012: 8).

justification (Indefeasibility), and that this justification was sufficient to rule out the possibility of rational false belief about what rationality requires (Sufficiency). As the following subsections argue, there are reasons to reject both Indefeasibility and Sufficiency. The following subsection addresses Indefeasibility and argues that there are good reasons to reject it.

4.2 Indefeasibility

The Indefeasibility claim says that our justificatory assets for the truth about what rationality requires are indefeasible.

> *Indefeasibility*[75]*:* Our justification for the truth about what rationality requires is indefeasible.

To make the argument from Justificatory Assets to the Impossibility Thesis, Indefeasibility needs to be true. If it is false, then rational mistakes about what rationality requires would be possible in cases in which our Justificatory Assets were defeated. Indefeasibility rules out this possibility and is thus essential for making the argument from Justificatory Assets to the Impossibility Thesis. However, Indefeasibility faces two problems: it is ill-motivated and it gives an unsatisfactory explanation of cases of misleading evidence about what rationality requires.

4.2.1 Ill-motivated

There are two principal ways to motivate the idea that our justification for the truth about what rationality requires is indefeasible: by appeal to its being a priori, and by appeal to its being propositional. A priori justification is, by definition, justified independent of experience. Some have thought that that any

---

[75] Indefeasibility about a priori justification is defended by H. Field ((2000); Ichikawa & Jarvis (2013); Kitcher (1980); Smithies (2015); Titelbaum (2015). In so far as the truths about what rationality requires are a priori, these views would also be committed to indefeasibility about justification for truths about what rationality requires.

misleading evidence that one could acquire for a priori claims must be experiential[76], licensing the following argument for the Indefeasibility claim:

> 1. If P is justified a priori, then it is justified independent of any experience.
>
> 2. If justification is independent of experience, then it cannot be defeated by experience.
>
> 3. The misleading evidence in the cases is derived from experience.
>
> 4. So, misleading evidence cannot defeat our propositional justification for the truth about what rationality requires (Indefeasibility).

Both Premises 2 and 3 are contentious. Premise 2 is made implausible by cases such as Logic 101. While students may be justified in believing that rationality prohibits contradictions before they take the class and consider persuasive arguments for dialetheism, it is implausible that their justification remains undefeated after they have the experience of hearing the professor's testimony. The fact that it is an experience that provides the source of this defeating evidence does not seem to be relevant to whether it is defeated[77].

Premise 3 is contentious because it must rely on a non-standard understanding of `experience' if it is to exclude cases such as Logic 101. The students acquire the

---

[76] For example, see Smithies' claim, relating specifically to a priori justification for logic: "apriori justification for beliefs about logic has its source in logical facts, rather than psychological facts about experience, reasoning, or understanding." (2015: 2270). Since misleading evidence about logic is not a logical fact, its source must be rather a 'psychological fact about experience, reasoning, or understanding'. This view dates back to Frege (see particularly his essay, "Logic" (1897). Ichikawa and Jarvis (2013) defend a more general version of this view for all a priori truths.

[77] Defenders of Indefeasibility typically explain away apparent cases of defeat such as this by appeal to a distinction between defeat and some weaker defeat-like status. §4.2.2 argues that we should resist this strategy.

misleading evidence by thinking about logic and philosophical arguments: through a priori reasoning. However, those who make this argument claim that the methods by which misleading evidence is acquired are not a priori, but rather a posteriori experiences of thinking through arguments[78]. On this view, only true a priori beliefs can be justified. False beliefs formed by abstract thinking and reasoning are mere experiences; any misleading evidence they might produce is merely empirical and so cannot defeat propositional a priori justification. On this, view when abstract thinking results in truth, then one is doxastically a priori justified. When abstract thinking goes wrong, that thinking was 'empirical', and one is not justified in its conclusions. This involves taking a disjunctive view about whether an instance of abstract thinking constitutes a priori justification. Intuitively, this seems like the wrong way to determine whether justification is a priori. This disjunctive view permits the implausible possibility that instances of abstract thinking that appear to be of the same kind in fact have very different epistemic statuses depending on their truth value[79].

The other principal way to motivate Indefeasibility is by appeal to its being propositional, via the following argument:

> 1. Propositional justification for P supports the truth of P.

> 2. Defeating propositional justification for P requires either undermining that support, or providing countervailing support for not-P.

---

[78] This view is held by, for example, H. Field (2000); Ichikawa and Jarvis (2013); Kitcher (1980); Smithies (2015)).

[79] For some further arguments in support of the intuitive view that justification is a priori when it involves a priori thinking and reasoning, see Casullo (1988); Jeshion (2000); Summerfield (1991). Nevertheless, evaluating this fully would require delving into a vast literature, and deciding on the appropriate definition of 'experience' and 'a priori', which is beyond the scope of this discussion. For some related arguments against disjunctivism, particularly disjunctivism about perception, see Sturgeon (1998).

3. The misleading evidence involved in the cases cannot provide such support.

4. Misleading evidence cannot defeat our propositional justification for the truth about what rationality requires (Indefeasibility).

If this is right, then it would establish that no misleading evidence about what rationality requires could defeat the agent's propositional justification for the truth about what rationality requires. However, (3) is false. The misleading evidence that you acquire in Logic 101 simultaneously justifies your false belief about what rationality requires and defeats any a priori justification you might have had for the true view. On a standard reading of the relationship between doxastic and propositional justification, doxastic justification implies propositional justification. For example, according to the `orthodox view' identified by Turri, P is propositionally justified for S if and only if S has justifiers (reasons or evidence) J sufficient to justify P (2010: 314)[80]. Standardly, S is doxastically justified in believing P when the following three conditions hold:

a) P is propositionally justified for S

b) S believes P

c) S believes P on the basis of J

The agents in the misleading evidence cases seem to meet all three of these conditions. In Logic 101, you pay attention, and come to believe that rationality sometimes requires contradictory belief on the basis of propositions that are offered to you by your professor, and that you arrive at by thinking through the

---

[80] For further defenders of this view, see Kvanvig (2003: §B1); Pollock (1986: 36-7); Swain (1979: 25)).

arguments. According to standard accounts of defeat, justification can be defeated by both undermining and rebutting evidence, regardless of the strength of the justification (Brown 2018; Pollock 1986). So, any prior justification you may have had for the truth about what rationality requires could be defeated by these propositions. These propositions do support the false conclusion that rationality sometimes requires contradictory belief, albeit not to a maximal degree, so they would seem to provide rebutting defeat of any justification you might be thought to have for the truth[81]. Examples of propositions that plausibly constitute your evidence include: "Dialetheism offers an elegant solution to the Liar paradox."; "Dialetheism allows us to preserve a naive truth schema."; "Logical validity is a plausible guide to what epistemic rationality requires."; "There are some compelling reasons to doubt that ex falso quodlibet ('anything follows from a contradiction') is a valid rule of inference."; "Classical logic is not the only possible system of logic.". All these propositions are plausibly true, and in absence of evidence to the contrary, these propositions would seem to serve as justifiers for the further claim that rationality sometimes requires contradictory belief[82]. If this is correct, then given the `orthodox view' of the relationship between doxastic and propositional justification, in Logic 101 you would seem to be both doxastically and propositionally justified in the claim that rationality sometimes requires contradictory belief – you believe it, you believe it on the basis of justifiers, and your justifiers do support the truth of the claim. So, not only are you justified in believing the false view about what rationality requires, but this justification also defeats any other justification you can be presumed to have for the true view of what rationality requires.

---

[81] This is consistent with views that take us to have defeasible entitlement to believe the truth about what rationality requires (see Wright (2004a; 2004b; 2014)).

[82] For a more fleshed out version of the argument, see Priest (1985; 2005; 2006).

Furthermore, other students in Logic 101 who are less diligent in studying the arguments and paying attention to the professor would also seem to be propositionally justified in the false belief about rationality. These less-than-diligent students have the same misleading evidence, but fail to make use of it[83]. They are not doxastically justified in the false claim, because they do not believe it; but they are propositionally justified, since they have the same evidence available to them as the more diligent students. If this is right, then Premise 3 of the argument for Indefeasibility is false, since the cases can provide propositional justification for the negation of the truth about what rationality requires, and so would seem to provide rebutting defeat of our putative propositional justification for the truth about what rationality requires. If this is right, then it is not true that our putative justificatory assets for the truth about what rationality requires are indefeasible – cases such as Logic 101 are counterexamples. So, neither the fact that the justificatory assets are a priori, nor the fact that they are propositional provides adequate motivation for the indefeasibility claim. Additionally, there is a further problem – a commitment to the Indefeasibility claim rules out a natural explanation of the misleading evidence cases at hand, without permitting an adequate alternative explanation.

4.2.2 Inadequate explanation of the cases

Commitment to Indefeasibility means that we are left without an adequate account of cases of misleading evidence about what rationality requires. A natural explanation of cases of misleading evidence about what rationality requires is that the misleading evidence defeats any prior justification the agent

---

[83] Depending on what we think is required for them to 'have' evidence, we might imagine that they believe the propositions, but fail to draw the inference, or fail to believe them even though they easily could (see Ch. 1, §2.2 for discussion).

might have had to believe the truth about what rationality requires[84]. Since this explanation is not available to defenders of Indefeasibility, they an alternative explanation of cases of misleading evidence about what rationality requires. However, as the rest of this subsection argues, the explanations compatible with Indefeasibility do not hold up. Given this, we should reject Indefeasibility.

Defenders of Indefeasibility typically explain away apparent cases of defeat such as this by appeal to a distinction between defeat and some weaker defeat-like status – the `disabling' of justification (Smithies 2015), or the agent's being `rationally compromised' (Ichikawa & Jarvis 2013). Views that distinguish between defeat and a weaker defeat-like phenomenon (henceforth `disabling'), say that when S's justification for P is disabled she is unable to use that justification in her reasoning, she is unable to form a justified belief that P on the basis of that justification. Nevertheless, S's justification for P is not defeated. The problem with this is that the options it permits for characterising the evidential situation of agents in cases of misleading evidence about what rationality requires are implausible.

One immediate point worth noting is that this appeal to the notion of disabling offers a way of making the possibility of rational mistakes about rationality compatible with Indefeasibility, thus making the route from Justificatory Assets to the Impossibility Thesis even more difficult. If our justification can be disabled, then this can offer a sense in which we might be rational in failing to make use of it, despite having access to it, and despite its being indefeasible. If, in Logic 101, you fail to make use of your a priori propositional justification, and instead trust the testimony of the professor and accept the conclusions of the arguments you study in class, then you will end up rationally believing the false view that

---

[84] See Wright (2004b, 2014) on defeasible entitlements, as well as Bonjour on a priori justification (1998).

contradictory belief in sometimes rationally required. So, the account defenders of Indefeasibility typically give of misleading evidence cases serve to permit, rather than rule out, the possibility of rational mistakes about rationality. Supporters of Indefeasibility might nevertheless protest that failing to make use of disabled justification is not a way to be rational in the fullest or most ideal sense of the term[85]. Nevertheless, the explanation offered by `disabling' of cases of misleading evidence about what rationality requires is unsatisfactory.

To show why the indefeasibility view has a problem explaining cases of misleading evidence about what rationality requires, recall that according to this view of a priori justification, there are only two ways that propositional justification can be disabled – either by your making an error due to cognitive incapacity akin to Chomskian 'performance limitations', or by your having empirical evidence that you currently lack the cognitive capacity to make use of propositional justification. A pedestrian example of genuine cognitive incapacity is limitation in cognitive processing capacity. Some logical truths are too complicated to be deduced by ordinary human agents, even though they are entailed by their evidence, and follow from things they already believe[86]. Another example is distraction by one's prior beliefs and biases, which frequently results in errors in relatively simple probabilistic reasoning[87]. Biases can cause performance error by distracting agents from a priori facts about probability, even when they know these facts to be true and would assert them on reflection[88].

---

[85] For example, as Ichikawa and Jarvis (2013) put it, failure to make use of one's justification constitutes a compromise of at least one sense of rationality - `p-rationality', that one manifests by making the transitions in reasoning permitted by the rational relations that hold between a priori propositions.

[86] Assuming a Bayesian framework, according to which all logical truths are entailed by any evidence set.

[87] Such as such as inferring that a conjunction is more likely than one of its conjuncts alone, as in the well-known Linda the bank teller case.

[88] See Kahneman (2011); Kahneman et al. (1982); Tversky & Kahneman (1993). Although it is orthogonal to the present discussion, it is worth noting that there is much discussion to be had

In cases such as these, the agent nevertheless retains her propositional justification, even while it is disabled by the cognitive limitations that prevent her at that moment from being doxastically justified in correct conclusions.

However, cases of misleading evidence about what rationality requires cannot always be characterised as either involving *genuine* cognitive incapacity or *evidence of* cognitive incapacity. In fact, the opposite is often the case. In Logic 101, a significant factor in your coming to have misleading evidence that rationally supports the false belief about rationality is that you are exercising your cognitive faculties well. You will only be able to acquire the misleading evidence about what rationality requires by employing your capacities for abstract philosophical reasoning well. In contrast, students in the class who do exhibit poor reasoning faculties would most likely avoid acquiring the misleading evidence. So, performance limitation does not seem to capture what is going on in this case.

One possible line of response to this is that, in fact, the students' faculties are limited in a way that could be construed as a *genuine* cognitive incapacity, and in at least two respects. Firstly, their capacities for abstract philosophical reasoning are very rudimentary – it is stipulated that this is their first philosophy class. Secondly, one might think that reasoning using paraconsistent logic, as they are doing in class, is itself a kind of incapacity[89]. However, even if this is right, there are further possible cases of rational mistake about rationality that are even more difficult to characterise as limitation in capacity. For example, cases of mistaken views about rationality held by more experienced philosophers, such as Disagreement (Ch. 1, p. 15). This mistake does not seem attributable to genuine cognitive incapacity, but rather the contrary. By employing your capacities for

---

about how experimental results indicating systematic errors in human reasoning should affect our theories of rationality (see Cohen (1981); Sperber (2011)).

[89] Boghossian (2003) and Peregrin (2014) defend this view, according to which reasoning according to classical logic is the exercise of basic reasoning competence.

abstract philosophical reasoning you have arrived at the conclusion that conciliationism offers the best story about how we should respond to peer disagreement, on the grounds that it preserves the theoretical virtues that you think are important. Even if this is the wrong conclusion to draw, it does not seem right to say that it is a mistake attributable to genuine cognitive incapacity, since your capacities for abstract reasoning are excellent, in fact some of the best that are humanly possible[90]. Of course one could insist that having some of the best capacities for philosophical reasoning humanly possible does not show that those capacities are not limited, since ideal agents would have even better capacities. However, the problem with this line of argument is that it risks generating a theory that is irrefutable. If every case of mistake within a particular domain is an instance of limited capacity in which the agent has propositional justification for the truth that is merely disabled, then it seems unclear how we could challenge this theory. No considerations pertaining to the agent's epistemic situation could suffice to challenge the claim that the agent has indefeasible justification, and this makes the claim trivial. So, it cannot be right that all apparently rational mistakes about what rationality requires are instances of genuine cognitive incapacity.

Nor do the cases at issue seem to be instances of *evidence of* cognitive incapacity, the remaining option for how justification could be disabled. The following is a typical example of *evidence of* cognitive incapacity:

> *Coffee*. Suppose that I work out my proof of T after having coffee
> with my friend Jocko. Palms sweaty with the excitement of
> logical progress, I check my work several times, and decide that

---

[90] The history of mathematics offers various cases of people with some of the best conceptual and mathematical reasoning capacities of their day making mistakes about mathematics (see Jeshion (2000)). Frege's Basic Law V is an example from logic. Similar possibilities seem possible for reasoning about what rationality requires.

the proof is good. But then a trusted colleague walks in and tells me that Jocko has been surreptitiously slipping a reason-distorting drug into people's coffee--a drug whose effects include a strong propensity to reasoning errors in 99% of those who have been dosed (1% of the population happen to be immune). He tells me that those who have been impaired do not notice any difficulties with their own cognition--they just make mistakes; indeed, the only change most of them notice is unusually sweaty palms. (Christensen 2007: 3).

According to a view that accepts the Indefeasibility claim, misleading evidence that my cognitive abilities are impaired prevents me from making the correct logical inferences involved in the proof of T that are licensed by my propositional justification, but it does not affect my propositional justification. Were it not disabled, I could use this justification to draw the correct logical inferences.

However, there are important differences between cases like Coffee and cases like Logic 101 and Disagreement. In Coffee, the experience I have is one of receiving misleading evidence that my logical reasoning capacities are impaired. The evidence I have is empirical and about my own reasoning capacities. This evidence prevents me from being doxastically justified in forming beliefs based on propositional justification for the logical facts, and this is true for any piece of logical reasoning I might undertake while I have reason to suspect that I might be under the influence of the reason distorting drug. This is true regardless of the particular content of that reasoning, my propositional justification is disabled so long as I have evidence that I am under the influence of the drug. This is quite different to having misleading evidence about what rationality requires. In Logic 101 and Disagreement, your evidence consists of propositions that bear on the truth of the false claim about rationality. Unlike the evidence in Coffee these

propositions do not concern the agent's own reasoning capacities, nor do they concern the evidence the agent has. Instead, the propositions are themselves evidence – albeit non-entailing evidence – for the false belief about what rationality requires.

In summary, we have two reasons to reject Indefeasibility. It is ill-motivated, and it rules out a very natural explanation of cases of misleading evidence about what rationality requires, leaving us unable to provide an adequate explanation to fill the space. Furthermore, even if we were to accept Indefeasibility, this would not be sufficient to establish the Impossibility Thesis unless a further claim – Sufficiency – is true. However, as the following section argues, Sufficiency is false.

4.3 Sufficiency

Access and Indefeasibility are not sufficient by themselves to licence the transition from Justificatory Assets to the Impossibility Thesis. Establishing the Impossibility Thesis also requires Sufficiency.

> *Sufficiency:* Our justification for the truth about what rationality
> requires is sufficient to rule out the possibility of rational false
> belief about what rationality requires.

Without sufficiency, even if it were true that we have access to indefeasible justification for the truth about what rationality requires, further argument would be required to establish how and why this justification rules out the possibility of rational mistakes about rationality. This section argues that Sufficiency is false, since even if we do have access to indefeasible justification for the truth about what rationality requires, these justificatory assets would be too epistemically inert to rule out rational mistakes about what rationality requires.

For justificatory assets to rule out the possibility of rational false belief about what rationality requires, they need to play a role in determining whether or not an agent's beliefs are rational. However, it does not seem that they do. We can illustrate this by comparing agents in good and bad epistemic situations, but who have the same justificatory assets, according to Access and Indefeasibility. Suppose that Anita is a student who takes Logic 101. She has good logical and philosophical abilities, pays attention, and studies as well as can be expected of her. She becomes competent at using paraconsistent logic, and she studies the best arguments for the false view that sometimes rationality requires contradictory belief. She comes to hold this belief on the basis of evidence that supports – but does not entail – it, her professor's testimony and the arguments she studies in class. If Access and Indefeasibility are true, then Anita also has justification to believe the true proposition, that rationality prohibits contradictory belief.

However, Anita's justificatory assets, assuming she has them, are compatible with her being doxastically justified in believing falsely about what rationality requires. Her false belief is based on reasons that support it, so according to a traditional view of propositional justification, this would imply that she is also propositionally justified in this false belief. Furthermore, despite her propositional justification, Anita lacks a viable deliberative route to the truth about what rationality requires. The considerations that would support the truth about what rationality requires are not considerations that Anita could, given her current epistemic situation, be doxastically justified in believing. In order to reason rationally from her current beliefs to true beliefs about what rationality requires, Anita would need to come to believe propositions that are compatible with her current beliefs. However, to do this with justification, Anita would need sufficient reason to reject the testimony of her professor and the arguments she

studies in class. Given that she currently has reasonably strong evidence supporting her false beliefs about what rationality requires, this would require extensive independent research that it does not seem reasonable to expect her to undertake. Given her intellectual capacities of a reasonably intelligent beginner philosophy student, and her role as a student of the class, she can plausibly be expected to pay attention and complete the homework assignments, but not to spend an excessive amount of time in the library undertaking independent research in the philosophy of logic. So, Anita's putative propositional justification does not play much of a role in determining whether she is rational. This suggests that sufficiency is false.

This is true in the good cases as well as the bad cases. Consider one of Anita's peers, Gertrude. Gertrude takes a more traditional introductory logic class, which covers classical logic. Suppose that like Anita, Gertrude has typical logical and philosophical abilities for her level, pays attention in class, and studies as well as can be expected of her. She becomes competent at using classical logic, and she studies some of the proofs of the classical law of non-contradiction. She comes to believe on this basis that it is never rational to believe contradictions. If Access and Indefeasibility are true, then Gertrude also has justification to believe the true proposition that rationality prohibits contradictory belief. However, it is unlikely that her belief is based on this justification. Justification for the law of non-contradiction that supports its conclusion to a maximal degree, and is immune from the arguments of dialetheists, is most likely extremely complicated, and not obviously comprehensible by a beginner. So, even though Gertrude believes what is supported by her justificatory assets, it is not plausible that she does so on the basis of those assets[91]. If this is right, then these assets are inert – they are

---

[91] We might say that she believes in accordance with the reasons, rather than for them. For further discussion, see Mantel (2018).

unnecessary for the agent's being justified in believing true propositions about rationality. Furthermore, according the Justificatory Assets claim, Gertrude has justification for both very simple logical truths, such as modus ponens, and very complicated logical truths. However, there is clearly a difference in Gertrude's epistemic situation with respect to the simpler and the more complicated logical truths, but appealing to her putative justificatory assets does not allow us to make this distinction. This implies that our justificatory assets are inert in the sense that they do not make a difference to which beliefs we should have. Without something that connects us to our allegedly vast store of justificatory assets, they can play no epistemic role in either good or bad cases – they are epistemically inert. In summary, even if we do have justificatory assets compatible with both Access and Indefeasibility, this is not sufficient to rule out the possibility of rational mistakes about rationality.

5. The Enkratic Principle

Finally, the Impossibility Thesis has frequently been defended by appeal to the Enkratic Principle. Assuming that the true requirements of rationality impose strict liability (as Externalism claims), then it is impossible to be rational while failing to conform to them. If rationality requires believing P, then any failure to believe P is a failure to be rational. If rationality also requires level coherence, then this will mean that in so far as the agent takes any attitude towards what rationality requires, she must believe truly that rationality requires P. In other words, she must have the following beliefs:

Rationality requires believing P

P

If she instead believes falsely that rationality forbids or does not require believing P, then this would be an instance of failure to conform to the Enkratic Principle,

and so an instance of irrationality on the assumption that the Enkratic Principle is a requirement of rationality. Furthermore, there are reasons to think that the Impossibility Thesis is a consequence of the Enkratic Principle (see Titelbaum 2015a, 2015b). If this is right, and if we had independent reasons for accepting the Enkratic Principle, then we should also accept that there is a requirement not to believe falsely about what rationality requires. However, this result can be read in two ways. We can either take it to show that we are committed to there being a requirement not to believe falsely about what rationality requires, or we can take it to show that since the Enkratic Principle commits us to this requirement, we ought to give up the Enkratic Principle.

The problem with this way of arguing for the Impossibility Thesis is that it would require us to establish antecedently that the impossibility of rational mistakes about rationality is more plausible than giving up the Enkratic Principle. However, the foregoing sections have shown this thesis to be difficult to justify. Even friends of the thesis admit it is counterintuitive. Littlejohn introduces his solution to the puzzle, which involves a commitment to the thesis, as 'the best of a bad bunch' (2015: 11). What he means by this is that if one wants to retain a commitment to Externalism, and one is uncomfortable with the idea that rationality sometimes generates dilemmas, then one faces a choice between giving up the Enkratic Principle or accepting the Impossibility Thesis. For this line of reasoning to work, it must be established that the costs of the Impossibility Thesis are less problematic than giving up the Enkratic Principle.

In fact, as the following chapter argues, giving up the Enkratic Principle is the best strategy for resolving the Puzzle of Rational Requirement. Chapter 3 argues that the Enkratic Principle is best thought of not as a requirement of rationality, but rather as a defeasible indication of the agent's deserving non-negative

appraisal. This strategy allows us to resolve the puzzle while allowing for the possibility of rational mistakes about what rationality requires.

6. Summary

This chapter has considered and rejected arguments for the Impossibility Thesis, namely: an argument from the more general claim about the impossibility of false beliefs (§2), an argument from the nature of rational competence (§3), an argument that appeals to our putative justificatory assets for the truth about what rationality requires (§4), and finally an argument from the Enkratic Principle (§5). Since there is no good argument for the Impossibility Thesis, we should reject it. The falsity of the Impossibility Thesis means that rational mistakes about rationality are possible, and so the puzzle presented in Chapter 1 cannot be dissolved by denying the possibility of the kinds of mistakes that generate it. The following chapter argues that the best option for solving the Puzzle of Rational Requirement is to deny that the Enkratic Principle is a requirement of rationality, and think of it instead as a principle of appraisal.

Part 2

Chapter 3

Demoting the Enkratic Principle

The previous chapter argued that we ought to accept that rational mistakes about what rationality requires are possible, since there are no good arguments for the view that they are impossible. This means that we face pressure to solve the Puzzle of Rational Requirement introduced in Chapter 1. This chapter argues that the best way to solve the puzzle is by giving up the Enkratic Principle, and thinking of it not as a requirement of rationality, but instead compliance with it as a defeasible indication of positive epistemic appraisal. This change in status allows us to solve the puzzle by diagnosing it as arising from an attempt to employ two distinct kinds of evaluation simultaneously. When we keep these two kinds of evaluation separate, the puzzle does not arise. This chapter outlines and defends this proposed solution, and shows how this solution allows us to accommodate the intuitions that have contributed to the Enkratic Principle's popularity as a requirement of rationality, while avoiding the costs of existing alternative solutions.

Section 1 outlines the traditional motivations for the Enkratic Principle. Section 2 distinguishes evaluations of requirement from evaluations of appraisal, and argues that these two kinds of evaluations ought to be kept separate. Section 3 makes use of the distinction between requirement and appraisal to diagnose the conflict of the puzzle. Section 4 defends the claim that it is the Enkratic Principle, rather than any of the other conflicting principles, that ought to be associated with appraisal rather than requirement. Section 5 illustrates in more detail how agents with misleading evidence about what rationality requires are to be

evaluated, according to the view defended here, and highlighting the distinction between evaluations of requirement and appraisal.

1. The Enkratic Principle

The Enkratic Principle is one of the three commitments needed to generate the puzzle of intra-domain conflict.

> *Externalism:* What rationality requires is completely determined by the facts about what rationality requires ("the rational requirements").

> *Evidentialism*: S(e)p → O(S(Bp)). If S has evidence that supports P, then S is rationally required to believe that P.

> *Enkratic Principle*: O (BOΦ → Φ). Rationally requires that agents do as they believe they ought to do.

The Enkratic Principle prohibits having the attitude mentioned in the antecedent without also having the attitude mentioned in the consequent. As it is stated here, the agent always has two options for complying with the principle – either forming both the antecedent and the consequent attitudes, or forming neither. To recap, suppose that rationality requires believing P, and S has some misleading evidence that indicates that rationality prohibits believing P. By Evidentialism, S is required to believe "rationality prohibits believing P". By the Enkratic Principle, S is required to be level coherent, so if S believes she is prohibited from believing P (as she should, if given her evidence), then S should not believe P. But, by the true requirements of rationality, S is also required to believe P. These leads to the apparent intra-domain conflict that motivates the puzzle: S appears to be both required to believe and *not* believe P.

The Enkratic Principle has traditionally enjoyed a largely uncontested status as a requirement of rationality[92]. This chapter argues that we ought to reject this orthodoxy. Rather than thinking of the Enkratic Principle as a requirement of rationality, we should view it as a principle of appraisal, and compliance with it as a defeasible indication that the agent has exhibited epistemic conduct that deserves positive epistemic appraisal. By demoting the enkratic principle in this way – from requirement of rationality to principle of appraisal – we can solve the puzzle of epistemic rationality while accommodating the intuitions that have motivated the view that the Enkratic Principle is a requirement of rationality. This section outlines some key motivations for thinking that the Enkratic Principle is a requirement of rationality, and argues that all but one of them ought to be disregarded. Section 4.2 argues that even the remaining motivation can be accommodated by a view that takes the Enkratic Principle to be a principle of appraisal and not a requirement of rationality.

One motivation for thinking that the Enkratic Principle is a requirement of rationality is an appeal to the apparent absurdity of the belief combinations that violate it. This has led some to argue that denying it would fail to explain the absurdity of Moorean assertions such as "P, but I ought not believe that P" (Smithies (2012)). The apparent absurdity of these kinds of assertions has led others to argue that level incoherent belief could only be rational in cases where our minds are fragmented (Greco 2014; Davidson 2004). Others have been so confident of the Enkratic Principle's plausibility that they have relied on it in order to argue for more controversial conclusions, such as the denial of Evidentialism (Littlejohn 2015), and the impossibility of rational mistakes about rationality (Titelbaum 2015)[93]. However, such considerations are not sufficient

---

[92] For some dissent, see Coates (2012); Lasonen-Aarnio (2014); Weatherson (2019).
[93] Additional advocates include Broome (2013), Feldman (2005), Horowitz (2014), Kolodny (2005), Littlejohn (2015), Way (2011) Wedgwood (2002).

alone to establish that rationality requires compliance with the Enkratic Principle, because it is not plausible that the intuition driving these considerations is fine-grained enough for us to be sure that it is tracking rational requirement rather than some other positive epistemic status, such as blamelessness. As §4.2 argues, understanding the Enkratic Principle as a principle of epistemic appraisal is compatible with accommodating this intuition when we interpret it as tracking blamelessness rather than rational requirement[94].

A second motivation rests on the thought that epistemic akrasia is a necessarily irrational phenomenon. If this were correct, then in so far as the Enkratic Principle prohibited epistemic akrasia, it would indeed be a requirement of epistemic rationality. However, as the following paragraphs argue, epistemic akrasia is not a necessarily irrational phenomenon. 'Akrasia' is usually translated as 'weakness of the will', 'incontinence', or 'lack of mastery'. Davidson (2004), for example, identifies an epistemic equivalent to weakness of will, which he terms `weakness of warrant'. According to him, this plays a central role in further quintessentially irrational states, such as self-deception and wishful thinking. Understood in this way, akrasia – epistemic or otherwise – is a failure to do what one believes one ought to do[95], and a failure to exercise appropriate control over one's first order actions and attitudes[96]. This kind of failure is typically thought

---

[94] This is analogous to a similar argument often made about intuitions of justification and blamelessness. See Littlejohn (forthcoming); Sutton (2007); Williamson (forthcoming, 2017)).

[95] For example, as Davidson characterises it, an agent acts incontinently if and only if (*a*) the agent does *x* intentionally; (*b*) the agent believes there is an alternative action *y* open to him; and (*c*) the agent judges that, all things considered, it would be better to do *y* than to do *x*" (Davidson (2001: 22)).

[96] For example, Owens claims that genuine epistemic akrasia must meet the following two conditions: "(a) a person's (first-order) beliefs … diverge from their higher-order judgements about what it would be reasonable for them to believe and (b) these divergent (first-order) beliefs are freely and deliberately formed" (2002: 1).

of as necessarily irrational[97]. However, as the rest of this section argues, violations of the Enkratic Principle need not involve a failure to exercise control, and failures to exercise control over one's attitudes need not always be irrational.

First, belief combinations that violate the Enkratic Principle – level incoherent belief combinations – need not, and in fact perhaps never, involve a failure of control because it is not clear that our beliefs are under our control at all. That said, while it is clear that we are not able to control our beliefs directly, as we control our actions, some have suggested that we do have an alternative kind of control that is applicable to belief. For example, Owens describes a notion of control, applicable to both intentions and beliefs, according to which "we exercise control over our agency by forming a view of the merits of the proposed course of action." (2002: 388)[98]. We act akratically when we act against this view. In the case of epistemic agency, we must judge that there is some (epistemic) reason in favour of a belief in order for that belief to count as under our control (2002: 398). On this view, epistemic akrasia would involve a failure to conform to our own judgment about the merits of particular beliefs. So, if we judge that believing P has the most, epistemically, to recommend it, then we fail to exercise appropriate control over our beliefs if we fail to believe P. However, the problem with identifying epistemic control with conformity to one's own higher order judgments about what one ought to do is that it lends undue rational gravitas to higher order judgments. Conformity to one's higher order judgment might not

---

[97] As Adler sees it, irrationality is an essential feature of what it is to be akratic (a "pillar", as he puts it) (2002: 12). He goes on to argue that epistemic akrasia is impossible on the grounds that the nature of theoretical reasoning precludes the possibility of an agent's having an epistemic state that exhibited the kind of irrationality necessary for a genuinely akratic state.

[98] Relatedly, Angela Smith argues that in thinking about moral responsibility, we ought to replace the notion of control with one of rational accountability: "to say that an agent is morally responsible for something, on this view, is to say that that thing reflects her rational judgment in a way that makes it appropriate, in principle, to ask her to defend or justify it." (A. Smith 2008: 369). See also Scanlon's 'judgment-sensitivity' (1998: Ch. 6).

always be the best way to exhibit rationality, particularly when one has many false higher order judgments. In particular, if one believes falsely that one ought to believe contradictions, or refrain from believing what is supported by the evidence, it is not clear that a way to become more rational is to bring the rest of one's first order beliefs in line with these higher order mistakes[99]. This is precisely what is at issue in the puzzle of intra-domain conflict that we are focussed on solving.

Nevertheless, one might think that level incoherence exhibits a failure to manage beliefs appropriately, even while agreeing that correct belief management does not require conformity with one's higher order beliefs. For example, Davidson identifies `weakness of the warrant' – intended as the epistemic analogue to weakness of the will – as the "sin" against a requirement to believe what one's total evidence supports (2002: 201), and we might think that level incoherent belief always exhibits this `sin'. We might think that a level incoherent state could never be supported by one's total evidence, and so level incoherent belief could never be a rational response to one's evidential situation. However, this is false. While states in which the evidence supports level incoherent states are rare, they are not impossible. When evidence is misleading about itself, the evidence will support level incoherent states (see (Lasonen-Aarnio, 2010, 2014; Weatherson, forthcoming)). Evidence is misleading about itself when P is very likely on one's

---

[99] Davidson makes a parallel point with respect to practical akrasia - practical rationality is not always best served by following our own higher order judgments about what it is best to do, we can also act irrationally and against our better judgment while following bad higher order judgments, for example: "I have just relaxed in bed after a hard day when it occurs to me that I have not brushed my teeth. Concern for my health bids me rise and brush; sensual indulgence suggests I forget my teeth for once. I weigh the alternatives in the light of the reasons: on the one hand, my teeth are strong, and at my age decay is slow. It won't matter much if I don't brush them. On the other hand, if I get up, it will spoil my calm and may result in a bad night's sleep. Everything considered I judge I would do better to stay in bed. Yet my feeling that I ought to brush my teeth is too strong for me: wearily I leave my bed and brush my teeth. My act is clearly intentional, although against my better judgement, and so is incontinent." (2001: 30).

evidence, but it is also likely that P is not likely on one's evidence. Typical examples of such situations include cases in which agents have good evidence for P, but misleading evidence about their abilities to assess the evidence for P effectively. For instance, recall Medicine (p. 18), in which you make the correct diagnosis while very sleep deprived. In this case, it would seem that what your evidence supports is a high credence in the patient having Disease D, but a low credence in your evidence supporting Disease D, because it is highly unlikely that your diagnostic skills will be on form[100]. Cases such as these show that it is not clear that level incoherent states cannot be supported by the evidence, so this cannot be used to motivate the Enkratic Principle as a requirement of rationality.

A third motivation for thinking that rationality requires level coherence is the idea that rationality supervenes on the mental. Some have thought that rationality is the epistemic virtue of thinking well (Reisner (2013); Wedgwood (2017)), or that to be rational is to be deserving of a particular kind of epistemic praise, or immune from a particular kind of epistemic criticism (Kvanvig (2014); Lord (2018)). Those persuaded by this association of rationality with appraisal and epistemic virtue might have thought that this implies that rationality supervenes on one's mental states, because in the appraisal of conduct it is often mental phenomena – beliefs, intentions, and desires – that are thought to be most relevant in determining whether the agent deserves praise or blame. Those sympathetic to this idea might thus be pushed to endorse the Enkratic Principle on the grounds that it requires mental coherence, and mental coherence is generally epistemically praiseworthy. However, as I argue in later chapters, epistemic appraisal makes demands of us that go beyond the limits of the mental.

---

[100] Of course, that this is the correct response to the evidence is controversial. For some disagreement, see Brown (2018, Ch. 5, 6); Horowitz (2014); Sliwa and Horowitz (2015). To some extent, this turns on whether or not higher order evidence `screens off' first order evidence. For discussion of this, see Feldman (2007); Fitelson (2012); Roche & Shogenji (2013); Roche (2018).

According to the account of appraisal that I defend, avoiding blame requires us to respond appropriately to our reasons, where these are not necessarily mental phenomena. Furthermore, the idea that the Enkratic Principle expresses immunity from criticisability is entirely compatible with the solution defended in this chapter. The idea that compliance with the Enkratic Principle is to be associated with positive epistemic appraisal completely vindicates this motivation. This leaves only a residual verbal disagreement between the view defended in this chapter, and views that insist that rationality supervenes on the mental. Unlike views that take rationality to supervene on the mental, the view defended here does not equate 'rational' with 'blameless'. The following section presents a way to resolve this disagreement, by distinguishing two distinct kinds of evaluation involved in judgments of epistemic rationality, only one of which concerns the kind of epistemic appraisal that Kvanvig, Lord, and Wedgwood associate with judgments of rationality.

2. Two Kinds of Evaluation

This section diagnoses the puzzle's conflict as the result of conflating two distinct kinds of evaluation, requirement and appraisal, and argues that these ought to be distinguished. Distinguishing these two kinds of evaluation permits a resolution to the conflict. A judgment of the form "S is epistemically rational" may involve either or both of the following claims:

Requirement: S has the attitudes required by the requirements of rationality.

Appraisal: S has exhibited good epistemic conduct.

Evaluations of requirement are focused on the agent's attitudes, and whether they are the attitudes she is required to have. Evaluations of appraisal are focused on the agent and what she does in managing her beliefs. Agents succeed in

meeting the requirements when they have the attitudes required by rationality (evaluation of requirement), and they deserve the positive appraisal – praise, rather than blame - associated with rationality when they have exhibited good epistemic conduct (evaluation of appraisal). These can come apart – doing what is required does not always deserve praise and failing to do what is required does not always deserve blame. This distinction is well-established in ethics. For example:

> *Kant's Prudent Grocer*. A grocer prices his wares fairly, as morality requires him to do. However, he does this not out of a motivation to do what is fair, kind, or morally right, but out of a motivation to maximise his profits. He knows that if he does not price his wares fairly, his customers will go elsewhere. If he could make more profit by pricing his wares unfairly, then he would do this instead.

Kant's prudent grocer does what is required – he succeeds in complying with the requirements, but he does so in such a way that does not deserve praise[101]. Likewise, failing to meet requirements does not always deserve blame:

> *Toes.* I step on your toe in a crowded lift, and in doing so cause you pain. Causing others pain for no good reason is prohibited by the requirements of morality. However, I step on your toe not out of any intention to cause you pain, but because the lift is crowded and I am not aware of where your toe is. Had I known your toe was there, I would not have stepped on it.

Here, I do something that is prohibited – I fail to comply with the requirements, but my conduct does not deserve blame[102]. Applying an analogous distinction in epistemology, an evaluation of requirement concerns whether the agent complies with the prohibitions and requirements of epistemic rationality, while an

---

[101] This does not necessarily imply that he deserves blame. The point is that he does not deserve a positive appraisal, although he may deserve a neutral appraisal (see Arpaly (2002a) for discussion).

[102] Again, it also does not necessarily deserve praise. I may deserve a neutral evaluation, or an excuse.

evaluation of appraisal concerns whether and to what extent the agent's epistemic conduct deserves epistemic praise or blame. Considerations relevant to the appraisal of the agent's action might include what could have been expected of her, the quality of her will, or the kinds of habits and characteristics she exhibits. The exact determinants of hypological facts are discussed further in the following chapter. The rest of this section uses anti-luminosity considerations to argue that requirement and appraisal should be distinguished in epistemology.

Anti-luminosity considerations offer one important reason to distinguish requirement and appraisal in epistemic rationality. Williamson (2002) has argued for anti-luminosity, according to which there is no non-trivial condition for which it is always possible to know whether or not one has met that condition[103]. As a result, it is not always possible to know when one has met the requirements of rationality (Srinivasan (2015b); Williamson (2002)). Being mistaken about what rationality requires is one way to fail to be in a position to know whether one has met the requirements, and being unable to know when one is complying with the requirements makes it very difficult to comply with them. Failure to know what is required means that whether or not one is doing what is required is not under one's control, because we cannot tell which actions would comply with the requirements (Srinivasan (2015b)). It is implausible that one could be blameworthy for failing to conform to a requirement if one was in no position to know one was failing to conform to it. We should distinguish requirement and

---

[103] According to Williamson, trivial conditions immune to anti-luminosity are those that hold in either all or no cases, and conditions for which one cannot change from being in a position to know that it obtains to not being in such a position (2002: 108).

appraisal because doing so would allow us to avoid this implausible commitment[104].

This picture is at odds with views that understand deontic and hypological facts as closely aligned in the epistemic case. According to these views, the failure to meet one's epistemic obligations is sufficient for blameworthiness, and so requirement and appraisal are the same kind of evaluation – the deontic facts are to be identified with the hypological facts[105]. However, identifying deontic and hypological facts in this way forces us into a dilemma. Either we are stuck with an implausibly harsh view of blameworthiness, or we risk trivialising rationality in the most likely hopeless pursuit of a set of luminous requirements of rationality.

On the one hand, anti-luminosity implies that for any non-luminous requirement, there will be cases in which agents fail to meet that requirement, but were in no position to know that they were failing to meet the requirement. So, if meeting the requirements were necessary for avoiding blame, then this would generate an implausibly harsh view of blameworthiness. Agents would sometimes be blameworthy for failing to do what they were in no position to know they were failing to do. Since it would be unfair to hold agents blameworthy for failing to do what it would be very difficult for them to do, in the epistemic case as well as the moral case, this view is implausibly harsh. To avoid this, many views of epistemic normativity will permit some cases of norm violation in which the agent is blameless, for example envatted subjects when we

---

[104] It is thus broadly in line with other views that distinguish appraisal and requirement in both ethics and epistemology (see Arpaly (2002b); Graham (2010); Strawson (1962)). For applications of the distinction to epistemology, see Hawthorne and Srinivasan (2013); Lasonen-Aarnio (2010); Littlejohn (2012); Sutton (2005; 2007); Williamson (forthcoming)).

[105] Supporters of this view include Alston (1989); Cohen (1984); Ross (1939); Steup (1999).

endorse a truth or knowledge norm of belief[106]. However, identifying the determinants of blameworthiness with requirements of rationality, which will not be luminous, forces us to evaluate these agents as blameworthy, which is unpalatable.

It might have been thought that this unfairness could be avoided by amending our account of what rationality requires, such that the requirements of rationality are luminous. To this end, one might be tempted to adjust the requirements of rationality such that they depend very closely on the agent's mental states, or how things seem to her by her own lights. However, traditional arguments for anti-luminosity show that this would not be sufficient to avoid the worry about demandingness articulated above. Anti-luminosity arguments show that no non-trivial condition is luminous, not even conditions for which compliance depends on how things seem from our own lights (Williamson 2002; Srinivasan 2015a). Any genuinely luminous condition would need to be extremely trivial to ensure that agents were always in a position to know whether they were meeting it, and it is not obvious that there is any such condition. Williamson's anti-luminosity arguments apply even to the feeling of being cold. Furthermore, some have suggested that it might be even more difficult for us to discern from the inside when we meet conditions that require us to be in particular mental states than when we meet external conditions (Schwitzgebel 2006; Srinivasan 2015b). Given this, it is not at all clear that a set of luminous requirements of rationality is possible[107]. The pressure to avoid both overly demanding and overly trivial – and

---

[106] See, for example, discussion of how externalist views can respond to the New Evil Demon argument (Brown (2018); Cohen (1984); Fantl and McGrath (2009); Kelp (2016); Littlejohn (2009); Sutton (2007); Williamson (forthcoming)).

[107] Even if such requirements were possible, they would need to be very different from the typical requirements of rationality that have been traditionally endorsed, and it is difficult to see how such a requirement could be of any use in adjudicating between competing consistent belief sets, or guiding agents towards more rational belief sets. See Lasonen-Aarnio (2010) on this point.

most likely impossible – requirements of rationality means that we should distinguish requirement and appraisal. The following section outlines how such a distinction would allow us to satisfactorily resolve the conflict.

3. Solving the Puzzle

Distinguishing evaluations of requirement and appraisal means that we need not think that agents in the puzzle are subject to conflicting requirements of rationality. Instead, when agents have misleading evidence about what rationality requires, we can distinguish the question of which epistemic attitudes are rationally required or prohibited from the question of which epistemic appraisal the agent would deserve for adopting those attitudes. Since these distinct evaluations are determined by different kinds of considerations, it will sometimes be possible for agents to deserve positive (or non-negative) appraisal for adopting attitudes that are prohibited by rationality, and negative (or non-positive) epistemic appraisal for adopting states that are required. This allows us to explain why agents who have misleading evidence about what rationality requires seem to be `rational' in having attitudes that are prohibited by the requirements of rationality. Although they have attitudes that violate the requirements of rationality, they deserve positive or non-negative appraisal for doing so.

Employing this strategy, the three conflicting claims – Externalism, Evidentialism, and the Enkratic Principle – no longer conflict, so long as we do not think that all three are associated with the same evaluation, whether requirement or appraisal. In the following section I argue that it is the Enkratic Principle, rather than any of the others, that ought to be associated with appraisal rather than requirement, while Evidentialism and Externalism should be retained as requirements of rationality. Demoting the Enkratic Principle in this way would mean that rationality requires, in all cases, that agents believe what

is supported by their evidence, and that they refrain from adopting any prohibited epistemic states. When agents have misleading evidence that rationality requires avoid believing P, when it in fact requires believing P, then rationality requires the following beliefs:

> I am rationally required to avoid believing P.
>
> P

Agents who have misleading evidence that rationality requires *not* believing P, when it in fact requires believing P, would deserve positive epistemic appraisal for refraining from believing P at the first order, at least on the plausible assumption that doing what one has good reason to believe is required is usually something that deserves positive appraisal. However, in not believing P, they would be failing to conform to the rational requirement to believe P. In distinguishing requirement from appraisal, we need not also think that they deserve negative appraisal for this failure, at least in so far as their failing to believe P can be attributed to non-blameworthy epistemic conduct. This strategy of giving up the Enkratic Principle as a requirement of rationality has been used as a solution to similar puzzles involving misleading higher order evidence (see Weatherson (2019), Lasonen-Aarnio (2014)). According to this solution, there are possible cases of rational belief that violate the Enkratic Principle, namely cases in which the evidence is misleading about itself. So long as it is possible both that P is likely on one's evidence, and that it is likely that P is not likely on one's evidence, then on an evidentialist conception of epistemic rationality, rationality will sometimes require level incoherent belief. Likewise, this solution allows us to understand puzzle cases such as Logic 101 (Ch. 1, p. 12) and Disagreement (Ch. 1, p. 15) as cases in which rationality requires a normative belief about what rationality requires that conflicts with a rationally required first order belief.

I have now outlined how denying that the Enkratic Principle is a requirement of rationality provides a solution to the puzzle. Nevertheless, it would also be possible to resolve the puzzle by demoting either of the other two conflicting principles to the status of principle of appraisal. As such, we need some argument for thinking that it is the Enkratic Principle in particular that ought to be demoted in this way. The following section makes this argument.

4. Demoting the Enkratic Principle

This section argues that it is the Enkratic Principle, rather than Externalism or Evidentialism that should be demoted to the status of principle of appraisal. There are two reasons to think that it is the Enkratic Principle that should be thought of as a principle of appraisal rather than a requirement of rationality. First, that it is the least theoretically costly of the available options, and second, that one of the main motivations for thinking of level coherence as a requirement of rationality is equally, if not better, accommodated by a view that takes it to be a principle of appraisal.

First, giving up the enkratic principle is the least theoretically costly of the available options. While the puzzle could be solved by giving up any of the three commitments, rejecting either of Externalism or Evidentialism involves significant theoretical costs. This leaves rejecting the Enkratic Principle as the least costly solution available[108]. The rest of this section discusses the costs associated with giving up Externalism and Evidentialism, and argues that giving up the Enkratic Principle is comparatively cost-effective. First, the costs of giving up Externalism. Recall that Externalism is the following claim:

---

[108] Alternative solutions and dissolutions of the puzzle that do not involve rejecting any of the three commitments were rejected in Part 1.

> *Externalism:* What rationality requires is completely determined
> by the facts about what rationality requires.

According to Externalism, the requirements hold independently of what agents subject to them believe, what evidence they have, what they are in a position to know, and all other features of the agent's epistemic situation. So, if rationality requires you to refrain from believing contradictions, then it requires you to refrain from believing contradictions regardless of whether you believe that you are required to refrain from believing contradictions. Rejecting Externalism means denying that what rationality requires depends on facts that hold independently of your perspective. Instead, what it is rational for you to believe would depend on facts about your perspective – facts such as how things appear to you, what you already believe, and the evidence you have available[109]. However, as Chapter 1 (p.13-15) argued, there are significant costs to accepting such a view. The following paragraph reiterates and expands upon these.

First, denying Externalism means denying that there are facts about what rationality requires that are external to the agent's perspective. In other words, any attitude at all could in principle count as rational, provided that one has sufficient evidential support for it. This is a problem because it makes somewhat mysterious why we should think that rationality is valuable, particularly why we should think it is valuable for agents who already have many false beliefs to be rational. If being rational is a matter of being consistent from your own perspective, then there is no reason to think that being rational will lead you to epistemic goods such as truth and knowledge, since rationality demands only that you have beliefs that are supported by your perspective. In fact, there is

---

[109] See, for example, H. Field (2009); Gibbons (2013); Lord (2018); Kolodny (2005); Kiesewetter (2011); Kiesewetter (2013); Kvanvig (2014); Raz (2005); Whiting (2014); Way & Whiting (2017); Zimmerman (2008).

reason to think that a perspectival requirement of rationality will exasperate the epistemically negative consequences of false belief by further isolating us from epistemic goods[110].

Second, a traditional motivation for perspectivism is the thought that it should always be up to you whether or not you meet the requirements of rationality, and that this is only possible if what rationality requires of you depends on your perspective. For example, this consideration motivates Kvanvig to sharply distinguish the 'normative' and the 'evaluative' dimension, and hold that the normative dimension depends entirely on the agent's perspective (2014: 49). However, when we recall the consequences of anti-luminosity, we see that this motivation cannot support Perspectivism. Williamson's anti-luminosity arguments show that no non-trivial condition can satisfy this desire[111].

Third, endorsing a perspectival view about requirements of rationality means endorsing an error theory about traditional requirements, such that a statement of what rationality required of an individual would be impossible to make in advance of considering the agent's precise situation. This would, surprisingly, make cases of misleading evidence about what rationality requires counterexamples to particular putative requirements of rationality. This would constitute a significant revision to how we ordinarily think of rational requirements, a revision that would seem unwarranted without good reason to think it necessary[112].

---

[110] This is a version of a familiar objection to coherentist theories of justification, namely that entirely false belief sets could count as justified if all that is required for justification is coherence (see Sosa (1980: 19)). The objection is particularly worrying in light of the consideration that those who believe conspiracy theories often have beliefs that are largely consistent and well-supported from their perspective (see Nguyen (forthcoming (a); forthcoming (b))).

[111] See also Srinivasan (2015b) on this point.

[112] See Lasonen-Aarnio (2014: 332) on this point.

Fourth, giving up external requirements of rationality would make it difficult for epistemic rationality to guide agents towards epistemically better belief sets. This is because if we think that rationality depends entirely on how things seem from one's perspective then no particular traditional requirement of rationality will be genuinely binding for all agents. Instead, the agent's perspective determines which requirements that agent is subject to. One unwelcome consequence of this is that requirements of rationality cannot provide information about which kinds of epistemic states are rational. This means that they cannot be used for epistemic guidance, since many different states could be supported by one's perspective, depending on the specifics of the agent's epistemic situation. These specifics are not easily formulated into general principles that can be used by agents to guide their epistemic activities. These are serious theoretical costs that should make us hesitant to give up Externalism without good reason.

Similarly serious theoretical costs come with giving up Evidentialism. While some have taken this option in order to resolve the conflict (Littlejohn 2015), their key motivation have typically been to preserve the Enkratic Principle (Lasonen-Aarnio (forthcoming); Littlejohn (2015)). For this to be a plausible motivation, we would require some independent reason to think that the Enkratic Principle is a more important commitment of epistemic rationality than Evidentialism. However, this is implausible. Evidentialism has had at least as great a following as the Enkratic Principle, if not greater. Various otherwise distinct accounts of epistemic rationality retain a commitment to it[113]. So, the fact that giving up Evidentialism would allow us to preserve the Enkratic Principle as a requirement of rationality does not itself give us a reason to prefer this strategy. Giving up

---

[113] See, amongst many others, Feldman and Conee (1985); Greco (2014); Huemer (2011a); Joyce (2009); Kelly (2002); Lasonen-Aarnio (2014); Smithies (2012); Williamson (2002); Weatherson (2019).

either Externalism or Evidentialism incurs significant theoretical costs, which are to be avoided if possible. Furthermore, the key motivation for thinking of the Enkratic Principle as a requirement of rationality can be accommodated just as well, if not better, by a view that takes it to be a principle of appraisal. This makes it the most appropriate of the three principles to 'demote' to the status of a principle of appraisal.

A further significant reason to think that the Enkratic Principle is the most appropriate of the three to demote is that the key motivation for it can be just as well, if not better, accommodated by thinking of it as a principle of epistemic appraisal rather than a requirement of rationality. Section 1 dismissed all but one of the key motivations for the Enkratic Principle, namely that to be rational is itself a kind of appraisal. According to this idea, to be rational is to be deserving of a particular kind of praise, or at least undeserving of a particular kind of criticism (Kvanvig (2014); Lord (2018: 4); Wedgwood (2017)). If we think that to be rational is to deserve epistemic praise, then agents who comply with the Enkratic Principle would seem to deserve epistemic praise in virtue of having managed their beliefs well. For example, violating the Enkratic Principle is likely to correlate with epistemic mismanagement, since being level incoherent would seem to often involve ignoring reasons to revise one's beliefs that ought to strike one as salient. An agent who believes 'I ought to believe P', but does not revise her beliefs so that she comes to believe 'P'; or indeed who believes 'I ought not believe P', but continues to believe P anyway, is failing to manage her beliefs well (Horowitz (2014); Lasonen-Aarnio (forthcoming)). Complying with the Enkratic Principle, one might think, ensures that you have done everything that could be expected of you from your perspective, and this protects you from criticism. However, this is not a reason to think that the Enkratic Principle is a requirement of rationality: it is a reason to think that it is a principle closely associated with

positive epistemic appraisal. Furthermore, violating the Enkratic Principle seems to involve the misuse of higher order beliefs. A plausible description of the function of higher order beliefs is to regulate our first order beliefs, and agents ignore this function when they fail to comply with the Enkratic Principle (see Christensen (2010a; 2010b; 2009); Littlejohn (2015)). Again, this consideration, while plausible, can be equally well – if not better – accommodated by a view that thinks of compliance with the Enkratic Principle as indicative of positive epistemic appraisal. As a result, giving up the Enkratic Principle, and thinking of it instead as an indication of positive epistemic appraisal involves no particular theoretical costs. This makes it a preferable solution to the alternative solutions of giving up either Externalism or Evidentialism.

Having presented the argument for this solution, the following section illustrates how this solution evaluates agents who have misleading evidence about what rationality requires, and shows how compliance with the Enkratic Principle can serve as defeasible indication that the agent deserves positive epistemic appraisal, thus solving the puzzle.

5. Compliance as a Defeasible Indication

This section illustrates how agents who have misleading evidence about what rationality requires are to be epistemically evaluated under the proposed solution.

Complying with the Enkratic Principle usually coincides with good belief management, and so can serve as a defeasible indicator that the agent deserves positive appraisal. For example, it usually indicates that she has responded to pressure to resolve apparent inconsistencies, or adopted the recommendations of her most well considered views about what she ought to believe. Despite this, being in an epistemic state that complies with the Enkratic Principle is neither

necessary nor sufficient for being deserving of positive evaluation. It is not *sufficient*, because merely having the beliefs specified by the Enkratic Principle is not something to be appraised positively. To illustrate this, consider two different agents, Diligent and Lazy, who have misleading evidence about what rationality requires:

Diligent's beliefs:	*Lazy's beliefs:*
I am rationally required to believe P	Rationality requires *not* believing P
P	~B(P)

Suppose that both are students of Logic 101, the misleading logic class from Chapter 1 (p. 12). Both agents are in situations where their evidence misleadingly supports the false view that rationality requires, in this case, belief in a contradiction. P represents an instance of the Liar sentence, and rationality (we suppose) in fact forbids believing P. Although both agents have attitude combinations that comply with the Enkratic Principle, the appraisal they deserve depends on the wider story of how they arrived at these beliefs. Diligent, let us suppose, considers her evidence seriously, and arrives at the false belief that rationality requires believing P. Then, following the recommendations of this false belief she forms the attitude recommended by her false views about what rationality requires. She conforms to the Enkratic Principle, and she has exhibited good epistemic conduct in doing so. In this instance, her conformity to the Enkratic Principle coincides with epistemic conduct deserving of epistemic appraisal. However, this is not the case for Lazy. Lazy, let us suppose, responds very differently to his evidence. He finds the idea that rationality could ever require believing P intuitively implausible, and so instead of carefully considering his evidence, he disregards it and reasons from the intuitive

implausibility of P to the belief that the view of what rationality requires that is supported by his evidence must be false. Although, in fact, he arrives at the correct result, he does so in a way that is to be appraised negatively. In reasoning "upstream"[114] he disregards his evidence, and he does so only because he finds the view of rationality supported by his evidence intuitively implausible. This is not a good reason to dismiss misleading evidence, and so he deserves negative epistemic appraisal, even though his beliefs conform to the Enkratic Principle[115]. What matters for appraisal is not whether the agent's beliefs conform to the Enkratic Principle, but rather what the agent has done, epistemically, to arrive at those beliefs.

Neither is conformity to the Enkratic Principle *necessary* for non-negative appraisal. To see why, consider a third agent:

> *Incoherent's beliefs:*
>
> Rationality requires *not* believing P
>
> P

Incoherent has a level incoherent epistemic state – she fails to conform to the Enkratic Principle. However, the appraisal that she deserves depends on how she has ended up in this state. If we suppose that she has arrived at this belief combination by failing to fully consider the implications of her beliefs about what rationality requires, then she would seem to deserve negative appraisal, although to a lesser degree than Lazy. However, we can also imagine an alternative backstory according to which Incoherent arrives at her belief combination via

---

[114] As Kolodny (2005: 529) puts it. See also Schroeder's 'symmetry' objection to thinking of the practical Enkratic Principle as wide scope (Schroeder (2004: 339)), which points out that only some of the ways one could bring oneself in line with the Enkratic Principle intuitively seem rational.

[115] Lewis (2004) also responds to the arguments of dialetheism in this way, but we can assume that as an experienced philosopher he has more reason to do this than the mere untrained intuitions that beginners would have to rely on.

conduct that deserves positive appraisal. For example, suppose that epistemologists in the future develop a device (the Excellent Evidence Evaluator) that can perfectly evaluate what one's evidence supports in any case, and that everyone uses these devices and comes to depend on them. Suppose that Incoherent has one of these devices, and that her higher order evidence suggests that her first order evidence is misleading, when in fact it is not. In this case, her total evidence supports both P and that her evidence does not support P. Assuming that rationality prohibits believing what is not supported by the evidence, it is plausible that in trusting the device and violating the Enkratic Principle, Incoherent would deserve positive appraisal for managing her beliefs well. In fact, this possibility is consistent with the position of some of the Enkratic Principle's defenders. Horowitz, for example, argues on evidential grounds that compliance with the Enkratic Principle is necessary for rational belief in the majority of cases, but concedes that there are some cases in which the higher and first order evidence support incompatible propositions, and that in these cases the Enkratic Principle does not apply (2014: 735-40).

In summary, when an agent has misleading evidence about what rationality requires, she will be irrational if she comes to have the first order beliefs recommended by those false beliefs, in virtue of violating the true requirements of rationality. In fact, she is required to have a combination of beliefs that violates the Enkratic Principle. She is required to believe falsely about what rationality requires, because this is what her evidence supports, and she is required to refrain from forming any rationally prohibited beliefs at the first order. So, if her evidence supports the false proposition that she is rationally required to believe a contradiction then she is required to believe this false proposition, but refrain from actually believing any contradictions. Since, on this view, the Enkratic Principle is also not a requirement of rationality, she would not be violating any

requirements of rationality in having this combination of beliefs. However, she may deserve positive epistemic appraisal if it is the case that she has managed her beliefs well. Very often, complying with the Enkratic Principle is a way to manage one's beliefs well. This means that we can take compliance with the Enkratic Principle as a defeasible indication of positive epistemic appraisal. Nevertheless, compliance with the Enkratic Principle is neither necessary nor sufficient for positive epistemic appraisal – conformity to the Enkratic Principle contributes nothing, on its own, to an agent's epistemic appraisal.

6. Summary

I have argued that the way to solve the conflict arising from misleading evidence about what rationality requires is to think of the Enkratic Principle as a principle of epistemic appraisal, and not a requirement of rationality. This change in status permits a resolution to the conflict that diagnoses the conflict as arising from a conflation of two distinct kinds of evaluation – requirement and appraisal. This approach allows us to both accommodate the intuitions that have motivated others to think of the Enkratic Principle as a requirement of rationality, while also resolving the apparent conflict. The following chapter further explicates this strategy by suggesting that the correct appraisal of agents who fail to comply with the requirements of rationality because they have misleading evidence about what rationality requires is that of excuse. This means we require an account of epistemic appraisal, and specifically excuse, that is independent of an account of what it is to meet the requirements of a particular domain. I suggest that a promising account that does this is the account that appeals to what it is reasonable to expect. Part 3 goes on to develop this account.

Chapter 4

Giving an Account of Excuse

This chapter gives an account of when agents can be excused from blame for failing to comply with requirements, which builds on the solution to the puzzle presented in the previous chapter. Specifically, agents can be excused from blame provided they have done what it is reasonable to expect of them. This serves to clarify the claim of the previous chapter that the puzzle is a case in which agents violate requirements of rationality but deserve non-negative appraisal. This chapter clarifies what kind of non-negative appraisal the agents deserve, and why: they deserve an excuse for epistemic blame they would ordinarily deserve for violating requirements of epistemic rationality, because they have done what it is reasonable to expect of them.

Section 1 argues that the kind of non-negative appraisal that the agents in the puzzle deserve is an excuse, and the remaining sections discuss competing accounts of epistemic excuses. Section 2 discusses and rejects one account of excuse, Williamson's derivative norms account. Section 3 argues that an account of excuse should include a requirement that the agent formed her beliefs responsibly, and discusses and rejects a way of fleshing out this notion by making reference to habits. Section 4 considers a strategy for fleshing out the notion of responsible belief formation based on Rosen's (2002, 2004) account of what it is to meet the standards of epistemically responsible belief. It argues that there are two respects in which this account is incorrect: in its claim that responsible belief formation never requires agents to adopt or avoid any particular belief, and in its claim that we are always permitted to believe what is obvious. This motivates the account of epistemic appraisal that I defend and present in Section 5. According

to this account, agents are blameworthy when, and only when, they fail to do what it is reasonable to expect. On this view, false normative belief can sometimes, but not always, excuse, and the Puzzle of Rational Requirement is one example of this. Agents who form a false belief about rationality that is supported by their evidence are rationally required *not* to follow the recommendations of this false belief, but would not necessarily deserve negative appraisal for doing so – provided they had done what it was reasonable to expect of them. This means that the solution to the puzzle proposed in the previous chapter relies heavily on its being the case that false normative belief can, at least sometimes, excuse. However, this is controversial. In Part 3 I go on to consider whether false normative belief can excuse.

1. Epistemic Blamelessness

This section argues that the kind of non-negative appraisal we make of agents in the puzzle is one of excuse, rather than the alternative possibilities of justification or exemption. Subsequent sections discuss competing accounts of epistemic excuse that offer different accounts of the conditions under which agents deserve epistemic excuses.

Traditionally, there are three ways to avoid blame: by having a justification, an excuse, or an exemption[116]. These are typically thought of as ways to defend an action that is (or is held to be) bad[117]. A justification offers a defence of the agent that shows how the action was not, in fact, bad; an excuse offers a defence that admits that the action was bad, but denies that the agent is criticisable for doing

---

[116] This taxonomy is also found in Baron (2007); Littlejohn (forthcoming); Strawson (1962).

[117] As Austin, for example, puts it, excuses are relevant, "where someone is said to have done something which is bad, wrong, inept, unwelcome, or in some other of the numerous possible ways untoward" (Austin (1979: 175-6)). Husak (2005: 566) denies that excuses are only applicable to actions deemed to be bad, and lacking justification, drawing on linguistic data involving the English phrase "excuse me". On this point, I follow Baron (2007: 24) in denying the relevance of this linguistic data for understanding ways of avoiding blameworthiness.

it. Exemptions show that the agent is not subject to the demands that prohibit the action (Baron (2007: 32))[118]. This taxonomy can be applied to moral, epistemic, and other kinds of blameworthiness[119].

Applying this traditional picture to considerations of agents who meet or fail to meet requirements, we can explicate the three ways of avoiding blame in the following way, understanding agents as blameworthy when they fail to do as they are required to do in absence of a justification, excuse, or exemption:

> *Justification:* The agent does what she is required to do.

> *Excuse:* The agent does not do what she is required to do, but there is some consideration(s) such that she does not deserve blame for failing to do as she is required to do.

> *Exemption*: Requirements that would normally apply do not apply in this case.

Given this taxonomy, when an agent is blameless, it is a further question whether this is because she is justified, excused, or exempt. The following example illustrates the differences between these ways to be blameless. Suppose you promise to go to watch Aston Villa play their big game of the season with your brother. Morality, let us suppose, requires you to keep this promise (all things being equal). So, you would deserve blame for breaking this promise, unless you have a justification, excuse, or exemption. Suppose you break the promise – you fail to show up at the game. You might be able to avoid the blame you would ordinarily deserve, but this depends on why you failed to show up. Suppose you

---

[118] Those who hold this taxonomy typically take exemptions to be grounded in a lack or compromise of capacity, see for example Littlejohn (forthcoming: 10); Strawson (1962).

[119] For example Bjornsson holds that blame can be appropriate in response to any activity that can be performed with "greater or lesser excellence", including "remembering things, or solving mathematical problems" as activities that can be deserving of "skill blame" (Björnsson (2017: 153)).

failed to show up because you were saving a drowning child. Assuming that the moral requirement to rescue drowning children is more important than the requirement to keep your promises, this is a case in which you would be justified in failing to meet the requirement to keep your promise to your brother – in breaking the promise you were fulfilling a more important requirement. In this case, morality did not require you to keep the promise to go to the game, since you had a more important requirement that conflicted with it, to save the child[120]. Alternatively, suppose that you failed to show up because you confused Aston Villa with Arsenal, and turned up at the wrong football game. Suppose also that this was an easy confusion for you to make, such that it would not have been reasonable to expect you to realise the mistake, and that you were trying to go to the game. In this case you deserve an excuse. You still failed to do what was required – you did not keep your promise to go the game, but the mistake is a consideration that means you do not deserve blame. Alternatively, suppose you break your leg on the morning of the game, and this prevents you from going. This would be grounds for an exemption, releasing you from the requirement to keep your promise. Although you did not keep the promise, in this case you are no longer required to keep the promise[121].

The previous chapter argued that the right solution to the Puzzle of Rational Requirement was to distinguish between evaluations of requirement and appraisal. When agents follow misleading advice about what rationality requires, and violate requirements of rationality, they will end up with irrational

---

[120] A plausible view that is consistent with the view I defend here is that one deserves praise for having acted rightly. However, a full discussion of praise is orthogonal to this discussion. For further discussion of the relationship between praise and moral requirement see Johnson-King (2018); Arpaly (2002a); Markovits (2010); Mantel (2018).

[121] There is some debate to be had over whether exemptions are really a sui generis category, or rather a kind of full excuse. Since the puzzle does not directly concern full excuses, I leave this aside.

beliefs. However, they do not necessarily deserve negative appraisal for this – sometimes these violations are blameless[122]. However, for this to work as a solution I will argue that it needs to be the case that the non-negative appraisal they deserve is one of excuse, rather than justification or exemption. As the following paragraphs show, neither of the alternative non-negative appraisals (justification or exemption) would offer a satisfactory solution to the puzzle.

If the non-negative appraisal the agents deserved was justification, then this would not allow us to solve the puzzle. Being justified means that one does what one is all things considered required to do. If the agents in the puzzle are justified in having the first order attitudes recommended by misleading views about rationality, then this would generate a view that was both inconsistent with the assumption that requirements of rationality apply universally, and that begs the question against the arguments of the previous chapter. If the agents in the puzzle were justified in violating the requirements of rationality, then this would imply that when they follow the recommendations of misleading evidence about what rationality requires, there is some further rational requirement, the Enkratic Principle, that they fulfil that justifies them in following the misleading advice. This would imply that some requirements of rationality can sometimes be overridden by a further requirement to comply with the Enkratic Principle, and furthermore that the requirement to be enkratic is so important that it can sometimes override other putative requirements, nullifying those requirements in that case[123]. This is inconsistent with the solution advocated in the previous chapter, and furthermore it is implausible that the Enkratic Principle's merits

---

[122] This is in line with other accounts that separate requirement and appraisal, including Graham (2010); Hawthorne and Srinivasan (2013); Lasonen-Aarnio (2014); Williamson (forthcoming; 2017).

[123] See also Lasonen-Aarnio (2014), who presses this objection to accounts of rational requirement that endorse this picture.

could be sufficient to outweigh all other possible rational requirements. So, if the agents in the puzzle can avoid the epistemic blame they would deserve for violating requirements of rationality, it is not because they are justified in violating those requirements.

Nor is it plausible that the agents in the puzzle deserve non-negative appraisal because they are exempt from the requirements. Being exempt means that one does not deserve blame for failing to do as the requirements of rationality require, because one is not subject to those requirements. Exemptions are traditionally associated with a lack of capacity to comply with requirements – someone in a coma might be granted an exemption for failing to meet a legal requirement to pay their taxes, on the grounds that they lack the capacity to do so. The requirements of rationality are universally binding, so apply to all agents with the relevant capacities for epistemic rationality[124], but we might think that agents who have significantly compromised rational capacities can be exempt from the demands of rationality, either temporarily or permanently. Having a mental illness or being under the influence of drugs are examples of the kinds of things that might lead to one's being exempt from the requirements of rationality. Exemptions such as these would be analogous to one's being exempt from keeping a promise to go to a football match because of a broken leg. However, being in a situation where one has rational support for a false belief about what rationality requires is not an incapacity that constitutes grounds for an exemption. The students in Logic 101 (Ch. 1, p. 12) are not incapacitated – quite the contrary. It is through the normal functioning of their intellectual capacities that they come to believe something false about what rationality requires. So, it

---

[124] I will not take a stand on what this category includes. Suffice to say that it will certainly include most adult humans, and perhaps also some children and animals.

is not plausible that they are blameless in virtue of being exempt from the requirements of rationality.

This leaves excuse as the only possible non-negative appraisal for agents in the puzzle. Being excused means that one does not do what one is required to do, but there is some consideration such that one does not deserve blame for this failing[125]. So, if the agents in the puzzle are excused from epistemic blame then there will be some consideration that provides the excuse. Solving the puzzle in this way demands an account of the considerations that can provide epistemic excuses for failures to meet requirements of rationality. The rest of this chapter discusses competing accounts of the considerations that can provide epistemic excuses, beginning with Williamson's account according to which norms governing excuses are derivative from the primary epistemic norms.

2. Williamson's Derivative Account of Excuse

Williamson (forthcoming) takes the notion of excuse to be governed by a set of norms that are derivative from a primary norm, where the violation of that primary norm would ordinarily incur blame. This section outlines the account and argues that it cannot offer an adequate account of when agents can be excused for failing to meet requirements of epistemic rationality: instead, an adequate account of excuse must be independent, rather than derivative, from the requirements of epistemic rationality.

According to Williamson agents can be excused for violating norms when they do what someone who had the general disposition to comply with the relevant norm would do in that particular situation. Importantly, in developing the

---

[125] For some examples of appeals to the notion of excuse in epistemology, see Hawthorne and Srinivasan (2013); Littlejohn (forthcoming); Sutton (2005, 2007), Williamson (forthcoming; 2017).

account Williamson has in mind a picture that involves only one, "truth-related" epistemic norm, such as the Knowledge Norm for belief (forthcoming: 12):

Knowledge Norm: Believe P iff you know that P.

Williamson's claim is that agents blamelessly violate a norm when they do what someone disposed to follow the norm would do. Doing what someone disposed to follow the norm would do is a sufficient condition for blamelessness. For example, this can grant brains in vats excuses for violating the knowledge norm, assuming that they manage their beliefs responsibly. Typically, ordinary non-envatted subjects who conform to the knowledge norm manage their beliefs responsibly, following their evidence and avoiding wishful thinking. The brain in the vat is blameless for forming non-knowledge beliefs when she also does these things, or is disposed to do these things.

However, while this may be a plausible enough explanation of why brains in vats might be excused for violating the knowledge norm, it does not generate the right results in cases of misleading evidence about what rationality requires. To see this, let's suppose that the primary norm in this case is the following:

No P Norm: Do not believe P.

Agents who have misleading evidence about whether or not one ought to believe P might come to violate this primary norm while managing their beliefs appropriately, by following the recommendations of false beliefs about whether rationality requires believing P[126]. As the previous chapter argued, when this is the case, agents deserve positive epistemic appraisal, despite violating the requirements of rationality (see "Diligent", Ch. 3, p. 95). However, Williamson's appeal to dispositions to conform to this primary norm, or to do what someone

---

[126] This is the situation with respect to Non-Contradiction in the Logic 101 case (p. 12).

who would conform to the norm would do, would not be helpful in explaining this. One reason for this is that it is not clear that agents who deserve non-negative appraisal do have a disposition to conform to the primary norm. Recall Diligent who believes, falsely, that rationality requires her to believe P, follows the recommendation of that belief, and so ends up believing P. If these agents do have a disposition to conform to No P Norm, then they are not manifesting it by believing P, and were they to manifest it, perhaps by believing not-P, then it is less clear that they would deserve non-negative appraisal. One way to manifest a disposition to conform to the No P Norm would be to disregard or fail to acquire misleading evidence that supports the belief that rationality sometimes requires believing P. However, as the previous chapter argued, agents who do this would be most likely to deserve negative appraisal – they disregard their evidence (see "Lazy", Ch. 3, p. 95). Given this, it is unclear why such a disposition should be relevant in an explanation of how agents can be excused from blame[127].

A further reason that the account is not helpful is that it considers only cases in which there is only one primary norm, from which the norms governing excuse are derived. However, the puzzle is motivated by cases in which there is conflict between more than one norm with equal claims to primacy. In Logic 101, for example, the agents are subject both to a requirement not to believe contradictions, and a requirement to believe what their evidence supports.

Non-Contradiction Norm: Do not believe both P and ~P.

---

[127] There is much more to be said about what it is to manifest a disposition, and which dispositions could be compatible with the epistemic states of agents such as Diligent and Lazy (p. 95). However, since there is a further problem with Williamson's account, I leave this aside.

> *Evidential Norm*: If you have evidence that supports P, believe
> P[128].

Agents who follow the recommendations of their false beliefs about what rationality requires do what someone disposed to conform to the Evidential Norm, but not what someone disposed to follow the Non-Contradiction norm would do. It is unclear what Williamson's account should say when there is more than one primary norm, and the agent is disposed to follow one but not all of these (or does what someone who was disposed to follow one but not all of these would do). This unclarity means that the account is not particularly helpful in solving the puzzle.

One way that one might attempt to apply Williamson's account to cases in which agents are subject to more than one primary norm would be to say that agents are excused when they do what someone disposed to conform to all of the primary norms would do. So, in Logic 101, agents would need to be disposed to conform to *both* the Non-Contradiction and the Evidential norms in order to be excused. This is unhelpful for solving the puzzle, since it is not at all clear what someone disposed to conform to *both* the Non-Contradiction and Evidential norms would do. Furthermore, that this is unclear should not be surprising since this is just a redescription of the original conflict that generated the puzzle. The puzzle began by asking what it would be rational for someone to do, given that rationality requires us to both believe what is supported by the evidence, and to refrain from having prohibited attitudes, such as contradictory belief. So, Williamson's account is unhelpful as a way of explaining why agents in the puzzle are excused, and the problem is that it derives the norms of excuse from the requirements of epistemic rationality. This suggests that what is needed is an

---

128 This is a simplification of Evidentialism (p. 13). As the puzzle is initially stated, the Enkratic Principle could generate a further potential primary norm. I leave this complication aside here.

account of appraisal that gives independent conditions of excusability. The following sections discuss accounts that purport to offer this, beginning with an account based on good habits.

3. Good Epistemic Habits

One consideration that might be thought to ground epistemic excuses is that the agent has employed good epistemic habits in forming her beliefs[129]. Typically, habits understood to be epistemically good when they, in general, lead to epistemic goods such as knowledge and bad when they, in general, lead to failure to achieve these epistemic goods. For example, Hawthorne and Srinivasan identify being epistemically blameworthy as "exhibiting poor habits of mind" (Hawthorne and Srinivasan 2013: 18). Similarly, Lasonen-Aarnio identifies a category of reasonable epistemic subjects, who "manage their beliefs in a way that makes sense given the goal of knowledge acquisition" (Lasonen-Aarnio, 2010: 12). These are distinct from those who actually succeed in knowing, and so we might think of them as the agents who deserve epistemic excuses.

However, we might be suspicious of this appeal to habits on the grounds that excusability and whether or not an agent employs a good habit can come apart, since agents are not always epistemically blameworthy when they exhibit bad epistemic habits. Determining whether an agent who exhibits a bad habit seems to depend, at least to some extent, on why they exhibit that habit[130]. Some agents seem excusable for their bad epistemic habits. For example, agents who are the victims of cults are often manipulated in such a way that they end up relying on poor epistemic methods – placing undue trust in authority figures or particular texts, for example – but in such a way that it would be very difficult for anyone

---

[129] For views that take epistemic blamelessness to depend on the agent's employing good epistemic habits, see Lasonen-Aarnio (2010; 2014) Alston (1989); Hawthorne and Srinivasan (2013).

[130] See also Watson (1996), who articulates a more general, non-epistemic, version of this point.

in their position to avoid doing so. Similarly, most ordinary human agents do not exhibit good habits of reasoning when thinking about probability, due to reliance on hard-wired cognitive heuristics and biases. However, it is less clear that we are epistemically blameworthy for this. Since these bad reasoning habits are the result of our evolutionary inheritance, it would seem unreasonable to blame us for failing to avoid them[131]. Furthermore, some agents do not seem excusable despite their exhibiting good epistemic habits. For example, consider someone who has a clairvoyant ability, but no reason to believe that she has such an ability. Although it would be a good method for reaching the truth, we might think that would be epistemically reckless to rely on this ability without further reason to think that it is reliable[132].

In summary, it seems that the epistemic habits the agent exhibits cannot serve as considerations that grounds an agent's having an epistemic excuse. This is because exhibiting good habits and being blameless do not always align. Agents can sometimes be blameless despite employing bad habits, and blameworthy despite employing good habits. This suggests that more specificity is needed in an account of blameless epistemic conduct. The following section discusses Rosen's account of blameless false normative belief, which specifies in more detail the conditions under which false normative belief can excuse, by appealing to obligations to avoid epistemically irresponsible belief.

---

[131] See the debate over what evidence of widespread error in human reasoning about logic and probability should mean for our theories of rationality Cohen (1981); Hawthorne and Bovens (1999); Kahneman, et al. (1982); Tversky and Kahneman (1993).
[132] For further discussion of clairvoyant cases, see Bonjour (1998).

4. The Demands of Responsible Belief Formation

Rosen (2002, 2004)[133] understands (morally) blameless belief formation to require compliance with moral obligations to avoid epistemically irresponsible belief formation[134]. When agents comply with these obligations in forming their beliefs[135], they can be excused for any wrongdoing they do as a result of those beliefs, and this applies to both normative and non-normative beliefs. So, compliance with these obligations to avoid irresponsible belief can provide a consideration that can excuse agents from blame. This section argues that while Rosen's account of the when normative belief can excuse is along the right lines, there are some important objections to his understanding of what epistemically responsible belief formation requires.

Rosen understands responsible belief formation to consist in the following specific obligations on belief formation:

> (a) That agents take steps to "inform [them]selves about matters
>
> relevant to the moral permissibility of [their] conduct" (2002: 63).
>
> (b) That agents "reflect" to the degree deemed appropriate by the
>
> situation (2002: 65).

Exactly what each of (a) and (b) demands in practice will depend on the particular situation. Complying with (a) requires agents to take certain actions towards managing their beliefs, for example, "to ask certain questions, to take careful notes, to stop and think, to focus one's attention in a certain direction". Complying with (b) requires agents to be as reflective as the situation demands.

---

[133] Additionally, Rosen holds that, in fact, this excuse holds in the vast majority of cases of wrongdoing (2002: 62). His commitment to this claim depends on an independent commitment to the Belief Condition, which I will argue against in Chapter 6. This aspect of Rosen's view is orthogonal to this discussion of which epistemic obligations agents are subject to.

[134] Rosen refers to these as 'epistemic obligations', but as he clarifies, these are really moral obligations that pertain to belief formation (2002: 63 fn.3).

[135] Although Rosen's view is concerned with moral blame, it is also applicable to epistemic blame. Here, I will be concerned only with the possible application of the view to epistemic blame.

Rosen takes some situations to demand more thought, curiosity, and reflection than others: for example, hard moral cases require more reflection than easier ones, and more reflection is demanded in response to "serious criticism", "known diversity of opinion", or "perceived tension in one's moral view" (2002: 65).

However, the demands of epistemically responsible belief formation are purely procedural. Epistemic responsibility, according to Rosen, does not require that we come to any particular belief. Instead, it is "procedural", and does not involve obligations to "know or believe this or that" (2004: 301). Agents who (a) take steps to inform themselves about morally relevant features of the situation, and (b) are as reflective as the situation demands, have met the obligations to be epistemically responsible. This means that they are not to blame if they do wrong as a result of their beliefs, regardless of what they end up believing. Epistemically responsible belief does not require or forbid particular beliefs. This core aspect of Rosen's view can be expressed in the following claim:

> *Procedural Claim*: Epistemic Responsibility does not require or forbid any particular belief.

Furthermore, Rosen understands the obligations on responsible belief formation to always permit us to believe what seems obvious to us, so long as these beliefs do not meet with any friction. That is, so long as it does not seem to us that we have any reason to think these beliefs are mistaken. As he puts it: "[W]hen what one takes to be a transparently correct moral verdict meets with no such friction, one is neither negligent nor reckless in failing to subject that verdict to special scrutiny" (2002: 65)[136]. In other words, Rosen is committed to the view that

---

[136] Rosen does not spell out exactly what constitutes 'friction'. His examples are generally considerations that he supposes would cause most ordinary people to doubt that beliefs are

responsible belief formation always permits agents to believe what seems obvious+ to them (where obvious+ is what seems obvious and for which there is no friction). Obviousness+ is sufficient to guarantee a permission to believe:

> *Obviousness Sufficiency Claim*: P is obvious+ for S   S is permitted
>
> to believe P[137].

However, I will argue that both of these core claims of Rosen's view are false. The Procedural Claim is false – sometimes responsible belief formation does require and forbid particular beliefs; and the Obviousness Sufficiency Claim is false – that P is obvious+ for S is not sufficient to guarantee that S is permitted to believe P.

A first problem with Rosen's position is that the Procedural Claim is false. That it is false can be seen by considering cases in which the agent has evidence that we can reasonably expect them to understand undermines their belief, even though that belief is obvious to them and meets with no friction. Part of what it is to believe in a way that is epistemically responsible is to avoid holding beliefs that are undermined by one's evidence, at least when it is reasonable to expect you to understand that these beliefs are undermined by this evidence. This indicates that the Procedural Claim is false – sometimes believing in a way that

---

correct, such as disagreement from those one takes seriously, and lack of support from one's moral culture.

[137] The view that responsible belief formation always includes a permission to believe what is obvious+ also finds some support in epistemology (see Alston (1989); Cohen (1984); Lewis (2004); Carroll (1895); Huemer (2007); Bonjour (1985; 1998). Obviousness, for the purposes of this discussion, is to be understood as a three-place relation between an individual, a proposition, and a time. S finds p obvious at T if and only if, at T, P seems true to S. This understanding is to be distinguished from more restrictive accounts of obviousness. For example, those on which to find P obvious is for P to seem true on the basis only of one's conceptual understanding of the concepts involved in p (see Jeshion (2000: 345)), or 'objective' notions of obviousness, according to which propositions can be obvious, or not, independently of whether or not anyone finds them obvious (Dummett (1974); Lewis (2004).

is epistemically responsible means avoiding particular beliefs when they are undermined by one's evidence.

Endorsing this view of what it is to believe in a way that is epistemically responsible would mean that agents would give a different verdict about whether or not agents can be excused for wrongdoing in some of the cases that Rosen discusses. For example:

> *Sexism.* Smith is a run-of-the mill American sexist circa (say) 1952. Like any decent middle class father he has encouraged his sons to go on to college, setting aside money for the purpose. But like any run-of-the-mill sexist he has done nothing comparable for his daughters. This differential treatment is not malicious. But it is unfair and therefore wrong. But of course Smith doesn't know this. He doesn't know that his daughters deserve equal consideration in this respect (2002: 66-69).

As Rosen describes him, Smith "believes what he believes because he finds it obvious, and […] he finds it obvious because he was raised to find it obvious and because the people he takes seriously find it obvious". Furthermore, while he knows that there are some people who disagree with his point of view (feminism predates 1950, after all), speaking to them does not change his mind in any way. Rosen's reading of this case is that Smith is not to be blamed for his sexist actions because he has met the obligation to form his beliefs in an epistemically responsible way – he has fulfilled the procedural requirements to (a) take steps to inform himself of the morally relevant features of the situation, and (b) reflectively considered the matter to an adequate degree. Rosen cashes out what it is to have reflected sufficiently in terms of time and mental energy: expending enough time and mental energy in forming one's beliefs is sufficient to make it

the case that one *has* reflected sufficiently and therefore holds one's beliefs in a way that is epistemically responsible.

However, if responsible belief formation requires us to sometimes avoid holding beliefs that are undermined by further evidence we have, then it is less clear that Smith's beliefs are held responsibly. Smith does have some evidence that counts against his views: for example, the testimony of those who disagree with him. When Smith consider this testimony, he does not consider it evidence against his sexist beliefs. As Rosen puts it, "what he confronts when he reflects is a sensibility very different from his own, a sensibility that strikes him as wrong-headed" (2002: 67). While this is an explanation of why Smith responds inappropriately to his testimonial evidence, it neither establishes that he lacks evidence against his views, nor that he lacks the capacity to realise that if the feminists are right, then what they say would constitute evidence that undermines his beliefs. So, it is not clear that Rosen is correct in holding that Smith's beliefs are epistemically responsible.

Additionally, it is plausible that Smith has further evidence against his sexist beliefs acquired via his emotional and psychological faculties[138]. For example, Smith lives in a world where he interacts frequently with women, and has daughters whose well-being he cares about. It is not clear that any complicated moral reasoning is required in order to establish that the women he interacts with are of equal moral status to their male counterparts, so it is not clear that he lacks the capacity to appreciate this evidence or that it undermines his views[139]. While the fact that Smith's sexist beliefs are part of a worldview that seems obvious to

---

[138] Buss (2014) and Railton (2014) describe how agents might use these to come to appreciate moral reasons.

[139] We need not think that having false moral beliefs that are obvious+ is always a barrier to doing what is in fact morally required. As others have pointed out, agents can sometimes act akratically and do the right thing for normative moral reasons (see Arpaly (2002b); Buss (2014); Greenspan (2016); Railton (2014)).

him and meets with no friction[140] may make it more difficult to take his evidence seriously in the way that epistemic responsibility demands, this is not beyond his capacities. It not clear that it would be unreasonable to expect him to notice that he has evidence that undermines his beliefs. In so far as we think that epistemic responsibility requires us to avoid holding beliefs that are undermined by our evidence, we should endorse a different conclusion about Smith to the one that Rosen endorses. The conclusion we should endorse is that when Smith retains his sexist beliefs he is failing to believe in a way that is epistemically responsible. Thus a first reason to reject Rosen's position is that it fails to accommodate an important aspect of what epistemic responsibility demands, namely that it sometimes requires us to avoid holding beliefs when we have evidence that undermines those beliefs.

A second problem with Rosen's position is that the Obviousness Sufficiency Claim (OSC) is false. The OSC says that the fact that P is obvious+ for S is sufficient to guarantee that S is permitted to believe P[141]. However, the OSC is incompatible with the plausible claim that epistemically responsible belief requires that we believe what is supported by evidence that it is within our capacity to recognise. This means that the demands of epistemic responsibility will not always permit us to believe what is obvious+. Specifically, it will not permit us to hold beliefs that are obvious+ for us when we can be expected to appreciate that we have evidence that undermines them. This is shown by the fact that the failure of the following bi-conditional fails in both directions, and it is reasonable to expect most epistemically responsible agents to be aware of this failure:

---

[140] Assuming this is true. In fact, if he has interacted with feminists who disagree with him, it is not clear that his beliefs are as frictionless as Rosen interprets them as being.

[141] Recall: *Obviousness Sufficiency Claim*: P is obvious+ for S   S is permitted to believe P.

Obviousness Bi-conditional: P is obvious+ for S   S's evidence

supports that P.

The failure of this bi-conditional can be noticed, since in various domains, it is

clear that obviousness, even in the absence of friction, is neither an indication of

what the evidence supports, nor itself evidence of truth. I consider first the left to

right direction:

D1: P is obvious+ for S   S's evidence supports that P.

If D1 were true, then every case in which P is obvious+ would be a case in which

P is supported by evidence. However, it is clear that this is not true. Within some

domains we have good reasons to think that what is obvious+ is not supported

by the evidence. One illustration of this is our beliefs about morality.

Consideration of historical moral beliefs about slavery or the rights of women

shows that many adequately reflective and very well-informed people have

managed to find various things obvious+ despite the evidence failing to support

them[142]. Another example is judgments about probability. It strikes many people

as obvious+ that under certain descriptions a conjunction is more likely than a

single event (see Tversky & Kahneman (1974)), but more precise consideration of

probability shows this to be a fallacy[143]. So, at least in these domains, that

something is obvious+ does not mean it is supported by the evidence – D1 fails,

and in general it would be reasonable to expect most agents who are otherwise

---

[142] See Anderson's discussions of historical philosophical defences of slavery, many of these written by some of the most well-educated people of the day (see Anderson (2014, 2015, 2016)). We might also think that implicit bias is an instance of obvious+ness, although this is less clear since implicit beliefs are typically not noticed at all by the believer. See Holroyd (2012); Robin (2016) for discussions of whether implicit bias is morally blameworthy.

[143] For further examples of the unreliability of obvious+ness in moral thinking, see Anderson (2014, 2015, 2016); for more examples of the role of obvious+ness in fallacious reasoning Kahneman et al. (1982); Kahneman (2011).

competent in managing their beliefs to realise that it fails. The bi-conditional also fails in the opposite direction:

D2: S's evidence supports P   P is obvious+ for S

If D2 were true, then every case in which P is supported by the evidence would be a case in which P is obvious+. It is clear this is not the case. The success of detective dramas is sufficient to show this – Miss Marple figuring out whether the butler did it takes time and is often rich with dramatic twists and turns that would be extremely dull if D2 were true. Such cases are examples of how P is supported by the evidence, but is not obvious+. That D2 fails is also shown by more complicated examples in which our evidence supports P, but also that it is unlikely that our evidence supports P (Christensen, 2010; Cohen & Comesaña, 2013; Elga, 2013; Lasonen-Aarnio, 2014; Weatherson, forthcoming). The failure of both directions of the bi-conditional shows that there is no connection between what is obvious+ and what is supported by the evidence.

Having established the lack of connection between obvious+ness and what our evidence supports, we might nevertheless wonder how to accommodate cases in which we are permitted to believe what is obvious+. Rejecting the OSC would mean that when S is permitted to believe P, and P is obvious+ for S, there is necessarily some other consideration that grounds this permission. So, if we reject the OSC, then we need an account of how this permission is grounded in the cases for which one is permitted to believe what is obvious+. For example, very simple arithmetical propositions seem obvious+ to most of us, and we are indeed permitted to believe that they are true. However, we need not think this obvious+ness is itself evidence that grounds this permission, since we have various other reasons to believe simple arithmetical propositions such as 2+2=4. For example, we can check by counting, it is supported by our further understanding of numbers, and we have reason to trust our mathematical

reasoning abilities. Although obviousness+ typically coincides with correct mathematical reasoning, his does not mean that it itself is evidence of mathematical truth. When we are permitted to believe what is obvious+, this is because of some further consideration[144]. In this case it because we have reason to trust our mathematical reasoning methods, and have calibrated them by checking.

In summary, we should reject both the Procedural Claim and the OSC, rendering Rosen's account of the demands of epistemic responsibility incorrect. Epistemic responsibility sometimes requires agents to avoid particular beliefs, and it does not always permit us to believe what is obvious+. However, this is compatible with Rosen's more general claim that agents can be excused for wrongdoing that is a result of false normative belief when they have complied with these obligations to avoid epistemic irresponsibility in forming those beliefs. It is only Rosen's understanding of what epistemic responsibility demands that is in need of revision. In stating the considerations that can excuse agents who do wrong due to false normative belief from wrongdoing, we should amend Rosen's account so that it more accurately captures the demands of epistemic responsibility.

One way that we might be tempted to amend Rosen's account is by including a requirement to believe what one's evidence supports as a further condition that agents must meet in order to count as believing in a way that is epistemically responsible. An amended account such as this would say that epistemic responsibility, and therefore eligibility for an excuse, requires not only that we (a) be reflective and (b) take steps to inform ourselves, but also that we believe

---

[144] One might nevertheless think that when one has really nothing else to go on, other than obvious+ness, then obvious+ness is what grounds a permission to believe. If so, this is plausible for only a very restricted range of fairly strange cases. Accommodating this does not threaten the overall view that in ordinary cases, obvious+ness is not to be relied upon.

what our evidence supports. However, this would not give the right verdict for all cases. Consider the following example, in which the agent's failure to believe what the evidence supports does not seem to be something for which he deserves epistemic blame:

> *Detectives.* Holmes and Watson have just arrived at the scene of a crime, and are examining the evidence. The evidence supports that Mary is the thief, but no one has yet worked this out. The police have already spent many days examining the evidence without reaching a conclusion. Suddenly, Holmes exclaims, "Aha! It was Mary". Holmes has exceptional powers for working out what evidence supports that surpass the capacities of most ordinary detectives. Watson nods, but does not see how the evidence supports this.

Although both Holmes' and Watson's evidence supports that Mary was the thief, it is implausible that Watson deserves epistemic blame for failing to work this out. After all, it is a complicated matter and a whole team of police detectives has already failed to do so. Unlike Holmes, Watson has merely ordinary reasoning capacities. He does not seem to deserve blame for failing to believe what his evidence supports because it seems unreasonable to expect him to work out what his evidence supports – he lacks Holmes' expert reasoning capacities in this area. Given this, a requirement to believe what one's evidence supports is implausible as a criterion of epistemic responsibility.

Such a requirement would also be implausible in cases in which one's evidence is misleading about itself[145]. For example:

---

[145] See, for example, Elga (2013); Lasonen-Aarnio (2014); Williamson (2011; 2014); Christensen (2007; 2010a; 2010b).

Pilot. You're alone, flying a small plane to a wilderness airstrip. You're considering whether you have enough fuel to make it safely to an airstrip 50 miles further away than your original destination. You make some fuel calculations, and you become extremely confident that you do have plenty of fuel. In fact, you do have enough fuel, and you have performed the calculations correctly. But then you notice that your altimeter indicates that you are at an altitude in which it is extremely likely that you are suffering from hypoxia. When suffering from hypoxia, victims' judgment is severely compromised, making them highly likely – but not guaranteed – to go wrong in simple calculations without being able to notice their errors (Christensen 2010b: 126).

In this case, and cases of a similar structure, the agent's evidence is such that some P is both likely on her evidence, and unlikely to be likely on her evidence. In Pilot, your evidence about how much fuel you have supports the belief that you have enough fuel, but your evidence about how likely your calculations are to be correct suggests that your calculations are very unlikely to be correct. So, if P is the proposition "I have enough fuel to reach the airstrip", then your evidence supports the following combination of beliefs:

P

P is very unlikely on my evidence.

However, this pattern of evidential support is compatible with its being inappropriate to blame agents for failing to adopt this belief combination. As in Detectives, a natural way to explain why blame would be inappropriate is by appeal to what it is reasonable and unreasonable to expect of agents given their capacities. For example, some argue that such belief combinations are impossible without psychological division (Greco 2014), and we might think that it is

unreasonable to blame agents for failing to do the impossible[146]. Others argue that if level incoherent belief is possible, then it constitutes Moorean absurdity (Smithies (2012)), and it is unclear whether we have the capacity to sincerely believe Moorean absurdities.

Most importantly, adopting level incoherent belief combinations would very often be epistemically irresponsible. This suggests that not only are such combinations not always excusable, they may themselves deserve epistemic blame. For example, some argue that it would involve irresponsibly ignoring epistemic pressure to revise either the first or the higher order belief (Lasonen-Aarnio, forthcoming; Littlejohn, 2015)[147]. If this is right, then believing what the evidence supports is not always a mark of epistemic responsibility, and is sometimes the very opposite. This means it cannot itself be a requirement of responsible belief formation. However, we can explain all this by appeal to what it is reasonable to expect of agents – it is reasonable to expect agents to respond to epistemic pressure to revise either the first or higher order belief in an akratic belief set, and so the failure to do this is something that would deserve epistemic blame in the absence of a justification, excuse, or exemption.

With this in mind, it is worth considering an alternative proposal for establishing considerations that can ground epistemic excuses: one based on the notion of what it is reasonable to expect of agents. It is this proposal that I ultimately defend as part of a solution to the Puzzle of Rational Requirement. This approach avoids the shortcomings of Rosen's account, without encountering the problems

---

[146] See also the arguments of Adler (2002); Owens (2002), who argue that epistemic akrasia is impossible, but it is not clear that they understand epistemic akrasia and level incoherent belief to be the same thing. Recall Chapter 3 (p. 72-73).

[147] See also Horowitz (2014), who argues that akratic belief combinations involve taking epistemically irresponsible attitudes towards the evidence, since the higher order evidence recommends believing the negation of what the first order evidence recommends.

faced by an amended version of Rosen's account that includes a requirement to believe what the evidence supports. The following section outlines this approach.

5. Reasonable Expectations

This section argues that agents can be excused from wrongdoing when they have done what it is reasonable to expect of them, given their volitional and epistemic capacities and their roles.

On this view, agents are blameworthy only when they fail to do what it was reasonable to expect of them. When agents have done what it is reasonable to expect of them, but nevertheless failed to do what is required, they can be excused from the blame that this failure would ordinarily incur. Accounts of blamelessness based on what it is reasonable to expect can be found in both the moral domain (FitzPatrick (2008); Goldberg (2017; 2018); Nelkin (2009); Rosen (2002; 2004); Sher (2009)), and the epistemic domain, where blamelessness (in the sense of either justification or excuse) is often linked to reasonableness (see Alston (1989), Cohen (1984), Fantl and McGrath (2009), Williamson (forthcoming)). This approach also preserves the key features of Rosen's view. On his view, agents can be excused for wrongdoing that is the result of false normative belief, provided that they have met the obligation to form those beliefs responsibly[148], because it is unreasonable (in the sense of being unfair) to expect agents to refrain from doing what they believe is permissible when they believe this responsibly (see Rosen (2002: 74)). So, what is doing the work in Rosen's account of how false normative belief can excuse is the notion of what it is reasonable to expect. Obligations to believe responsibly are only relevant because they can help determine whether or not the agent has met the expectations it is

---

[148] As Rosen understands this notion (see p. 102).

reasonable to have of her. This indicates that having done what it is reasonable to expect is worth exploring as a consideration that can ground excuses.

If we take seriously the idea that blame is appropriate only when agents have failed to do what it is reasonable to expect them to do, then determining when agents can be excused requires determining what it is reasonable to expect of agents.

The concept of reasonableness used here is to be distinguished from the legal use of the concept of the reasonable person. This concept is often used to determine legal culpability for wrongdoing (see Baron (2011), Moran (2010)), and has been borrowed by other accounts of blamelessness based on meeting reasonable expectations (Sher (2009)). However, there are two important differences to note between that concept and the one used here. Firstly, what it is reasonable to expect of agents is determined by an agent's mental as well as her physical capacities. Orthodoxy in legal philosophy has generally been reluctant to take an agent's mental capacities into consideration when attributing legal culpability, using instead a concept of 'the reasonable person' that assumes the possession of various particular mental capacities[149]. In contrast, the view presented here takes into account the agent's epistemic capacities, which are primarily mental, and does not assume that these are the same for all agents. This makes the account more applicable to epistemic appraisal, and allows it to accommodate a wide range of agents. Additionally, the concept used here should be distinguished from its usage in connection with what it is common or typical for people of a particular society or social group to believe. The legal concept has sometimes been used in this way to argue for the excusability of discriminatory but

---

[149] For example, *Vaughan vs. Menhove,* in which it was deemed impermissible to take into account the defendant's low intelligence (Moran (2010: 1238)).

commonplace beliefs and actions[150]. However, for the same kinds of considerations that led us to reject the view that responsible belief formation does not always permit one to believe what is obvious+, the view presented here sometimes takes it to be reasonable to expect agents to reject what is commonly believed. For example, on this view it is usually reasonable to expect agents to be responsive to epistemic considerations that do not support what is commonly believed.

In determining what it is reasonable to expect of agents, it is useful to think of agents as being subject to two kinds of expectations: volitional expectations, pertaining to their efforts to control their actions, and epistemic expectations, pertaining to their efforts to believe in an epistemically responsible way. For both of these sets of expectations, what it is reasonable to expect of the agent is determined jointly by her capacities and her role. Generally, the less capacity an agent has, the less is expected of her – but particular roles she might occupy can alter what is expected, and thereby alter what she must do to avoid blame. What it is reasonable to expect of agents constitutes a minimal standard that agents are to meet in order to avoid blame. When agents fail to do what it is reasonable to expect of them, they are blameworthy to the degree that they fall short of the minimal standard set by their reasonable expectations. The greater the failure to do what can reasonably be expected, the more blameworthy the agent is.

For example, it is not reasonable to expect agents who have very limited powers of motor control to refrain from stepping on toes, and it is not reasonable to expect severe kleptomaniacs to refrain from stealing. These are volitional expectations. Reasonable epistemic expectations imply that it is not reasonable to expect colour-blind people to know what colour the traffic lights are and so know

---

[150] For example, to defend white men for attacking black men whom they (wrongly) suspected of posing a threat, or to defend perpetrators of sexual harassment Baron (2011), Moran (2010).

on the basis of traffic light colours that they should stop – although it would be reasonable for them to know where the light meaning 'stop' is normally located in the traffic light system (i.e. it is the uppermost light in the UK). Either kind of expectation can be affected by the agent's role(s). Parents can be expected to take more care than others in knowing the whereabouts of their children, and diplomats can be expected to take greater care in their interpersonal interactions with foreign dignitaries. This is true even if the agents who occupy these roles have limited volitional or epistemic capacities to fulfil the requirements of that role. For example, diplomats are required to remember codes of etiquette for countries that they interact diplomatically with, and this is true even if their epistemic capacity to know about these codes is limited by their being especially forgetful. Diplomats who are especially forgetful are required to either employ strategies to help them fulfil the requirements of their role, or refrain from occupying that role[151].

Usually, meeting one's reasonable epistemic expectations requires one to believe what one's evidence supports. For example, if you are looking at a red wall under normal conditions and you have no visual or mental impairments, it is reasonable to expect you to believe that the wall is red; you have evidence that it is reasonable to expect you to respond to in this way. However, conditions are not always normal – you might be blind, or distracted with more important matters. This affects the extent to which it is reasonable to expect you to respond to your evidence. For example, if you are blind, and so lack the capacity to see the red wall, then it is not reasonable to expect you to believe that the wall is red, because

---

[151] This raises the question of what kinds of things can be roles, in the sense used here. On the one hand, fairness seems to demand that roles be optional, such that the agent can opt out rather than incurring the burden of fulfilling the requirement. On the other hand, some roles that seem legitimate are not always optional in a meaningful sense. Being a parent is one example – while it is often (but not always) an option that one takes voluntarily, once one has become a parent it is not a role one can opt out of. While important, this is not a question I engage with directly here.

you cannot easily access the evidence that would support this belief in the way that a sighted person can. However, if you cannot access this evidence because your eyes are closed, it is reasonable to expect you to open them and access it. So, in this case it would be reasonable to expect you to believe that the wall is red. Your roles also contribute to determining what it is reasonable to expect. If you are employed as an interior designer, and the red wall is in the home of one of your clients, then it is reasonable to expect you to make an effort to know what colour the wall is, even if at this moment you do not have evidence that supports believing it is red. You should, for example, make the effort to visit the home of your client and acquire evidence about the colours of her walls. It is reasonable to expect this of you even if you are blind, and lack the capacity to learn about the colours of walls in the ordinary way. For example, you could be expected to use a code mapping colours to textures, or rely on a sighted friend's testimony.

This account allows us to explain how agents can be excused for failing to do what rationality requires, thus helping us to flesh out the solution to the puzzle offered in the previous chapter. The following paragraphs illustrate how the reasonable expectations view presented here evaluates agents in the puzzle. Recall the agents Diligent, Lazy, and Incoherent (Ch. 3, p. 95). All three are beginner students taking the misleading logic class, Logic 101. All have misleading evidence that supports a false belief about what rationality requires. They arrive at the following distinct belief combinations:

Diligent's Beliefs	*Lazy's Beliefs*	*Incoherent's Beliefs*
I am rationally required to believe P	Rationality requires *not* believing P	Rationality requires *not* believing P
P	~B(P)	P

As the previous chapter told it, Diligent has considered her evidence seriously, arrived at the false belief that rationality requires believing P, and then arrived at the first order belief P by following the recommendations of her false belief about what rationality requires. Lazy, in contrast, disregards the misleading evidence about what rationality requires because what it supports is intuitively implausible. He does not carefully consider his evidence, and reasons from the intuitive implausibility of P to the belief that it is not the case that rationality requires P. Incoherent, let us suppose, has arrived at this belief combination by failing to fully consider the implications of her beliefs about what rationality requires. Given these backstories, the previous chapter suggested that Diligent deserves non-negative appraisal, Lazy deserves negative appraisal, and Incoherent deserves negative appraisal, but to a slightly lesser degree than Lazy. Using the reasonable expectations account of blamelessness, we can now explain why these agents deserve these particular appraisals.

Diligent can be excused from the blame for failing to meet the requirements of rationality because she does what it is reasonable to expect of her given her intellectual capacities and her role as a beginning philosophy student. The fact that she is a non-ideal reasoner studying philosophy for the first time means that her *capacity* for philosophical reasoning is limited. This makes it unreasonable to expect her to work out why it is false that rationality requires P using independent reasoning – although this would perhaps not be an unreasonable expectation of ideal beings, or professors. Diligent's *role* as a beginning student makes it reasonable to expect her to pay attention and complete the homework assignments, but not spend hours in the library researching the finer points of what rationality requires; though this might be a reasonable expectation to have of the professor teaching the class. In so far as Diligent meets the expectations it

is reasonable to have of her, she does not deserve blame for violating the requirements of rationality by believing P.

In contrast, Lazy fails to do what it is reasonable to expect of him, and so cannot be excused. He fails to believe what the evidence supports, even though he has the *capacity* to do so, and he has no good reason for this failure – he just finds the conclusion counterintuitive. He also fails to meet the expectations that it is reasonable to have of him given his *role* as a student in the class. He does not pay attention, and he does not complete the homework assignments. So, he fails to have evidence that it is reasonable to expect him to have. Since he fails to do what it is reasonable to expect of him, he is not excused from blame for failing to do what is required and believe what his evidence supports about what rationality requires. Furthermore, while he has the attitude required of him at the first order, he fails to do what it is reasonable to expect him to do in forming this attitude, and so nevertheless deserves blame. This is a case of blameworthy compliance with the requirements. Incoherent has the attitudes that rationality requires of her (according to the arguments of the previous chapter), but fails to do what it is reasonable to expect of her. This means that she is blameworthy, but to a lesser extent than Lazy. Although she does pay attention in class, and form the beliefs that her evidence supports, she fails to make a relatively simple transition that it is reasonable to expect her to make, since it is well within her capacity. Furthermore, she fails to notice or respond to any tension between her first and higher order beliefs, which we also suppose she has the capacity to do. Her failure to meet the expectations it is reasonable to have of her means that she is blameworthy. This account of excusability as doing what it is reasonable to expect allows us to explain and justify the distinct epistemic appraisal of agents such as Diligent, Lazy, and Incoherent, lending support to the solution to the puzzle offered in the previous chapter.

6. Summary

This chapter has argued for an account of excusability according to which agents can be excused from blame if they have done what it is reasonable to expect of them. This enables the distinction between requirement and appraisal, introduced in the previous chapter to solve the puzzle to be drawn in a principled way, providing a consideration that determines when agents can be excused from epistemic blame. This account is an improvement on Williamson's dispositional account, the habits based accounts, and Rosen's account of excuse, and it allows false normative belief to excuse agents from blame – for example, in the Puzzle of Rational Requirement. On this view, having misleading evidence about what rationality requires is one example of when false normative belief can excuse, so long as it is the case that the agent has done what it is reasonable to expect of her. However, this result is controversial. In Part 3, I address two alternative views on the question of whether false normative belief can excuse: the view that false normative belief can *never* excuse, and the view that it *always* does (at least when the belief is itself blameless). Both of these positions disagree with the view defended here – that false normative belief can sometimes, but not always, excuse. As such, both kinds of view constitute important objections to this account. Part 3 rejects both of these views, and defend the view that false normative belief can (only) sometimes excuse. The following chapter deals with the view that false normative belief can never excuse, and the final chapter deals with the view that false normative belief always excuses.

Part 3

Chapter 5

False Normative Belief Sometimes Excuses

This part discusses whether false normative belief can excuse agents who would ordinarily deserve blame for failing to do what is required. This chapter considers and rejects the view that false normative belief can never excuse, and the following chapter considers and rejects the view that false normative belief always excuses (at least when it is itself blameless). In particular, this chapter considers and rejects an influential argument for the view that false normative belief can never excuse, the 'De Re Argument'. The De Re Argument is based on an appeal to the putative significance of responsiveness to a particular kind of reason, normative reasons de re. I argue here that, in fact, normative reasons de re do not have any particular significance for appraisal, and this is true even if we accept the Strawsonian account of appraisal typically used to explain their significance.

Section 1 outlines the De Re Argument for the view that false normative belief can never excuse. This argument relies on a claim about the significance of normative reasons de re for appraisal – the Blame Claim – and typically supports that claim by reference to a Strawsonian account of appraisal, according to which appraisal is determined by the agent's quality of will. Section 2 provides one reason to reject the Blame Claim, namely that it implies, implausibly, that factual and normative uncertainty have asymmetrical consequences for appraisal. Furthermore, as Section 3 argues, there are counterexamples to the Blame Claim – that is, possible cases in which some agents systematically fail to respond to normative reasons de re, but should not be evaluated as blameworthy, even if we

assume the Strawsonian account of appraisal typically used to motivate the Blame Claim. Section 4 argues that if that is right then there are cases in which false normative belief excuses the wrongdoing of some agents who fail to respond to normative reasons de re. So, false normative believe can sometimes excuse.

The arguments of Part 3 concern moral blameworthiness, since this is where the debate has been focussed. It is a further question how transferable the arguments discussed here are to epistemic blameworthiness, one that I do not engage with here. If they are not transferable, then the views discussed in Part 3 present no threat to the solution to the puzzle presented in Part 2. For the sake of argument, I assume here that the views discussed in Part 3 are transferable in some way to the epistemic domain, and so do constitute a serious objection to the proposed solution. The arguments of Part 3 therefore serve to dismiss this potential objection.

1. The De Re Argument

This section presents an influential argument for the view that false normative belief can never excuse (Arpaly (2002b); Harman (2011); Weatherson (2019)). Crucially, this argument depends on a claim – the Blame Claim – about the significance of normative reasons de re for appraisal. The Blame Claim is typically justified by appeal to a Strawsonian account of appraisal, according to which agents are blameworthy when they exhibit deficient quality of will. The following sections argue that both the Blame Claim and the Strawsonian account of appraisal should be rejected, thus undermining the De Re Argument.

De Re Argument

> 1. Agents are blameworthy to the extent that they fail to respond appropriately to normative reasons de re.

2. If agents do wrong due to false normative belief, then they fail to respond appropriately to normative reasons de re.

3. If agents do wrong due to false normative belief, then they are blameworthy. (From 1, 2).

4. If agents are blameworthy then they are not excused.

Conclusion: False normative belief cannot excuse wrongdoing.

The first premise of the De Re Argument makes a substantive claim about blameworthiness. Call this the Blame Claim:

Blame Claim[152]: Agents are blameworthy to the extent that they fail to respond appropriately to normative reasons de re.

In fact, this chapter argues, the Blame Claim is false. Normative reasons de re are of no particular significance for appraisal. The rest of this section outlines why some have taken the Blame Claim to be true. In the moral domain, normative reasons de re are features of a situation that make actions morally right or wrong, where this is read 'de re', and not 'de dicto'. Normative reasons de re are considerations whose content is some particular right making or wrong making feature such as 'that it is fair', or 'that it would be cruel'. Exactly what these features are depends on which is the correct first order moral theory. Normative reasons de re are to be contrasted with normative reasons de dicto, whose content is the moral value of the action, such as 'that it would be morally right', or 'that it would be morally wrong'.

Normative reasons de re have been understood to be of particular significance for various kinds of moral appraisal. Arpaly and Harman take them to be significant for appraisal of an agent in performing an action:

---

[152] See for example Alvarez and Littlejohn (forthcoming); Arpaly (2002); Harman (2011), Weatherson (2014; 2019). One might also think that Williams' disapproval of agents who exhibit 'one thought too many' in their moral reasoning stems from related concerns (see Williams (1981)).

> An action is blameworthy just in case the action resulted from the agent's caring inadequately about what is morally significant – where this is not a matter of de dicto caring about morality but de re caring about what is in fact morally significant (Harman (2011: 460), paraphrasing Arpaly (2002b)).

Markovits takes them to be significant for appraisal of actions themselves:

> Morally worthy actions are those performed for the reasons why they are right (Markovits (2010: 202)).

Michael Smith takes them to be significant for appraisal of agents in general:

> Good people care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality, and the like, not just one thing: doing what they believe to be right, where this is read de dicto and not de re. Indeed, common sense tells us that being so motivated is a fetish or moral vice, not the one and only moral virtue (M. Smith 1994: 75).

Views that endorse the Blame Claim typically do so on the basis of what I will call a Strawsonian picture of appraisal, according to which the facts that determine how an agent is to be appraised are facts about the agent's quality of will. Agents exhibit a deficient quality of will when they have desires, intentions, and motivations that are deserving of the reactive attitude of blame (see Strawson (1962)). Specifically, an agent's desires, intentions, and motivations exhibit a deficient quality of will when and to the extent that they fail to respond appropriately to what is morally important, where this is understood as 'de re' and not 'de dicto' (Arpaly(2002b); Harman (2011))[153]. However, as the following three sections argue, both the Blame Claim and the Strawsonian view of appraisal should be rejected.

---

[153] Although the views committed to the Blame Claim emphasise the important of responsiveness to what is morally right de re, and not de dicto. However, it would be possible to defend a similar view according to which responsiveness to what is morally important de dicto is also sufficient for non-deficient quality of will (for example, see Johnson-King (forthcoming; 2018)).

2. Normative Uncertainty[154]

One reason to reject the Blame Claim is that it has the implausible implication that normative recklessness is not blameworthy. The Blame Claim implies that when agents are uncertain about the normative status of the possible outcomes of their conduct (whether it is permissible, required, or forbidden), they should not take this uncertainty into consideration when choosing between actions. Commitment to the Blame Claim implies that the agent's uncertainty about the moral value of possible outcomes is irrelevant to moral appraisal, because it is only her responsiveness to normative moral reasons de re that determines appraisal. The Blame Claim says that agents are blameworthy to the extent that they fail to respond appropriately to the normative reasons de re. When agents are uncertain about what morality requires, they are also uncertain about what the normative reasons de re are, and/or what an appropriate response to them would be. This section argues that this is the wrong result, and so we should reject the Blame Claim. Instead, normative recklessness is blameworthy just as factual recklessness is. To see this, consider the following example, where the agent seems to be required to take his normative uncertainty into account when choosing how to act:

> *Voting.* Bill is trying to decide who to vote for. He could vote for the Equality party, the Freedom party, or the Other party. The Equality and Freedom parties both promise to implement policies that promote one and only one political ideal – equality in distribution or libertarian freedom, respectively. The Other party would promote other goods such as the efficient running of public services. Bill knows that the political situation that would result from the election of either of the Freedom party or the Equality party would be either optimal or a disaster, depending on which moral theory turns out to be true, while the election of the Other party would have a neutral outcome. He has studied some political philosophy and finds himself pulled

---

[154] The arguments of Section 2 can also be found in C. Field (2019b).

in two opposing directions. On the one hand he suspects the ideals of libertarian freedom upheld by the Freedom party are valuable, and knows that these are incompatible with the ideals of equality upheld by the Equality party. On the other hand, he is persuaded by the idea that equality can sometimes promote more happiness for the underprivileged, and wonders if libertarianism's emphasis on freedom is misguided. He is entirely unmoved by the policies of the Other party, who do little to promote either freedom or equality, but he knows that were they to be elected they would at least do no great harm and are likely to make some small improvements in the efficiency of public services. He must vote tomorrow, and has no more time to deliberate. Bill lives in a political system where his vote will have a direct effect on which party will be elected, such that which way he votes is a genuinely moral question.

Voting is a normative analogue of a Jackson case (see Jackson (1991: 462)). In a Jackson case, the agent has three options. For one of the options, she knows that choosing it will have a very small positive outcome. For the other two, she is uncertain about the moral value of the outcomes. She knows that one will have a very good outcome, and the other will have a very bad outcome – but she is not sure which is which. The structure of a Jackson case can be represented in the following way:

Jackson Case

	Beliefs	Facts
A	Either +100 or -100	+ 100
B	+20	+20
C	Either +100 or -100	-100

A standard position is that agents should avoid moral recklessness, and take their uncertainty into account when deliberating. This means it would be

blameworthy to choose anything other than option B[155]. Option B carries the least risk of moral catastrophe, and even though more moral value could be achieved by choosing A, choosing A would be reckless given what the agent currently believes, and given that the agent knows she has available a morally safe option[156].

In Voting, Bill faces a choice between a solution that will resolve the situation completely, producing the best outcome; a moral catastrophe, and a partial resolution that avoids the risk of catastrophe. As in the original, Bill has no way of distinguishing between the best outcome and the moral catastrophe. Whether voting for the Equality Party would have a positive or negative moral value depends on the first order moral facts of the case – on whether it is equality or freedom that should be promoted in this case, and it is these first order moral facts that he is uncertain about[157].

The important point is that Bill cannot tell from his current information whether it is freedom or equality that is required in this situation, but he knows that one

---

[155] Defenders of this view include: Bykvist (2014); Graham (2010); Moller (2011); Sepielli (2009); Zimmerman (1997; 2008; 2014).

[156] There is further debate over whether option B is also the morally right thing to do, as Zimmerman (1997, 2008, 2014) argues, or whether really option A is the morally right thing to do, but there is some other consideration that explains why it would not be morally conscientious to choose it in this case. For example, according to Bykvist (2014), option A is the morally right thing to do, but because the morally conscientious person has a preference for acting so as to promote moral value and avoid moral disvalue it would be rational for her to choose option B, because this maximises expected moral value. Similarly, Graham (2010) claims that the morally conscientious person prudentially ought to choose option B, even though morally she ought to choose option A; and that furthermore it would be blameworthy to choose option A. Distinguishing between these views is beyond the scope of the present discussion, but what they can all agree on is that it would be reckless for the agent to choose anything other than option B, and such recklessness would deserve negative moral appraisal.

[157] It is compatible with this being a genuine Jackson case that both freedom and equality are indeed political goods, and promoting them would be morally good in some situations; but in this situation the outcome of promoting one would be a catastrophe, while the outcome of promoting the other would be best. Alternatively, it may also be the case that only one of freedom or equality is a genuine good, or that both are goods but one is significantly more important than the other.

would be the morally best option, one would be the morally worst option, and there is a third morally safe but sub-optimal option available to him. Intuitively, Bill would not be blameworthy for choosing the safe option and voting for the Other Party[158] – quite the opposite, he would be blameworthy for not doing so. However, a commitment to the Blame Claim is incompatible with this conclusion.

Harman (2015) and Weatherson (2014), endorsing the Blame Claim, argue that there is an asymmetry in how uncertainty affects appraisal. On their view, choosing either option A or C does not deserve blame if the uncertainty involved is normative – if it is uncertainty about what is morally permissible, required, or forbidden. The problem with taking normative uncertainty into account, according to Harman and Weatherson, is that choosing the morally safe option involves agents taking a step back from the genuine moral reasons de re and responding instead to the de dicto consideration of what moral rightness itself requires. If the Blame Claim is true, then responsiveness to de dicto moral considerations is not something that can affect the agent's blameworthiness. It is either freedom or equality that is morally important in Bill's situation, and the Blame Claim says that to avoid blame Bill must respond to whichever of these it is. So, in so far as the agent fails to respond to moral reasons de re, she is blameworthy. The morally safe option of voting for the Other Party is an appropriate response on neither of the possible answers to the question of which are the normative reasons de re, and Bill knows this. So, by choosing the morally safe option he would know that he was failing to respond to the normative reasons de re, whichever those are. This illustrates the unpalatable implication of

---

[158] We might also endorse the further claim that he *ought* to choose the safe option. See Guerrero (2007); Lockhart (2000); Sepielli (2009); Moller (2011) for endorsements of this conclusion.

the Blame Claim: that agents are blameworthy when they act so as to mitigate moral risk by choosing Option B, as Bill does when he votes for the Other Party.

Harman and Weatherson defend this implication by arguing that there are various asymmetries between factual and normative recklessness that suggest that we should not treat them in parallel ways. In the following subsections I examine two of these apparent asymmetries and argue that, in fact, they can be explained away by further consideration of the epistemic details of the cases. Once the apparent asymmetries have been explained away, we are left with no reason to think that moral recklessness is any less blameworthy than factual recklessness is.

2.1 Steak Sensitivity

The first apparent asymmetry is that there are cases in which it can seem morally permissible for an agent to perform actions that are normatively reckless, while it is not permissible for her to perform actions that would be factually reckless. To illustrate, consider Weatherson's example:

> *Dinner.* Martha is deciding whether to have steak or tofu for dinner. She prefers steak, but knows there are ethical questions around meat-eating. She has studied the relevant biological and philosophical literature, and concluded that it is not wrong to eat steak. But she is not completely certain of this; as with any other philosophical conclusion, she has doubts. As a matter of fact, Martha is right in the sense that a fully informed person in her position would know that in this instance it is permissible for her to order steak for dinner, but Martha cannot be certain of this (Weatherson (2014: 2)).

We might represent the Martha's options as follows:

Dinner	Eating meat is wrong		Eating meat is not wrong	
	Moral Value	Hedonistic Value	Moral Value	Hedonistic Value
Steak	-10	+20	0	+20
Tofu	0	+5	0	+5

Weatherson's claim is that it is not wrong for Martha to have steak for dinner, despite the fact that she is completely certain that she has an option that is morally permissible – tofu. If Martha chooses steak, then she risks incurring moral disvalue to gain the small pleasure of eating steak. Weatherson claims that intuitively, it would not be wrong for Martha to choose steak and this shows that moral recklessness cannot be wrong. If this is correct, then this would be one respect in which normative uncertainty differs from factual uncertainty. It would show that it is morally permissible to make reckless choices when the uncertainty is normative, but not when it is factual. In the rest of this subsection I argue that this is not the conclusion we should draw from consideration of this case, because there are alternative explanations for the intuition that Martha is permitted to choose steak.

Firstly, there is more than one way to explain the intuition that it is permissible for Martha to eat steak even though in doing so she risks some moral disvalue. The putative non-blameworthiness of moral recklessness is only one way. The following paragraphs outline two alternative explanations of the intuition: one based on the observation that morality does not necessarily require us to always maximise moral value at the expense of all other value, and another based on the difficulty in distinguishing when our intuitions are tracking blameworthiness and when they are tracking moral permissibility.

First, it is sometimes permissible for agents to risk small amounts of moral disvalue for prudential gains. So, we might agree that Martha is permitted to choose the steak, and view this as morally suberogatory but permissible[159]. Although it would not be the morally best thing for her to do, Martha is permitted to risk a small amount of moral disvalue for the pleasure she would gain from eating the steak. This would mean that Weatherson is right about the moral permissibility of Martha's choosing steak, but wrong to take this to imply that moral recklessness is never wrong. Instead, the right explanation is that morality does not always require us to maximise moral value at the expense of all other kinds of value, and this is a case in which morality permits us to prioritise pleasure over the risk of moral disvalue. So, there is no asymmetry with factual uncertainty – various examples from daily life suggest that it is at least sometimes permissible to risk some moral disvalue for our own pleasure[160]. For example:

> *Driving.* Amber is deciding whether to visit the art gallery or stay at home. To get to the art gallery she must drive. She would prefer to visit the art gallery, but she cannot be certain that she will not harm someone on the way. She knows that there is a small possibility that she will harm someone while driving.

Driving	Someone is harmed		No one is harmed	
	Moral Value	Hedonistic Value	Moral Value	Hedonistic Value
Drive to art gallery	-1000	+50	0	+50
Stay at home	0	-50	0	-50

---

[159] See also Harman's view that eating meat is a suberogatory but not morally prohibited action, a 'morally permissible moral mistake' (Harman (2016)).
[160] See also Guerrero (2007) on this point.

The case is structurally similar to Dinner, except that it involves factual rather than moral uncertainty. By driving to the art gallery, Amber is performing an action that she cannot be completely certain will avoid harm, because driving accidents are always possible. Nevertheless, we usually take risks of this kind to be morally permissible in the factual case, even if not the morally best thing one could do. This is true even though the harm risked is greater than in the Dinner case. This does not suggest an asymmetry between factual and normative uncertainty, but rather that there are simply some risks that it morally permissible to take, and this is true regardless of whether the uncertainty is factual or normative. If it is permissible to risk the possibility of causing a traffic accident for the small pleasure of going to an art gallery, then it should not be surprising that the risk of wrongdoing involved in choosing steak in Dinner is also morally permissible. However, if there really was nothing of any value to be gained from driving, or from eating the steak, then it is not so clear that we would have the intuition that these actions were permissible[161]. Again, factual and normative uncertainty are symmetrical in this respect.

The second explanation is that the intuition that leads us to judge that it would be permissible for Martha to choose the steak is not tracking moral permissibility at all, but rather an unwillingness to condemn agents who take very small risks. Weatherson interprets the intuition as one that indicates the moral permissibility of choosing steak, but it is not clear that our intuitions are sufficiently fine-grained for us to reliably distinguish intuitions of moral permissibility from intuitions of other moral evaluations, such as blamelessness or excusability[162]. An alternative explanation of the intuition is that, in fact, Martha is not permitted to choose the steak – she really ought to choose tofu – but we are unwilling to blame

---

[161] Guerrero (2007) also makes this point.
[162] Various epistemologists take this to be a common phenomenon in our thinking about justification in the epistemic case (see Littlejohn (forthcoming); Sutton (2007); Williamson (2017)).

her for this transgression. One might think that it is never permissible to prioritise one's own pleasure over any risk moral of disvalue – moral value is just too important to risk in this way. This would mean that, strictly, Martha is not permitted to choose the steak, and Amber is not permitted to drive to the art gallery. One might acknowledge this, but be unwilling to go as far as blaming them.

For example, we might think blame is a serious sanction that is appropriate only for relatively serious transgressions, in which more is at stake than just one steak at one dinner. Although Martha does make a reckless choice, it is only very slightly reckless. So slight, in fact, that describing the case as 'reckless' sounds a little odd; we typically reserve the term for more serious cases where greater risks are taken. Even on the assumption that eating meat is wrong, choosing steak once at one meal is not a great moral catastrophe. As such, risking this minor wrong is perhaps not something deserving of any of the typical interpersonal actions associated with blame, even if the agent ought not to have done so. This would explain the intuitions that Weatherson appeals to in the Dinner cases. Nevertheless, it is plausible that our intuitions might change were we to increase the stakes:

> *Veganism.* Martha is trying to decide whether to become a vegan or not. She enjoys eating meat and dairy products, so would prefer not to. She has studied the relevant philosophical literature carefully and concluded that it would not be morally wrong for her to choose to continue consuming meat and dairy products. However, she is not completely certain of this; as with any other philosophical conclusion, she has doubts. She also knows that if it turns out that she is wrong, the impact that a lifetime of consuming meat and dairy products would have would constitute a serious moral wrong. As a matter of fact, Martha is right in the sense that a fully informed person in her position would know that it is permissible for her to not become a vegan, but Martha cannot be certain of this.

The consequences of a lifetime of meat eating are much greater than the consequences of one steak at one dinner, and this is a case in which it is much less clear that Martha can be excused for choosing a lifetime of meat-eating, given her uncertainty.

Another feature of Dinner that might make one hesitant to blame Martha is the fact that her epistemic situation makes her choice only very slightly reckless. There are three features of her epistemic situation that serve to mitigate the recklessness of her choice. Firstly, Martha is reasonably confident in her belief that choosing steak is permissible. Martha in Dinner is 'not completely certain', making her choice to eat steak not as reckless as it might be[163]. We might represent her epistemic state as follows:

Dinner	Belief	Credence	Facts
Steak	Morally Permissible	0.9	Morally Permissible
Tofu	Morally Permissible	1	Morally Permissible

Martha's high credence in the permissibility of eating steak helps to mitigate the recklessness of her choice, and contributes to the sense that she does not deserve blame. Secondly, the doubts she has are comparable to doubts one might have about most philosophical conclusions. We might think it would be unreasonable to expect her to be any more confident. It is very difficult – if not impossible – for a conscientious epistemic agent to achieve complete certainty in most philosophical claims, and a greater degree of certainty might even make us suspicious. If this is right, then Martha's 0.9 credence makes her about as sure as it is reasonable to be about a moral claim, and so plausibly carries much more

---

[163] For example, compare Jackson's original case, in which Jill the physician must choose between three drugs, and 'there is no way that she can tell' which drug will cure and which will kill the patient (Jackson (1991: 463)).

weight than the same credence for a factual claim, where certainty is often more achievable. Compare Martha to a more reckless counterpart:

> *Lunch.* Louise is deciding whether to have steak or tofu for lunch. She prefers steak, but knows there are ethical questions around meat-eating. She has studied the relevant biological and philosophical literature, and is still unsure. She is inclined to think that meat eating is permissible, but she is not completely sure if this is what the evidence supports, and furthermore she suspects that wishful thinking might be influencing her evaluation of the evidence. So, she has doubts about the correct answer. As a matter of fact, Louise is right that it is permissible for her to order steak, and a fully informed person in her position would know this.

It seems far less plausible that Louise's recklessness is blameless, as it seemed that Martha's could be. While both Louise and Martha make reckless choices, Louise's choice is much more reckless because she is much less sure that her beliefs are correct. Louise could have easily chosen the low risk option until she had investigated the relevant considerations more carefully, and the fact that she does not plausibly contributes to her deserving blame. Thirdly, the agent's evidential state is also relevant in determining how reckless her choice is, and thereby how blameworthy she would be. Consider the following example:

> *Breakfast.* Nora is deciding whether to eat steak or tofu for breakfast. She prefers steak, but knows there are ethical questions around meat-eating. She has not bothered to study the relevant biological and philosophical literature, even though as a well-educated philosophy student, it would not be difficult for her to do so. Nevertheless, she is almost (but not completely) certain that it is permissible to eat steak. This is mostly because meat-eating seems natural to her, and most members of her community do not find eating meat morally problematic. As a matter of fact, Nora is right in that a fully informed person in her position would know that meat-eating was permissible.

We could represent Nora's credence in the permissibility of eating meat at 0.9, so just as high as Martha's. However, despite having identical credences, Nora seems much more deserving of blame than Martha. One explanation of this is

that Martha's evidential basis for her high credence is much better than Nora's, and this means that her choice is much less reckless than Nora's. Martha has done everything that we could expect of her in investigating the morally relevant features of the situation. In contrast, Nora could have done much more – as a philosophy student she has the capacity to be do much more to inform herself of the morally relevant features and better equip herself to reach a well-informed conclusion. Instead she has lazily relied on the opinions of others. While both are reckless, Nora's choice seems much more so because it is based on a poor evidential foundation. These various features of Martha's epistemic situation all contribute to mitigating the recklessness of her choice, and thereby the blame she deserves. This allows us to explain the intuition Weatherson identifies without needing to appeal to an asymmetry between factual and normative uncertainty, and without needing to deny that normative recklessness is blameworthy.

2.2 Inverse Akrasia

A second apparent asymmetry between factual and normative recklessness is that there seem to be cases in which agents who act against their moral beliefs are praiseworthy. If moral recklessness was blameworthy, then these agents would be prime examples of morally reckless agents and we would expect them to be blameworthy rather than praiseworthy. Arpaly suggests that Huck Finn is an example of an inversely akratic agent:

> *Huck Finn.* Huck Finn befriends Jim, a slave, and helps him escape from slavery. While Huck and Jim are together on a raft used in the escape, Huck is plagued by what he calls "conscience." He believes, as everyone in his society "knows," that helping a slave escape amounts to stealing, and stealing is wrong. He also believes that one should be helpful and loyal to one's friends, but loyalty to friends is outweighed by some things, such as property rights, and does Miss Watson, Jim's owner, not have property rights? Hoping against hope to find some excuse not to turn Jim in, Huck deliberates. Like many children (and adults), Huck is not very good

at abstract deliberation, and it never occurs to him to doubt what his society considers common sense. Thus, he fails to find a loophole. "What has poor Miss Watson done to me," he berates himself, "that I can see her [slave] go away and say nothing at all?" Having thus deliberated, Huck resolves to turn Jim in, because it is "the right thing." But along comes a perfect opportunity for him to turn Jim in, and he finds himself psychologically unable to do it. He accuses himself of being a weak-willed boy, who has not "the spunk of a rabbit" and cannot bring himself to do the right thing, and eventually shrugs and decides to remain a bad boy (Arpaly (2002: 75)).

We might represent Huck's options as follows:

Huck Finn	Freeing Jim is right	Freeing Jim is wrong
Turn Jim in	Very morally bad	Very morally good
Free Jim	Very morally good	Very morally bad

As Arpaly sees it, Huck's action is intuitively praiseworthy, despite the fact that he does it in the belief that it is wrong. He pays attention to the right kinds of things – Jim's humanity and welfare – and as a result cannot bring himself to turn Jim in to the authorities, even though he (falsely) believes that turning him in would be the right thing to do. This seems to suggest that moral recklessness is not blameworthy – Huck seems even more morally reckless than someone who fails to choose the morally safe low risk option when they are normatively uncertain – he actually believes that what he does is wrong, and he does it anyway. If moral recklessness was blameworthy, then we should blame him rather than praise him, some have taken this to suggest that there is an asymmetry between factual and normative recklessness. However, further consideration of Huck's motivations allows us to accommodate the intuition that Huck Finn is praiseworthy while maintaining that moral recklessness is blameworthy.

Actions can be more or less praiseworthy, and different features of those actions can exhibit different degrees of praise and blameworthiness. Acknowledging this allows us to grant that Huck's action has *some* praiseworthy aspects – for example, Huck acts in response to morally significant things: the recognition of Jim's humanity, the desire to help a friend, and so on. If he had instead freed Jim purely out of a desire to cause mischief, then it is not so clear that he would deserve praise. However, we can acknowledge this while maintaining that in so far as Huck acts morally recklessly, he is blameworthy. The blameworthiness of his recklessness is compatible with there being other features of his action that are praiseworthy, and it is compatible with the possibility that his action overall has more praiseworthy features than blameworthy features. Huck's action is praiseworthy, but it is not maximally so. Had Huck freed Jim without acting recklessly; for example, if he had freed Jim in the knowledge that his action was right, then his action would have been even more praiseworthy. This would have removed any risk of error, error which might occur due to his false and objectionable belief that property rights can sometimes apply to people[164]. Views that deny the blameworthiness of moral recklessness cannot capture this aspect of the Huck Finn case, and cannot distinguish between the sense in which Huck is praiseworthy and the sense in which agents who do the right thing because they know it is the right thing are praiseworthy. Those who deny that moral recklessness is blameworthy explain Huck's praiseworthiness by appeal to the claim that our moral beliefs are irrelevant to the moral worth of our actions. However, the alternative explanation that Huck's action is praiseworthy because it has features that are sufficiently praiseworthy so as to outweigh any other features of it that would otherwise make it not praiseworthy is just as good. Not only is it able to explain the praiseworthiness of Huck's action, it can also

---

[164] Sliwa (2016) also makes this point.

distinguish between degrees of praiseworthiness, and is consistent with the intuitive view that moral recklessness is in general blameworthy rather than praiseworthy. To illustrate the plausibility of this alternative explanation, compare an agent who acts just as recklessly as Huck, but without the praiseworthy qualities that redeem his action:

> *Tuck Finn.* Huck's cousin, Tuck, lives in the same society as Huck and has the same moral opinions. Like Huck, Tuck believes (as everyone in his society "knows") that helping a slave escape amounts to stealing, and stealing is wrong. Tuck harbors an intense and unreasonable dislike for Miss Watson, and in order to spite her conspires to help one of her slaves escape.

It seems fairly clear that Tuck is not to be praised for his action. This is the case even if we think he does the right thing. In so far as Tuck does the right thing, he does so accidentally. He is not to be praised for this accidental right action because the motivations out of which he does it are in no way admirable – he believes what he is doing is morally wrong, and his main motivation is a desire to spite Miss Watson. We might think of this as an example of a case in which praiseworthiness and moral rightness come apart[165]. To bolster the point, notice that akratic action only seems unproblematic when the agents end up doing the right thing. Compare the following case:

> *Mafia.* Dmitri, like most people in his society, believes that killing is wrong. He is a member of the mafia, and as well as believing that killing is wrong values friendship and loyalty to family very highly. An associate has wronged his brother, and it seems to him that the only respectable thing to do would be to kill the associate, even though this would be morally wrong. Although he knows it is wrong, he just cannot bring himself to let the wronging slide and shame his family, so he sets out to murder the associate.

---

[165] See Hawthorne and Srinivasan (2013) for further discussion of this.

This is a non-inverted akrasia case. While Huck Finn does the right thing akratically, Dmitri ends up doing the wrong thing akratically. Unlike Huck Finn, he seems blameworthy. This suggests that part of what the intuition of Huck's praiseworthiness is tracking is the fact that he does something good rather than bad in spite of his false beliefs. However, there are possible explanations for this intuition that are neutral on the issue of moral recklessness. For example, it is possible that we judge Huck as praiseworthy because in unknowingly responding to reasons that make freeing Jim morally right, he is taking a first step away from the false views held by his society, and so beginning to equip himself with the tools to do the right thing in the future. All this suggests that the Huck Finn case does not on its own demonstrate any asymmetry in the significance of factual and normative uncertainty, and nor does it show that moral recklessness is not blameworthy, since there are explanations of our intuitions in response to the Huck Finn case that are compatible with moral recklessness being blameworthy.

In summary, we do not have reason to deny that normative recklessness is blameworthy. The following two sections (§3 and §4) present a further reason to reject the Blame Claim, namely that there are counterexamples to it if we assume the Strawsonian view of appraisal that is typically used to motivate it.

3. Moral Limitations

This section presents three cases in which the agent has a pattern of desires and motivations that are not blameworthy according to a Strawsonian view of blameworthiness, even though they fail to respond appropriately to normative reasons de re. Each of these cases involves an agent who is in some way limited in their capacity to respond appropriately to morally salient features of the situation due to some neuroatypicality. As a result, this section shows that the Blame Claim faces counterexamples even if we assume the Strawsonian view of

appraisal that is typically used to motivate it. The following section (§4) argues for the further claim that, given these counterexamples, the Strawsonian view permits some cases of false normative belief to excuse.

A Strawsonian view of appraisal, as I use the term here, takes appraisal to be determined by the agent's desires and motivations – her 'quality of will'. Agents deserve blame to the extent that they exhibit a deficient quality of will, and they exhibit a deficient quality of will when their patterns of desires and motivations are blameworthy. A pattern of desires and motivations is blameworthy when it would be appropriate to have a reactive attitude of blaming in response to it (Strawson (1962); Arpaly (2002b)). Some take this to support the Blame Claim because they take the view that patterns of desires and motivations that fail to be appropriately responsive to what is morally important (de re and not de dicto) are blameworthy, in virtue of deserving the reactive attitude of blaming (Arpaly (2002b); Harman (2011)). However, as the following cases show, not every possible case in which an agent fails to respond to what is morally important (de re) is a case in which the agent exhibits a blameworthy pattern of desires and motivations.

Moral reasoning is a skill that not everyone excels at. Incompetence in this skill of moral reasoning comes in various forms and can be exhibited to various degrees. Some people exhibit moral incompetence because they are bad at thinking through the implications of morally relevant considerations and working out what the right thing to do is. Others exhibit moral incompetence because they are bad at noticing morally salient features of situations. Others do so because they are bad at appreciating the force and significance of some kinds of normative moral reasons[166]. The rest of this section describes three possible

---

[166] Of course, these three groups might overlap.

agents, each with neuroatypical psychological features that ground each of the three limitations in moral competence just described. The first case involves limited capacities for moral reasoning:

> *Attention Span*. Steve has a low attention span, and finds it difficult to concentrate on any one thing for very long. He desires to do what is fair. However, he is not very good at reasoning successfully from his desires to actions that effectively realize those desires. Sometimes, his low attention span causes him to fail to do the action that would be the appropriate response to the normative moral reasons de re. He is trying to decide how to allocate office space among a group of researchers. The available space is such that not everyone can have an ideal office. Some must share, some offices are in noisy locations, some are small, or damp. Various considerations contribute to determining the fair allocation of the available office space (how much each researcher uses the office, their accessibility needs, how well they get along with colleagues, perhaps the length of time they have been with the institution, and so on). Working out the fairest allocation is somewhat complicated, and part way through the decision process Steve becomes distracted. As a result, he does not allocate the office space fairly.

While Steve has morally admirable desires – he desires to promote fairness – he lacks the ability to reason effectively from these desires to the action that successfully realizes those desires in his situation. When he fails to determine the fair allocation of office space, he fails to choose the action that constitutes an appropriate response to the normative reasons de re in this situation. However, the Strawsonian account of blameworthiness should not evaluate him as blameworthy, because he does not exhibit a blameworthy pattern of desires and motivations. Quite the contrary, his desire to promote fairness indicates a good, rather than deficient, quality of will.

Supporters of the Blame Claim can accommodate this case by pointing out that in trying to promote fairness, Steve is responding appropriately to normative

reasons de re, and so is not blameworthy. The fact that he often becomes distracted does not affect the fact that he has a good will in virtue of desiring the right kinds of things (see Arpaly and Schroeder (2013: 201)). However, the case is worth discussing because it is relevantly similar to the third case, which is more difficult for supporters of the Blame Claim to deal with. The second case involves limited capacities for recognizing normative reasons de re.

> *Imagination.* Mike finds it difficult to imagine the internal mental lives of other people, and he finds it particularly difficult to imagine how others will feel in response to his actions[167]. Since he finds other people's emotions difficult to imagine, he also finds it difficult to see them as reasons for and against actions. Mike desires the well-being of his fellows, and he believes that the well-being is always served by their knowing the truth. Sometimes people's feelings are hurt when he is too truthful. However, he finds it very difficult to tell when this happens. He usually only notices that it has happened after the fact, when his closer friends and family explain to him why that person was upset, and tell him that he must try to be more sensitive. Although he desires to do whatever will promote the well-being of his fellows, he finds it hard to believe that the 'sensitivity' and 'hurt feelings' that others describe really track anything of genuine importance. As a result, he rarely responds appropriately to these features when they are morally salient.

Let us suppose that, ceteris paribus, that someone's feelings would be hurt is a normative reason de re not to do that which would cause the hurt feelings. Mike, as described, is someone who systematically fails to recognize some of the normative reasons de re, and often does wrong (for example, asserting hurtful truths) while failing to respond to appropriately to the possibility that these hurtful truths will hurt others. Responding appropriately to any particular normative reason de re would seem to at least presuppose being able to recognize

---

[167] As described in Barnbaum (2008); Richman and Bidshahri (2018).

it and distinguish it from other considerations, and this is something Mike cannot do for a particular subset of normative reasons de re[168]. However, he does not exhibit a blameworthy pattern of desires and motivations that could be described as 'ill will'. He desires the well-being of his fellows, and he acts so as to promote this in so far as he is able.

Arpaly and Schroeder, who support the Blame Claim, accommodate this case by saying that Mike's limitations are morally neutral; they do not affect his quality of will (2013: 201). While this is plausible, it is nevertheless false that Mike can be said to respond appropriately to normative reasons de re when he has no idea what those reasons are. Responding to a reason would seem to require, minimally, that the agent be able to recognize that reason and act for it (rather than merely in accordance with it). Mike cannot do this for some normative reasons de re, since he is unable to understand the emotions of others. He may be able to indirectly respond to these reasons, but allowing this is in tension with the Blame Claim. The Blame Claim places particular weight on direct responsiveness to normative reasons de re, and not de dicto. However, if indirect responses to more general reasons (for example, the well-being of others) can also count as appropriate responses to normative reasons de re, then this raises the question as to why responses to the de dicto reason 'that it is morally right' could not also be an appropriate indirect response to normative reasons de re.

The Blame Claim is also incompatible with the third case, which involves limited capacities to appreciate the force and significance of normative reasons de re:

> *Distinction.* Bonnie consistently fails to notice and understand the distinction between moral and conventional wrongness[169].

---

[168] Responding to a reason requires more than merely doing what it recommends. For more on the distinction between acting for and acting merely in accordance with a normative reason, see Mantel (2018).

[169] As described in Cleckley (1955); Blair et al. (2005); Dolan and Fullam (2010); Shoemaker (2011).

Bonnie desires above all her own well-being. She has managed to learn by heart most of the actions that her society consider wrong, and has found that avoiding these actions in general promotes her well-being. She also knows that some of these actions are considered 'morally' wrong, and that these are typically thought by others to be very important to avoid. However, she is unable to feel the force of the distinction herself. One day, she is waiting for a taxi in the rain when a pregnant woman arrives at the taxi rank, urgently needing a taxi to take her to the hospital. It does not occur to her that it might be wrong *not* to offer the first taxi to the pregnant woman, because the situation does not fall under the description of any of the actions that she has learned are wrong. As a result, Bonnie takes the first taxi when it arrives, and does not allow the pregnant woman to go first.

Suppose that there is a normative reason de re for Bonnie to let the pregnant woman take the first taxi, and that this would be clear to most neuro*typical* agents in the same situation. Bonnie does wrong in failing to let the pregnant woman take the first taxi, but it is not obvious that the pattern of desires and motivations she exhibits deserve blame on a Strawsonian view: they are not malicious – she does not desire to harm the pregnant woman in any way. They do not deserve praise either: if anything they are neutral. Furthermore, Bonnie has limited capacities to appreciate the force of morally salient considerations, such as that the pregnant woman needs the taxi, and here this leads her to fail to respond appropriately to the normative reason de re that makes it wrong for Bonnie not to offer the taxi to the pregnant woman. Like Steve and Mike, Bonnie exhibits a moral incompetence that is a result of her limited capacities to appreciate a subset of the morally salient features of the situation – the moral significance of particular considerations, such as the needs of others. If Steve and Mike are not blameworthy, despite failing to respond appropriately to normative reasons de re, then it is plausible that neither is Bonnie.

However, some would resist the claim that Bonnie exhibits a morally neutral pattern of desires and motivations. Some would argue that there is a salient difference between the limitations that Steve and Mike have, and Bonnie's limitations. Specifically, Steve and Mike care about the right kinds of things, while Bonnie does not. For example, Steve cares about morally important considerations, such as fairness. His limitations affect only the execution of his actions. Similarly, Mike cares in a general way about the well-being of his fellows, but is unable to respond to salient considerations relevant to what that involves in some situations – particularly when the emotions of others are involved. Bonnie does not care about many of the morally salient considerations, and some would argue that this lack of concern is precisely what it is to exhibit a negative quality of will (Arpaly (2002b); Arpaly and Schroeder (2013); Harman (2011); Mason (2015)). However, there are reasons to think that this is unfair.

In so far as the practice of blaming is a backward-looking activity, involving the imposition of sanctions, punishment, or negative judgment, it would seem that it ought to be governed by principles of fairness, such as the following:

> *Capacity Principle*: It is not fair to blame people for doing what
> they had no, or very little, capacity to do.

There are plausibly other blaming practices that are not of this kind, and not subject to principles of fairness, but these are not our focus[170]. In so far as responsiveness to normative reasons de re is something that some people find much more difficult than others due to robust features of their psychology, this is unfair. To blame any of the three neuroatypical agents described for not

---

[170] Pereboom (2013) also makes this point. There are numerous alternative views of what the practice of blaming involves that are able to sidestep this objection from fairness. For example, views on which blame is an evaluative tool for marking debits on one's moral account (Smart 1961), a way of acknowledging a change in an interpersonal relationship (Scanlon 2008), an expression of reactive attitudes (Strawson (1962)), or a tool for social coercion (Strawson (1962); Vargas (2012).

responding appropriately to the normative moral reasons de re would violate the Capacity Principle. So, the Blame Claim is in tension with the Capacity Principle, since it says that any failure to respond to normative reasons de re is blameworthy. This should give us reason to be suspicious of the view[171].

One way that Strawsonian views can respond to this charge of unfairness is to argue that the kind and degree of moral incompetence that the neuroatypical agents display means that they are less than full members of the moral community, and so exempt from moral appraisal, because they lack the capacity to do what is required to avoid blame (Shoemaker (2011); Strawson (1962)). For example, Strawson argues that we ought to take the 'objective stance' towards neuroatypical agents in our appraising practices (1962). While this would avoid the problem of endorsing unfair sanctioning practices, it involves banishing individuals from the moral community in a way that is morally unacceptable. Limitations in competence are not usually considered acceptable grounds to banish people from moral and social practices. For example, in cases of physical limitation, it is usually considered appropriate to take steps to accommodate individuals who lack the capacity to participate in the community in the ordinary way, and morally wrong to fail to do this. The expectation that wheelchair ramps will be provided to allow individuals to access public buildings is one example of this. Most people would think that taking Strawson's 'objective stance' towards those with limited physical capacities, for example by banishing wheelchair users from communities and practices that take place inside buildings with steps, would be very morally wrong. Given this, taking the objective stance towards individuals with psychological limitations such as Bonnie's would seem to be similarly morally wrong.

---

[171] See Brink and Nelkin (2013), and Rosen (2002) for similar points.

Not only this, but taking the objective stance towards neuroatypical agents like Bonnie would be impractical. Exempting neuroatypical people from the demands of morality depends on the possibility of drawing a clean line between neuroatypical agents and the rest of us. It is not obvious that this is possible. The psychological characteristics contributing to the limitations in moral reasoning – attention span, empathy, and appreciating moral salience – come in degrees, and vary among the population. Even the most neurotypical of us will have found ourselves in morally unfamiliar situations in which our epistemic access to the normative reasons de re is limited. This can happen, for example, when we travel in an unfamiliar culture, when moral understanding is unevenly distributed[172] or when we are puzzled or mistaken about the correct first order moral theory. One reason that it is useful to consider neuroatypical agents is that they can be thought of as extreme examples of these more everyday situations.

In summary, if we accept the Strawsonian view of appraisal that is typically used to motivate the Blame Claim, then Attention Span, Imagination, and Distinction are not cases in which the agent exhibits ill will, and so these are not cases in which the agent is blameworthy. In so far as the Blame Claim is motivated by this Strawsonian view, these constitute counterexamples to the Blame Claim. Furthermore, the attempt to dismiss these counterexamples by appeal to the Strawsonian objective stance generates a view that has unfair implications that are seriously morally questionable. These unfair implications are a reason to avoid interpreting the Strawsonian view in such a way that it is committed to them[173]. The following section shows that once we accept this, we see that there

---

[172] For example, see Calhoun's example of the uneven distribution of moral knowledge relevant to feminism and the wrongness of sexual harassment Calhoun (1989).

[173] Some who endorse both the Strawsonian view and the Blame Claim have nevertheless insisted, despite counterexamples such as the ones presented here, that failing to care about and be responsive to normative reasons de re just is what it is to exhibit ill will, and that unfairness is beside the point (see Arpaly (2002, 2013); Harman (2011, 2015); Weatherson (2014)). Though I

are some cases in which false normative belief can excuse, even when agents fail to respond to normative moral reasons de re.

4. Bad Advice Cases

Building on the argument of the previous section, this section argues that there are some cases in which false normative belief can excuse agents who do wrong while failing to respond to normative moral reasons de re. These cases involve agents who have limitations in their capacities for responding appropriately to normative reasons de re, of the kind exhibited by the three neuroatypical agents described in the previous section. In these cases, the agents seek out and act on moral advice about what to do in a difficult moral situation where there are high moral stakes.

Even if one cannot respond directly to normative reasons de re, one can often still act in accordance with them by performing the action that would constitute an appropriate response to those reasons. If one is aware of one's limitations and also fortunate enough to have available some method of working out what the right thing to do is, then it is nevertheless possible to respond to the reason "that it is the right thing to do", even if one is unable to respond directly to the normative reasons de re that make that the right thing to do. This is the situation that Steve, Mike, and Bonnie seem to be in with respect to morality. While they have limitations that make acting morally difficult, they will at least sometimes be able to do what the normative reasons de re demand. For example, some of them will have friends who are more skilled at moral reasoning than them, who they could ask for advice about what they ought to do.

---

have offered some reasons to resist this view, at some point the disagreement seems to come down to a clash of intuitions that it is not clear is resolvable.

To illustrate, compare my situation with respect to cooking. Just as Steve, Mike, and Bonnie find it difficult to respond appropriately to normative reasons de re in the moral domain, I find it difficult to respond appropriately to normative reasons de re in the culinary domain. Just as this limitation makes them bad at moral reasoning, it makes me bad at cooking. For example, suppose that there is a normative reason de re to add salt to pasta, a consideration such as "that it improves the texture of the pasta", or "that it improves the taste of the pasta". However, I find myself unable to fully appreciate this reason – I have never been able to discern the difference between pasta cooked with added salt and pasta cooked without it. If there is indeed a gastronomical reason to add salt to pasta, then I consistently fail to recognize, care about it, and respond appropriately to it. However, this limitation in my capacity to respond appropriately to normative reasons de re in the culinary domain does not prevent me from acting in accordance with them, since I can seek out advice from others who are more skilled than me. I know that it would be wise to do this because I know that I am bad at cooking. I know this from the reactions on my friends' faces when they taste my food, and I can work this out without being able to work out why my cooking is bad, or what I would need to do to improve it. Furthermore, I know that there are various gastronomically significant features – normative reasons de re – that I am failing to respond to and this contributes to my cooking being bad. Suppose that I am in a situation in which the gastronomical stakes are high – perhaps I am throwing a large dinner party. It is somewhat important that I make a tasty meal, since my friends' well-being is riding on it[174]. Fortunately for my dinner guests, I do not need to rely on my own abilities to recognize culinary reasons in order to act in accordance with them and prepare a tasty meal. There are methods I can employ to help me to act in accordance with the normative

---

[174] This may or not be of genuine moral importance – nothing hangs on this.

161

reasons de re in the culinary domain, such as recipe books, or friends who are better at cooking. I can also use considerations of risk and stakes to discern when would be reckless not to seek advice.

Just as I should seek advice when if I am in a high stakes culinary situation – for example, if I am throwing a dinner party – it would seem that agents who have limited capacities for moral reasoning should seek advice when they find themselves in tricky moral situations. Furthermore, it seems that this would be the best thing they could do given what they know about their own limitations and the moral stakes of the situation. We might even think they are blameworthy if they do not seek advice, since it would be reckless for them to rely on their own moral reasoning given what they know about their limited capacities. Certainly, they would not be blameworthy for seeking advice. However, advice can be misleading. Even the most reliable advisors are sometimes mistaken. This presents a problem for view that endorse the Blame Claim, since they imply that if an agent with limited capacities to respond appropriately to normative reasons de re seeks out and acts on misleading advice, then they are blameworthy. This is because they do the wrong thing and they fail to respond appropriately to normative reasons de re, and so are blameworthy according to the Blame Claim. This is because the advice they receive will be of the following form:

"Do X, because X is morally right."[175]

The advice cites a consideration that is the action's moral value – that X is the right thing to do, and so is a normative reason de dicto and not de re. We can assume that this is the only reason the agents respond to when they do X, so in

---

[175] It might be contested that the testimony is in fact of the form 'S says that X is morally right'. Unless it is the case that the action would be made right merely by S's saying it is right, the morally salient information that the testimony provides is the de dicto consideration that X is to be done because it is morally right.

doing X they do not also respond to any normative reasons de re – we have stipulated that these are agents who have very limited capacities to respond to normative reasons de re. Responsiveness to this kind of consideration is a motivation that Michael Smith calls 'fetishistic', because it concerns moral rightness de dicto and not de re (see M. Smith (1994: 74)). However, in this situation, that X is morally right seems to constitute a good reason for morally limited agents to do X, and it is implausible that they would be blameworthy for acting on this consideration[176].

This is a problem for the Blame Claim because when morally limited agents act on this advice, they are not responding to normative reasons de re. This is because the advice is advice about what is morally right (de dicto), and so responding to it is not a way of responding to the normative reasons de re. This means that if this advice is misleading, then the agents are blameworthy for doing the wrong thing when they act on the advice. However, this does not seem right. Even assuming a Strawsonian view of moral appraisal, it does not seem right to say that morally limited agents are exhibiting ill will by acting on misleading advice while trying to do what is morally right. On the contrary, their desires and motivations are, if anything, admirable. In seeking out and acting on the advice they would be exhibiting a desire to do what is morally right. Furthermore, in seeking advice they are knowingly taking what they justifiably believe is the most efficient means for them, given their limited capacities, to do the morally right thing. This seems like conduct that should be praised – even if it is not to be praised to the maximal possible degree (see Hills (2009), Sliwa (2012)).

---

[176] Although, we might think that they would be blameworthy if, having received this advice, they failed to attempt X.

Some would object to this, arguing that in so far as the agents lack concern for the thicker content of what morality requires, they exhibit a deficient quality of will (see Arpaly (2002b); Harman (2011)). The implication is that desires and intentions to do what is morally right 'de dicto', when accompanied by unresponsiveness to what is morally right 'de re', is always an instance of deficient quality of will. However, close consideration of neuroatypical agents shows this to be false. It is false because it relies on an inference about the kinds of desires and intentions that underpin the motivation to do what is morally right de dicto. This inference is usually cogent when made about neuro*typical* agents, but it does not transfer straightforwardly to neuro*atypical* agents such as Steve, Mike, and Bonnie. While it is plausible that most neurotypical agents who act out of a motivation to do the morally right thing de dicto and who are unresponsive to moral reasons de re have a deficient quality of will, it is not so clear that this true of neuroatypical agents. When they seek out and act on misleading advice, the neuroatypical agents under discussion act on the motivation to do the morally right thing de dicto, and they are unresponsive to moral reasons de re. However, they do not exhibit a blameworthy pattern of desires and intentions, ill will, in doing so. They are responsive to normative reasons de dicto because they are unable or find it very difficult to respond to normative reasons de re. This does not seem to be an instance of ill will, or something that deserves blame. If this is right then the Blame Claim cannot be true, even if we accept the Strawsonian picture of moral appraisal that typically underpins it.

Addressing this worry, supporters of the Blame Claim divide cases of moral advice taking into two categories (Harman (2015), Weatherson (2014)). If taking moral advice is not blameworthy then this is because the advice taker is factually ignorant, and there are some non-normative facts that her moral advice giver has that she does not. She should therefore defer to their judgment not out of the de

dicto motivation to do the right thing, but as a way to take into consideration the non-normative facts that she is unaware of and that ideally she would investigate before reaching a judgment. All other cases of moral advice taking, they argue, are objectionable cases of motivation de dicto, which is never praiseworthy and can never excuse wrongdoing, because it exhibits a deficient quality of will. However, when the neuroatypical agents described above act on moral advice, they do not seem to fit comfortably into either category of ill will or pure factual ignorance. As noted above, they are not happily described as manifestations of ill will, because they are also cases in which the agents are trying to their very best to do what is right. However, nor are they cases of pure factual ignorance. For example, Steve does not lack factual knowledge, instead he lacks the ability to reason effectively from his desires and commitments to an action that successfully realizes those desires and commitments in this situation. This is consistent with his not being ignorant of any of the non-normative facts. Similarly, although the normative reasons de re that Mike fails to recognize are partially factual, they also have a moral dimension. The concept of 'hurt feelings' is at least partially moral. Since Mike is poor at imagining how others feel, he is unresponsive to the normative reason de re 'that saying P would be hurtful', because his understanding the concept 'hurt feelings' is shaky[177]. Nor can Bonnie's ignorance be described as factual. She has no trouble identifying morally wrong actions, it is the significance of their value that she fails to understand.

A supporter of the Blame Claim might seek to accommodate misleading advice cases by agreeing that the agents are blameless, but denying that they really are cases of unresponsiveness to normative reasons de re. Instead, it might be

---

[177] This highlights the difficultly of drawing a clear distinction between the factual and the normative, which provides a further reason to prefer an account that does not rely on any such distinction.

argued, our neuroatypical agents could and ought to seek advice not about what the morally right thing (de dicto) to do is, but about what the normative reasons de re are. This advice would be of the following form:

> "Respond to consideration Y, because it is one of the normative moral reasons de re".

However, it is not clear that this kind of advice would be of much use to morally limited agents. For this advice to be useful, the agent would need to be able to draw the right conclusions about what normative reasons de re require. So, this would not help the neuroatypical agents, since it is not clear that they can do this. It would be of no use to Steve, who is bad at working out what his reasons support; or to Mike, who is bad at understanding the relevance of some of the normative reasons de re to moral value; or to Bonnie, who is bad at appreciating the value of morally salient considerations. To insist that these agents seek only advice in de re form, regardless of its usefulness, itself suggests a kind of fetishism for moral reasons de re for which we would need more justification.

In summary, when agents who have very limited capacities to respond appropriately to normative reasons de re seek out and act on moral advice about what to do, they are not blameworthy. Sometimes, this advice will be misleading. When the advice is misleading, and they act on that advice, they will do the wrong thing. Such cases show that, sometimes, false normative belief can excuse. Specifically, when the agent does the wrong thing, is not blameworthy, and is not blameworthy in virtue of a consideration that means she does not deserve blame (see Ch. 4, p. 93), in this case that she believed falsely about what morality required and had limited capacity to respond appropriately to reasons that would have supported doing the right thing.

5. Summary

This chapter has argued that we should reject the view that false normative belief can never excuse, since the key argument for this view – the De Re Argument – fails. The argument fails because it has a false premise, namely the Blame Claim. I have offered two reasons to reject the Blame Claim: that it has very implausible consequences for action under normative uncertainty (§2), and that some neuroatypical agents present counterexamples to it (§3). These counterexamples are compatible with the Strawsonian view of blameworthiness typically used to motivate the Blame Claim, unless that view is interpreted in a way that is unacceptably unfair. Finally, I showed how when agents limitations in their moral reasoning capacities do wrong while acting on misleading moral advice, they present counterexamples to the view that false normative belief can never excuse (§4). Having established here that false normative belief can sometimes excuse, in the following, final, chapter I clarify the extent to which false normative belief can excuse by addressing and rejecting arguments for the view that blameless false normative belief always excuses.

Chapter 6

False Normative Belief *Only* Sometimes Excuses

This chapter argues against the view that false normative belief, at least when it is itself blameless, always excuses[178]. This view is in tension with my preferred view that false normative belief *only* sometimes excuses. This chapter considers and rejects two ways of arguing for the view that blameless false normative belief always excuses. Section 1 considers an argument from a putative control condition on blameworthiness, and Section 2 considers an argument from a putative belief condition on blameworthiness. Section 3 further clarifies the conditions under which, on my view, false normative belief can excuse. On this view, false normative belief sometimes excuses.

1. The Control Condition

One way to argue that false normative belief always excuses is by appeal to a control condition on blameworthiness.

> *Control Condition*: One is blameworthy for something only if one
>
> was in control of that thing.

From this the following argument can be made:

---

[178] Since it is typically only blameless false belief that is thought to excuse (see, for example, Zimmerman (2008, Ch. 4)), this will be the focus of this chapter. Henceforth, by false normative belief" I will mean only false normative belief that is itself blameless, unless indicated otherwise.

Control Argument

P1: For any action, A, done by S from a false belief that A is permissible, S is blameworthy for A only if S is blameworthy for the false belief from which it was performed (from P1)[179].

P2: One is blameworthy for something only if one was in control (directly or indirectly) of that thing (the Control Condition).

P3: For any action, A, done from the false belief that A is permissible, S is blameworthy for having done A only if she was in control (directly or indirectly) of the false belief from which she did A (from P1, P2, P3).

P4: One is never directly in control of whether one believes or does not believe something; that is, any control that one has over one's beliefs and disbeliefs is only ever indirect.

P5: If one is blameworthy for something over which one had merely indirect control, then one's blameworthiness for it is itself merely indirect.

P6: One is indirectly blameworthy for something only if that thing was a consequence of something else for which one is directly blameworthy.

P7: For any action, A, done from the false belief that A is permissible, S is blameworthy for having done A only if there was something else, B, for which she is directly blameworthy and of which the false belief from which she did A was a consequence.

P8: In all cases of action, A, done from the false belief that A is permissible, there is no such thing, B.

C: In all cases of action, A, done from the false belief that A is permissible, S is not blameworthy for A[180].

If this argument is correct, then cases in which S does A from a false normative belief and is blameworthy for A would need to be cases in which she has undertaken an action, B, that has resulted in the false normative belief. To be

---

[179] P2 is endorsed by Ross: "Now suppose that of two men one does that which he mistakenly believes to be his objective duty, and the other does that which is his objective duty, believing it not to be so, we should regard the former as at least less blameworthy than the latter; and in fact *we should not regard the former as directly blameable for the act, but only, if at all, for previous acts by which he has blunted his sense of what is objectively right.*" (Ross (1939: 163-4), emphasis mine).

[180] This is an adaptation of Zimmerman's (2008: 175) argument for the view that ignorance of wrongdoing, and by extension false normative belief, can excuse in nearly all cases. The full argument involves a commitment to a further epistemic condition on blameworthiness – that blameworthy action requires that the agent believe that the action is wrong, and so all blameworthy actions are cases in which the agent akratically did something she believed was wrong (see Rosen (2002, 2004); Zimmerman (2008). This epistemic condition is dealt with in §2.

directly blameworthy for something, it must be under our direct control. So, when we do something wrong because of a false belief, we are only blameworthy for that wrong thing if there is a further thing, under our direct control, that we are blameworthy for. Another way to put this is that anything for which the agent is blameworthy will be part of a chain of blameworthiness, of which at the origin is something for which the agent is directly blameworthy. If Premise 8 is correct, then in cases in which wrongdoing is attributable to a false belief there is no action, B, that the belief can be attributed to. Premise 8 is supported by the putative observation that most mismanagement of belief is not the result of actions that are under the agent's direct control. The rest of this section discusses two strategies for resisting the Control Argument – by rejecting Premise 8, and by rejecting Premise 2 (the Control Condition).

We can reject Premise 8 by pointing to the apparent numerousness of cases of blameworthy false normative belief in which the agent's false belief is the result of behaviour for which she is directly blameworthy. This would show that cases in which false belief leads to wrong action, and the agent's false belief can be attributed to some blameworthy behaviour of the agent, are not, as Premise 8 claims, rare. Even if we accept the argument from control, cases in which false normative beliefs do not excuse do not seem particularly rare. Assuming that Premise 4 is correct, and we are not in direct control of our beliefs[181], there are nevertheless various actions related to belief management that we do have direct control over. If we can be blamed for these, then according to the Control Argument, we are indirectly blameworthy for any false normative beliefs resulting from them. Some examples of belief management actions that we do have direct control over include choosing where to focus our attention, deciding

[181] This is well supported in the literature, see for example Adler (2002); Helm (1994; Velleman (2000); Wedgwood (2002).

which books to read, and devoting more or less time to inquiry. If we are in direct control of these actions, then the beliefs that result from them would be under our indirect control. This would mean that even if we accept the argument above, we would be indirectly blameworthy for false normative beliefs, and thus also for the actions that result from them[182]. This plausibly generates many cases in which false belief does not excuse wrongdoing, for example[183]:

> *Deficient Investigation*. Bertha is about to operate a dangerous machine. There is an instruction manual, but she does not read it. She falsely believes that it functions much as similar machines she has operated. Unfortunately, she is incorrect, this machine is quite different. She operates the machine incorrectly and injures someone.

> *Benighting Act*. Carrie is mildly short-sighted, and usually wears glasses when driving. Today, she leaves the house in a hurry, and leaves them behind. Because of this, while driving she swerves to avoid a pothole, falsely believing that it is safe to do so. Unfortunately, she hits a child whom she did not see crossing the road (and whom she would have seen had she been wearing her glasses).

In both of these cases, the agent's false belief can be explained by earlier actions of belief management that were under her control. If this is correct then the control argument would seem to be compatible with there being a substantial number of cases in which false belief does not excuse. So, the Control Argument does not establish that false normative belief always excuses.

Another way to resist the Control Argument is by rejecting the Control Condition (Premise 2). An initial reason to reject the Control Condition is that it is not very

---

[182] In fact, Zimmerman (2008) is committed to the view that we are not blameworthy for any of these actions related to belief management, or the beliefs that result from them, due to his commitment to the claim that agents are only blameworthy for doing what they believe to be wrong. So long as the agent believes that she is managing her beliefs appropriately she is blameless. This claim is rejected in the following section.

[183] Both Deficient Investigation and Benighting Act are adaptations of cases discussed by Holly Smith (1983: 544).

plausible in the case of specifically *epistemic* blameworthiness, because belief is not under our direct control. So, if meeting the Control Condition was a necessary condition for any kind of blameworthiness, then we could not be blameworthy for any of our beliefs. This is an unwelcome result in so far as we think that epistemic blameworthiness is a useful notion. However, this argument for rejecting the Control Condition is too quick, since there are possible interpretations of control that are applicable to the epistemic domain. For example, Angela Smith's notion of rational accountability, according to which attitudes not under our direct volitional control are nevertheless under our 'rational' control in the sense of being sensitive to reasons (A. Smith 2005; 2008), and Scanlon's notion of "judgment-sensitivity" (1998: Ch. 6). Nevertheless, even if there is a plausible reading of the Control Condition that would be applicable to the epistemic domain, the many cases in which we typically hold agents blameworthy for actions not under their control (direct or indirect) exert pressure to resist a control condition on blameworthiness. If the Control Condition was true, then the agent would not be blameworthy in the following cases:

> *Hot Dog.* Alessandra has gone to pick up her children at their elementary school. As usual, Alessandra is accompanied by the family's dog, Sheba, who rides in the back of the van. The pickup has never taken long, so Alessandra leaves Sheba in the van while she goes to gather her children. This time, however, she is greeted by a tangled tale of misbehaviour, ill-considered punishment, and administrative bungling which requires several hours of indignant sorting out. During that time, Sheba languishes, forgotten, in the locked car. When Alessandra and her children finally make it to the parking lot, they find Sheba unconscious from heat prostration. (Sher, 2009: 24).

> *On the Rocks:* Julian, a ferry pilot, is nearing the end of a forty-minute trip that he has made hundreds of times before. The only challenge in this segment of the trip is to avoid some submerged rocks that jut out irregularly from the mainland. However, just because the trip is so routine, Julian's thoughts have wandered

to the previous evening's pleasant romantic encounter. Too late,
he realizes that he no longer has time to manoeuvre the ferry.
(Sher, 2009: 24).

In both of these cases there is no action under the agent's direct control that can be identified as something for which the agent can be blamed and that the wrongdoing is a consequence of. Forgetfulness and becoming distracted are not actions under our direct control. As Sher puts it when discussing these cases:

"We all know what it is to be assaulted by an urgent problem that drives all other thoughts from our minds, to emerge from a reverie into which we have no recollection of choosing to enter […]. When such things happen, there is simply no point at which we are conscious of choosing to allow them to happen." (2009: 25)

Although there are perhaps further actions that the agents could have undertaken that would have prevented their forgetting or becoming distracted, it is implausible that they are to blame for failing to take these actions, given that they had no reason to suppose that they would become distracted. There are three ways that one might respond to this conflict between cases of apparently blameworthy negligence and the Control Condition:

1. Retain the Control Condition and deny that the agents are blameworthy.

2. Reinterpret the notion of control such that the agents can be said to be in control of the lapses in belief management.

3. Give up the Control Condition.

I argue here that we give up the Control Condition. Although it commands some intuitive support, particularly from those who take blameworthiness to depend on free will (see Levy (2011); Pereboom (2005)), it is a fairly common position taken by various authors who otherwise defend diverging positions (see (Arpaly

(2002a; 2002b); Hieronymi (2004); Mason (2015); Sher (2006; 2009); Sie (2005); A. Smith (2008); Strawson (1962)).

One reason to give up the Control Condition is that in so far as the agents in Hot Dog and On the Rocks are blameworthy, their blameworthiness seems to be 'non-tracing'– it is not traceable to any earlier actions or omissions that were under their control (see H. Smith (1983; 2011)). Assuming that it was permissible for Alessandra to bring the dog to school, to leave her in the car for a short period of time, and to not take any further steps to remind herself about the dog, Alessandra's forgetting the dog is not attributable to any earlier action under her control and for which she could be blamed. The same is true for Julian. Julian did not choose to become distracted at that moment, and it is difficult to see what more he could have intentionally done to prevent his becoming distracted. Nevertheless, as pilot of the ferry he would typically be blamed for the accident. If Alessandra and Julian are blameworthy, then their blameworthiness cannot be traced to earlier failures that were under their control. This suggests that the control condition is not playing a significant role in deciding whether or not agents are blameworthy, and so there does not seem to be a significant cost to giving it up. Without the Control Condition, the Control Argument does not go through.

In summary, there are good reasons to reject both Premises 9 and 3, and therefore good reasons to reject the Control Argument.

2. *The Belief Condition*

Another way to argue for the view that blameless false normative belief always excuses is by appeal to a belief condition on original blameworthiness.

> *Belief Condition:* S is directly blameworthy for doing A only if she
>
> believes that A was wrong.

The Belief Condition applies to direct blameworthiness. Zimmerman endorses the Belief Condition as part of his Origination Thesis:

> *Origination Thesis*: Every chain of culpability is such that at its origin lies an item of behaviour for which the agent is directly culpable and which the agent believed, at the time at which the behaviour occurred, to be overall morally wrong. (Zimmerman 2008: 176).

Both the Belief Condition and the Origination Thesis hold that blameworthiness must originate with some behaviour that S believes is wrong. S is directly blameworthy for this behaviour, and indirectly blameworthy for any downstream consequences of this behaviour. The Belief Condition says that S believing that A is wrong is a necessary condition of S being directly blameworthy for A[184]. It cannot be a sufficient condition, since it is implausible that S is blameworthy for doing actions that she believes is wrong but is forced to do. If the Belief Condition is true, then if S does not believe that A is wrong, and A does not originate from any item that S believes is wrong, then S is not blameworthy (directly or indirectly) for A. This means that blameless false normative belief can always excuse wrongdoing, since so long as S does not believe that A is wrong, does A because of this false belief, and this belief is not itself blameless, then S is blameless for A. The following section (§2.1) argues that the Belief Condition is false, since consideration of normative uncertainty provides counterexamples to it. In these counterexamples S is blameworthy for

---

[184] Some have defended the related, but importantly different, claim that the belief that A is wrong is a necessary condition of A's being wrong. For example: "it is not in any way illegitimate…simply to define morally wrong action as action that is (i) of a certain kind…, and (ii) believed by its performer to be morally wrong" (Strawson (1986: 220)). I do not evaluate that claim here.

A even though she does not believe it is wrong, and her doing A does not originate from any behaviour that she believes is wrong.

One might think that Zimmerman's Origination Thesis expresses the view that belief is not only a necessary but also a sufficient condition for original blameworthiness, perhaps with an additional clause such as "and the agent freely and willingly does A" to exclude cases in which the agent is unwillingly forced to do A. However, as §2.2 argues, this sufficiency claim is also false. Rejecting this sufficiency claim, while strictly orthogonal to our main concern of whether false normative belief always excuses, serves to bolster the rejection of the Belief Condition by giving us reason to doubt that there is any interesting connection between whether S believes A is wrong, and whether S is directly blameworthy for A.

2.1 Belief is Not Necessary for Blameworthiness

Rejecting the Belief Condition requires us to show that believing that A is wrong is not a necessary condition of being directly blameworthy for A. Consideration of uncertainty gives us reason to think that there are possible cases in which agents are directly blameworthy but do not believe that they are doing wrong. This section presents some counterexamples that show this. Consider the following case[185]:

> *Drugs.* A doctor has a patient with a minor but not trivial skin complaint. The doctor has three drugs to choose from, and careful consideration of the medical evidence available to her has led her to the following opinions. One of drugs A or C will completely cure the skin condition, the other will kill the patient. In fact, A is the cure. However, crucially, she has no way of telling which of A and C will cure the patient, and which will kill him. However, she knows that drug B is very likely to relieve the condition but will not completely cure it (see Jackson (1991: 462)).

---

[185] This case was also mentioned briefly in the previous chapter (p. 145, fn. 162).

A plausible description of the doctor's state is that she suspends on whether it is A or C that is the cure, and that she does not believe that it would be wrong to prescribe either. Given this uncertainty, it seems clear that she would be blameworthy were she to prescribe either, even if we assume that she does not believe that it would be wrong to prescribe A or C. This is because prescribing either of drugs A or C would be reckless. Her available information does not allow her to be sure that she is not prescribing poison, and curing a minor skin condition is not worth the risk of death, people appropriately concerned with moral value do not gamble with it in this way. There is widespread agreement on this point (Bykvist (2014); Graham (2010); Sepielli (2009); Moller (2011); Ross (2006); Zimmerman (1997; 2008)[186]. However, if the Belief Condition were true then the doctor would not be blameworthy for prescribing A if she did not believe that choosing A was wrong. This seems like the wrong result.

One might be tempted to respond that it cannot be the case that were the doctor to prescribe A or C, this action could not be traced to any behaviour that she believes it wrong, because surely she believes that it would be reckless to choose A or C, given her information. However, the following adaptation of Drugs suggests that this is irrelevant:

> *Packaging.* A doctor has a patient with a minor but not trivial skin complaint. The doctor has three drugs to choose from. The medical evidence available to her indicates that one of drugs A or C will completely cure the skin condition, and the other will kill the patient. The medical evidence does not indicate which of A and C will cure the patient, and which will kill him. She knows that drug B is very likely to relieve the condition but will not completely cure

---

[186] Some do deny this. Harman (2015) and Weatherson (2014) argue that one should do, in all cases including uncertainty cases, is whatever the right thing to do is, and furthermore that to act so as to maximise expected moral value is to exhibit a blameworthy motivation – the motivation to do whatever is morally right. An alternative objection is offered by Gustafsson and Torpman (2014), who argue that seeking to act so as to maximise expected value is impossible because it would require us to make use of inter-theoretic comparisons of value, which they claim are impossible.

it. The doctor's favourite colour is blue, and Drug A is packaged in a pleasant shade of blue. She knows that the colour of the bottle is not evidence for the medicinal properties of the drug, but she nevertheless finds herself believing that the Drug A is the cure. She believes that it would be right to trust this belief, reasoning that if she believes it so strongly, she must be right. She does not know that her strong belief is influenced by the bottle's colour.

In this case, the doctor acts in a way that is blameworthily reckless, and this is not mitigated by the fact that her action cannot be traced to any behaviour that she believes is wrong. Prescribing A is not at all supported by the medical evidence, and the doctor knows this. The fact that she also believes that A is the correct choice, and that her belief that A is the correct choice is trustworthy does not help to mitigate the recklessness of the choice. This indicates that the belief that one's behaviour is wrong is not a necessary condition of direct blameworthiness. If this is right, then the Belief Condition is false.

A further reason to reject the Belief Condition is that it gives the wrong verdicts about agents who are insufficiently reflective. The Belief Condition says that agents are only directly blameworthy for doing wrong if they believe that they are doing wrong. This means that one way to avoid blame is to avoid having beliefs about whether or not what you are doing is wrong. This has the implausible result that unreflective wrongdoing is excusable. Consider the following example:

> *Bullying.* Nora spends a lot of time cruelly teasing one of her classmates, calling this classmate names, making unkind comments about their appearance, and excluding them socially. Nora does not reflect on whether or not her behaviour is wrong, although there have been ample opportunities for her to do so (she has attended numerous school workshops aimed at raising awareness about what constitutes bullying, for example,

although most of the time during these she was playing games on her phone). Consequently, Nora does not believe that her behaviour is wrong.

If the Belief Condition is true, then Nora is not blameworthy for bullying her classmate. She neither believes that the bullying is wrong, nor believes that not listening to the school workshops is wrong. However, not only does it seem that Nora is blameworthy for bullying, but we might think she is also blameworthy for not being adequately reflective about a matter of moral importance. This blame might be mitigated if Nora lacked the capacity to appreciate the wrongness of bullying, or to pay attention, but this is not the case here.

The considerations of uncertainty and unreflective wrongdoing discussed in this section should lead us to reject the Belief Condition, indicating that the Belief Condition does not give us reason to think that false normative belief always excuses. The following section addresses the related claim that the belief that A is wrong is sufficient for being blameworthy for A, arguing that this should also be rejected. Establishing that the belief that A is wrong is neither necessary nor sufficient for one's being directly blameworthy for A gives us significant reason to doubt that there is any interesting connection between the belief that one's action is wrong and direct blameworthiness for that action.

2.2 Belief is not Sufficient for Blameworthiness

The sufficiency claim that is the target of this section says that doing wrong while believing that A is wrong is sufficient for blameworthiness.

Sufficiency Claim: S is blameworthy for doing A if she believes

that A is wrong, or her doing A originates from some behaviour

that she believes is wrong[187].

Some have attempted to defend the Sufficiency Claim by appeal to the putative blameworthiness of the kind of psychological states the agent would need to have to do what she believed to be wrong. As Holly Smith puts it, for such agents to "have done the objectively right act would have been for them to do what they believed to be wrong. Such an act would necessarily have stemmed from a worse configuration of desires" than the configuration of desires involved in their doing the act they believed was right (H. Smith (1983: 559)). If this is right, then it is always more blameworthy to do what one believes to be wrong[188]. However, as various others have pointed out, it is not true that doing what one believes to be wrong always involves a more blameworthy configuration of desires than doing what one believes to be right (Arpaly (2002b); Harman (2011); M. Smith (1994)). Consider again Huck Finn (p. 135). In being unable to turn Jim in, Huck seems to be exhibiting a range of morally admirable concerns, despite doing what he believes to be wrong. For example, he desires to help a fellow human being, and acts so as to reduce another's suffering. In contrast, were he to do what he believes is right, and turn Jim in, this would exhibit a set of desires that does not include the morally admirable concerns for which he does act, and so it is not clear that this would be a better set of concerns[189].

---

[187] Again, we must exclude from this cases in which the agent believes that what she does is wrong, but cannot do otherwise because she is forced to do A.

[188] See also Ross (1939) for an endorsement of this view.

[189] Although, as Sliwa (2016) point out, the set of desires that Huck does exhibit is not necessarily the best possible set of desires. The set of desires he would have were he to also know the truth about what is right would arguably be better.

Another reason to reject the Sufficiency Claim is that it commits us to saying that agents who accept very demanding moral standards are blameworthy when they fail to meet them. For example:

> *Saint.* Sarah [is] a saintly person who has overly demanding moral standards. She believes that it is her moral obligation to exhaust herself in the service of others, although this is in fact (let us suppose) supererogatory. Early one morning her alarm awakes her and, contemplating yet another exhausting day of labour, she collapses back into bed and decides to sleep in an extra hour, feeling guilty about doing so because she believes that this is the wrong thing to do. (Zimmerman (2008: 199-200)).

If acting from belief that one is doing wrong is sufficient for blameworthiness, as the Sufficiency Claim says, this means that Sarah is to be blamed for her decision to stay in bed. This seems implausibly harsh[190]. Not only this, but the Sufficiency Claim commits us to saying that inversely akratic agents, such as Huck Finn, are blameworthy. As various others have pointed out, the conclusion that Huck is blameworthy seems to be the wrong result. He does the right thing, and he seems to do the right thing in response to genuinely salient moral considerations. The fact that he does not have the philosophical competence to reject the beliefs of his society and come to the correct beliefs about morality should not detract from the fact that his action is one that is praiseworthy, and for which he does not deserve blame.

A plausible alternative explanation of both the Saint and Huck Finn cases is that blameworthiness does not depend wholly on what the agent believes. This means that agents who have false normative beliefs, but nevertheless manage to act

---

[190] In Zimmerman's terms, Sarah deserves an 'accuse'. While excuses remove blame that would otherwise be incurred for a wrong act, accuses remove praise that would otherwise be incurred for a right act.

rightly need not be blameworthy, because blameworthiness cannot be wholly settled by whether or not the agent believes that what she is doing is wrong. So, assuming that Sarah's very demanding normative beliefs are false, we need not blame her when she fails to conform to them, even though she herself believes she is acting wrongly. Since her action is right and exhibits morally admirable features, we need not evaluate her as blameworthy. Similarly, Huck's belief that he is acting wrongly need not lead us to evaluate him as blameworthy, given that he acts rightly and his action exhibits various morally admirable features.

So, we ought to reject both the Belief Condition and the Sufficiency Claim. The belief that A is wrong is neither necessary nor sufficient for S to be (directly or indirectly) blameworthy for A. It is not necessary because there are possible cases of blameworthiness in which the agent does not believe she is doing wrong, for example some cases of wrongdoing under uncertainty; it is not sufficient because there are possible cases in which the agent believes she is doing wrong but is not blameworthy.

So far, this chapter has considered and rejected two ways of arguing for the view that false normative belief always excuses. The Control Argument was rejected on the grounds that we should reject the Control Condition on blameworthiness, and the argument from the Belief Condition was rejected on the grounds that the belief that A is wrong is neither necessary nor sufficiency for (direct or indirect) blameworthiness. Since these strategies fail, we should think that false normative belief can sometimes excuse. Given this, the final section considers in more depth the conditions that must be met for false normative belief to excuse, and further develops the positive view of the conditions of blameworthiness, first presented in Chapter 4 (p. 101): the view that agents are blameworthy when they fail to do what it is reasonable to expect of them.

3. What is it Reasonable to Expect?

The arguments of Part 3 thus far have established that false normative belief neither always nor never excuses. This means that false normative belief *sometimes* excuses. Chapter 4 presented an account of appraisal, the Reasonable Expectations View, that both permitted false normative belief to sometimes excuse, and allowed us to solve the puzzle of epistemic rationality presented in Chapter 1.

The Reasonable Expectations View says that false normative belief excuses wrongdoing when it is the case that, despite doing wrong, the agent has done what it is reasonable to expect of her, given her capacities and roles. This final section illustrates this by outlining how the account deals with two kinds of wrong action due to false normative belief that have motivated much of the debate over whether false normative belief can excuse: false normative belief due to culture (§3.1) and false normative belief due to psychological limitations (§3.2). Finally, I show how false normative belief that is a result of misleading evidence about what rationality requires can sometimes, but not always, excuse agents for failing to comply with normative requirements.

3.1 Culture

First, false normative beliefs acquired through culture have prompted much discussion of whether false normative belief can excuse. For example:

> *Ancient Slavery.* In the ancient Near East in the Biblical period the legitimacy of chattel slavery was simply taken for granted. No one denied that it was bad to be a slave, just as it is bad to be sick or deformed. The evidence suggests, however, that until quite late in antiquity it never occurred to anyone to object to slavery on grounds of moral or religious principle. So consider an ordinary Hittite lord. He buys and sells human beings, forces labour without compensation, and separates families to suit his

purposes. Needless to say, what he does is wrong. The landlord
is not entitled to do these things. But of course he thinks he is.
(Rosen (2002: 65))[191].

This is often cited as a clear case in which false normative belief does excuse[192], on the grounds that the slaveowner's beliefs about slavery are widespread and typically go unquestioned. However, I will argue that these facts do not imply that it is not reasonable to expect the slaveowner to question them. For example, it is consistent with the case that the slaveowner's beliefs are not supported by his evidence. Furthermore, we can assume that the slaveowner has ordinary capacities for recognizing and responding to morally salient considerations such as human suffering, and so that it is reasonable to expect him to perceive the suffering of his slaves and appreciate its moral significance. Furthermore, since he occupies the role of a slaveowner, it is reasonable to expect him to take greater care than others in investigating whether or not what he is doing is right, since if he is wrong about this he risks a great moral harm[193]. In summary, it is not clear that he is blameless because it is not clear that he lacks the capacity to respond appropriately to the salient reasons of the situation. Furthermore, it is plausible that he occupies a role that means we should expect him to take more care than usual in determining whether he is right. For it to be the case that one's culture could excuse wrongdoing, it would need to be the case that one really did not have the capacities to avoid the wrongdoing. Arpaly's case of Solomon is a plausible such example:

> *Solomon*. Consider the case of Solomon, a boy who lives in a
>
> small, isolated farming community in a poor

---

[191] See also Rosen's similar case of Smith, a typical 1950s sexist father (Rosen (2002: 66-69)), discussed previously (Ch. 4, p.113).

[192] See, for example, (Rosen (2002); Slote (1982)), as well as similar examples discussed by (Donagan (1977); Wolf (1987); Zimmerman (2008)).

[193] See also (Lockhart (2000); Moller (2011); Sepielli (2009)) on moral risk.

country. Solomon believes that women are not half as competent as men when it comes to abstract thinking, or at least are not inclined toward such thinking. Solomon's evidence for his belief is the fact that all the women in his community, despite his attempts to engage them in learned conversation, seem to discuss nothing but gossip, family, and manual work, that the few people in his community who are interested in abstract thinking are all men, that no one he knows of has ever doubted that women are worse abstract thinkers, and that the community's small, outdated library contains abstract work written by men only. Solomon's belief is false, but it is not *particularly* or *markedly* irrational. It is not particularly irrational because Solomon is not exposed to striking counter evidence to it, and he is exposed to a consensus and "expert" opinion in its favour (just think how many of our own everyday beliefs are grounded simply on expert opinion and lack of clear counterexamples). True, if Solomon were to think more carefully, he might find reasons to change his mind, but in his case "thinking more carefully" would involve having the intelligence of John Stuart Mill without the advantage of having known at least one woman who is inclined toward abstract thinking. (Arpaly (2002: 103-4)).

Unlike the slaveowner, and Smith the 1950s father (Ch. 4, p. 114), it is plausible that Solomon does lack the capacity to recognize and respond to reasons that would count against his sexist behaviour and beliefs. As he is described, all of his evidence supports the view that women are not good at abstract thinking, and he has little to no opportunity to acquire evidence to the contrary. For example, he

has no opportunity to interact with any women who are good at abstract thinking, and plenty of chances to interact with women who confirm the view that they are not. This isolation limits his epistemic capacities, and this limited nature of Solomon's capacities makes it much more plausible that his false normative beliefs would excuse any sexist beliefs and behaviour[194].

Huck Finn (p. 135) is another example of an agent whose false normative belief can be attributed to his culture, and the case is usually used to show that agents with false normative belief can sometimes be praiseworthy when they act against these false normative beliefs[195]. One might worry that the Reasonable Expectations View threatens that conclusion. For example, one might think that what it is reasonable to expect Huck to do, given his belief that freeing Jim is stealing, is to *not* free Jim. This would mean that Huck was blameworthy, rather than praiseworthy, for freeing Jim.

However, this is not the correct interpretation of the view. In fact, according to the Reasonable Expectations View, Huck is not blameworthy, since it is not the case that he has failed to do what it is reasonable to expect of him. It is true that he can be expected to continue to hold the false normative belief that he ought not free Jim, since it would be very difficult for Huck to know that he was doing the right thing in this situation. However, it is also reasonable to expect him to respond to moral considerations such as Jim's well-being, since he has ordinary capacities to recognize and respond to morally salient considerations. This means that it is reasonable to expect Huck to free Jim, and so in meeting these expectations he does not deserve blame. However, Arpaly and others who emphasise Huck's praiseworthiness have perhaps been too positive in evaluating Huck as praiseworthy. According to the Reasonable Expectations View, he does

---

[194] This is consistent with both Arpaly's and Fricker's (2007: 33-4) reading of the case.
[195] See Arpaly (2002b).

not exceed his expectations, but merely meets them. As Sliwa points out, it would be even better if he had come to know that his action was right (Sliwa (2016)). In summary, the Reasonable Expectations View does not in general take false normative belief acquired through one's culture to excuse, although it may do sometimes if it is the case that one's culture severely restricts one's epistemic capacities to avoid wrongdoing.

3.2 Psychological Limitations

Cases of false normative belief acquired through psychological limitations are also often discussed in relation to the question of whether false normative belief can excuse. For example, another of Rosen's characters, Bonnie, commits a relatively minor moral wrong (stealing a cab by queue-jumping in front of a stranger) in full knowledge of the non-moral facts of the situation, out of a lack of understanding and appreciation of the significance of moral reasons (2002: 76). Due to features of her psychology that are beyond her control[196], she has no capacity to appreciate the significance of moral reasons. We might describe her as a psychopath. Psychopaths are typically understood as lacking the capacity to understand the distinction between 'moral' and 'conventional' norms, and thus being unable to understand and appreciate the significance of moral reasons (Cleckley (1955); Blair et al. (2005)). This psychological limitation typically leads psychopaths to falsely believe that morally wrong actions are permissible. The Reasonable Expectation View of appraisal says that what it is reasonable to expect is determined by one's capacities, so this lack of capacity to appreciate moral reasons affects which reasons it is reasonable to expect psychopaths to respond to, and therefore how they should be appraised when they act. It is not

---

[196] Rosen implausibly attributes Bonnie's lack of capacity to 'a virus'; current psychology suggests amygdala damage would be a more likely explanation. This is nevertheless beside the point – the interest of the case is that Bonnie's behaviour is grounded in robust psychological features that she has no power to change.

reasonable to expect them to appreciate and respond appropriately to moral reasons if they lack the capacity to do so. When they do wrong, and their wrongdoing can be attributed to false normative belief, then these are cases in which false normative belief can excuse.

Nevertheless, not all cases involving psychopaths doing wrong from false normative belief are cases in which false normative belief excuses. There are some possible cases involving agents with psychological limitations in which it is reasonable to expect the agent to have done more to respond to the reasons than she has. To illustrate these possibilities, the rest of this sub-section outlines some additional possible cases involving Rosen's character of Bonnie the psychopath (Rosen (2002: 76)). The first case is one in which Bonnie is excused:

> *Queue*. Bonnie is waiting in the rain, under an umbrella, for a taxi at a currently empty taxi rank. She is in no particular hurry. Soon she is joined by pregnant woman who needs to get to the hospital as soon as possible. A taxi pulls up. At no point does it occur to her that it might be morally right to offer the taxi to the pregnant woman, and morally wrong not to do so. Were this to be explained to her, she would find it difficult to care about these reasons.

In this case, Bonnie is excused from wrongdoing. By stipulation, she completely lacks the capacity to appreciate the force of moral reasons and so respond to any of the reasons that favoured offering the taxi to the pregnant woman, and there are no other reasons that she could have been expected to appreciate that would have indicated that giving the taxi to the pregnant woman was the right thing to do. So, according to the Reasonable Expectation View of blameworthiness, she is blameless and this is a case in which false normative belief excuses. The second case is one in which it is not clear that Bonnie is excused:

> *Murder*. Bonnie is waiting in the rain, under an umbrella, for a taxi at a currently empty taxi rank. She is in no particular hurry. Soon she is joined by pregnant woman who needs to get to the hospital as soon as possible. To lessen the boredom of the wait, she decides to murder the pregnant woman and hide her body

> before the first taxi arrives. At no point does it occur to her that it might be morally wrong to do this. Were this to be explained to her, she would be unable to appreciate the moral force of these reasons. She also knows that her meta-ethical views about what is valuable are idiosyncratic, and she is willing to entertain the idea that she might be wrong about what is valuable.

In this case, it is much less clear that Bonnie is excused from wrongdoing. The key difference between Queue and Murder is that in Murder the moral stakes are much higher, and it is reasonable to expect Bonnie to know this because she knows that her meta-ethical views are idiosyncratic and she knows that on rival meta-ethical views, murder is very morally wrong. As in Queue, Bonnie cannot appreciate the force of moral reasons. However, Bonnie's own perception of the relative force of reasons, moral and otherwise, are not the only considerations that could speak in favour of her refraining from murdering the pregnant woman. It is possible for her to appreciate facts about what other people consider to be moral reasons, in much the same way that it is possible for any of us to appreciate facts about what other cultures consider to be polite or impolite, even if we cannot fully appreciate the force of these facts. Furthermore, she is able to appreciate considerations that bear on how likely she is to be correct in her meta-ethical views about the normative force of moral reasons, and there are considerations of moral risk[197].

It is plausible that agents can be expected to refrain from bringing about disvalue unless they have a good reason to think that the risk is worth taking. We can thus suppose that agents are in general blameworthy for recklessly risking moral disvalue. An important feature of Queue that differentiates it from Murder is that the moral transgression involved is relatively subtle, such that it is plausible that

---

[197] For example, Greenspan (2016) argues that psychopaths can act morally despite their psychological limitations.

someone not sensitive to moral reasons would not even notice that it might be a moral transgression. Murder is not like this – it is a striking example of a moral transgression, and explicitly prohibited by laws in most jurisdictions. Even though Bonnie does not believe that moral transgressions ought to be avoided, we can suppose that she knows that most people who do believe that morality is valuable believe that murder is very morally wrong, and therefore that if it turns out that they are right and she is wrong, then she would be doing something very wrong indeed by murdering the old woman. As has been established, the mere belief that one is not doing wrong is insufficient to excuse wrongdoing (see §3.2, above). In Murder, Bonnie does not necessarily need the capacity to appreciate the force of salient moral considerations herself to realize that she is risking doing something that is very wrong, since she knows that it might turn out that she is mistaken about the normative insignificance of moral value. This means that whether or not Bonnie is blameworthy for the murder will turn on whether or not her action is a reckless risk of moral disvalue. If it is, then her false normative belief does not excuse the wrongdoing because it does not mitigate the recklessness of her action, for which we can assume she is blameworthy.

Whether or not Bonnie's action is blameworthily reckless depends on the epistemic status of her meta-ethical beliefs. By committing the murder, she avoids mild boredom while waiting for a taxi, but nothing more. However, she risks significant moral disvalue if she is wrong about the relative importance of morality and her own amusement. We could represent the relevant values of her options in the following way:

	Moral Value	Hedonistic Value	Total value	
			Moral value is important	Moral value is not important

Murder	-1000	+20	-980	+20
	0	-20	-20	-20

This risk *might* be non-reckless if Bonnie had very good epistemic reasons to believe that her meta-ethical views are correct. Specifically, if she had very good evidence that supports a high credence that she is right in valuing her own well-being over what others assign moral value. If this was the case, then the risk would not be reckless, and so Bonnie would not be blameworthy for the murder. However, as Chapter 5 (§2) noted, evidence that supports very high credences in any philosophical beliefs is extremely rare, and Bonnie's meta-ethical belief about the importance of moral value is no exception. This means that if she does murder the pregnant woman, she would need to be very sure, and have very good reason to be very sure, that her view of the comparative importance of moral and prudential value was the right one. To compare, were this balance of costs and benefits inverted, such that she gained a significant amount of prudential value from committing what would be only a minor moral wrong, it would be more plausible that she would count as doing what can be reasonably expected of her.

Finally, consider a third case, which focuses on how Bonnie's role affects what it is reasonable to expect of her:

> *Ethics Committee.* Bonnie is serving on a medical ethics committee, where she needs to evaluate complicated ethical cases and make decisions that significantly impact the lives of others. This is a task Bonnie is particularly unsuited to, and she knows this – for example, she knows that what others cite as morally salient considerations fail to move her. She ends up making some significant moral blunders which lead to some very bad decisions that negatively affect the lives of the patients whose cases are under discussion.

In this case, Bonnie's role means that what it is reasonable to expect of her, even given her limited capacities, is considerably higher than it otherwise would be. The fact that it is very difficult for her to meet these higher than usual expectations might mitigate her blameworthiness somewhat, but it is unlikely to be sufficient to excuse her completely, since it does not make it such that she does do what can be reasonably expected of her. What is more, it is plausible that she is blameworthy not only for making catastrophic decisions while sitting on the ethics committee, but also for accepting the position on the ethics committee – since she knows that her capacities for moral reasoning are limited.

4. Summary

This chapter has defended the view that false normative belief sometimes excuses by arguing against the view that blameless false normative belief always excuses. Two arguments for the view that blameless false normative belief always excuses were rejected (the Control Argument, and the argument from the Belief Condition). Finally, it was clarified under what conditions false normative belief does excuse, with reference to some often discussed cases.

Conclusion

I have aimed to solve the Puzzle of Rational Requirement, a puzzle that arises from the possibility of rational mistakes about what rationality requires. The strategy defended here connected three distinct issues: how our theories of what rationality requires should accommodate misleading evidence, the relationship between complying with requirements and deserving particular appraisals, and whether normative ignorance can excuse.

Rational false beliefs about what rationality requires present a puzzle for our theories of rational requirement. When agents have misleading evidence about what rationality requires, does rationality require that they do what they (falsely) believe they are rationally required to do? If yes, then rationality requires that they violate the true requirements of rationality. If no, then rationality requires that they be level incoherent, and violate the Enkratic Principle.

In Part 1, I considered this puzzle and rejected a variety of potential solutions, including embracing the puzzle as a dilemma of epistemic rationality, indexing the apparently conflicting requirements in such a way that avoids the apparent conflict, and denying that rational false beliefs about what rationality requires are possible. In Chapter 1 I argued that Dilemmism not only fails to solve the puzzle, but leads to triviality given Standard Deontic Logic, and that the indexing strategies so far proposed are unable to solve the puzzle, since none are able to draw on an appropriate, non-question begging distinction that can be used to separate apparent conflicts of requirement within the same normative domain. In Chapter 2 I showed that the puzzle cannot be solved by denying that rational false beliefs about what rationality requires are impossible, because there is no good argument for this thesis, the Impossibility Thesis. Specifically, I considered

and rejected arguments from: the more general claim about the impossibility of false beliefs, the nature of rational competence, our putative justificatory assets for the truth about what rationality requires, and the Enkratic Principle itself. Instead, I suggested we should investigate the second option, that rationality requires that agents with misleading evidence supporting false beliefs about what rationality requires are required to violate the Enkratic Principle, which I claimed is plausible once we distinguish the distinct evaluations of requirement and appraisal.

Part 2 further developed this solution. According to it agents with rational false beliefs about what rationality requires are required to avoid the recommendations of those false beliefs. Thus they end up with a set of beliefs that violates the Enkratic Principle. Chapter 3 argued for this solution on the grounds that it is the least theoretically costly of the remaining available solutions. Nevertheless, even though the Enkratic Principle is not a rational requirement, it still seems that an agent may be judged negatively for violating it. By contrast, if agents do follow the Enkratic Principle and end up violating a requirement of rationality, then it seems that they are to be positively evaluated in at least some circumstances, in particular if it is the case that they do what it is reasonable to expect of them. In Chapter 4 I provide an account of such appraisals of agents by looking at when agents are blameworthy for their beliefs. I examined and rejected a range of accounts of when agents are blameworthy for their beliefs before recommending my own reasonable expectation view. Applied to the original puzzle case, this showed that agents who follow the misleading recommendations of rational false beliefs about what rationality requires are not rational, but can be excused from blame for failing to comply with the requirements of rationality, because these agents meet, and perhaps exceed, what it is reasonable to expect of them in managing their beliefs.

This solution is committed to the claim that false normative belief can sometimes excuse violations of requirements. However, whether false normative belief can excuse at all is controversial. Part 3 argued that agents can be excused when it is the case that they have done what it is reasonable to expect of them, where this is determined by their capacities and roles. It defended the view that false normative belief can excuse in these cases against the two rival views: that false normative belief can never excuse, and that false normative belief always excuses – at least when it is itself blameless. The main arguments for each of these views are presented and rejected, leaving the only remaining possible answer to the question of whether false normative belief can excuse is that it can sometimes excuse. This result vindicates the solution to the puzzle set out in Part 2.

To return to the title, is it OK to make mistakes when it comes to one's normative beliefs? One sense of this answer is affirmative: rational false normative beliefs can sometimes serve as an excuse for further violations of requirements that these mistakes might lead one to make. Another sense of this answer is negative: having mistaken normative beliefs cannot change the fact that you would be violating your requirements in following their misleading recommendations. The upshot is that following rational false beliefs about what rationality requires can excuse you from blame, but not make you rational, for violating requirements of rationality.

Bibliography

Adler, J.E., 2002. Akratic Believing? *Philosophical Studies*, 110(1), pp.1–27.

Alexander, D.J., 2013. The Problem of Respecting Higher-Order Doubt. *Philosophers' Imprint*, 13, pp.1–12.

Alston, W., 1989. *Epistemic Justification*, Cornell University Press.

Alston, W.P., 1985. Concepts of Epistemic Justification. *The Monist*, 68(2), pp.57–89.

Alvarez, M., 2010. *Kinds of Reasons: An Essay in the Philosophy of Action*, Oxford University Press.

Anderson, E., 2015. Moral Bias and Corrective Practices. *Proceedings and Addresses of the American Philosophical Association*, 89, pp.21–47.

Anderson, E., 2014. Social Movements, Experiments in Living, and Moral Progress: Case Studies from Britain's Abolition of Slavery. *Lindley Lecture, University of Kansas*.

Anderson, E., 2016. The Social Epistemology of Morality: Learning from the Forgotten History of the Abolition of Slavery. In M. Fricker & M. Brady, eds. *The Epistemic Life of Groups*. Oxford University Press, pp. 75–94.

Arpaly, N., 2002a. Moral Worth. *Journal of Phliosophy*, 99(5), pp.223–245.

Arpaly, N., 2002b. *Unprincipled Virtue: An Inquiry Into Moral Agency*, Oxford University Press.

Arpaly, N. & Schroeder, T., 2013. *In Praise of Desire*, Oxford University Press.

Austin, J.L., 1979. *Philosophical Papers* J. O. Urmson & M. Warnock, eds., Oxford University Press.

Barnbaum, D.R., 2008. *The Ethics of Autism: Among Them, but Not of Them*, Indiana University Press.

Baron, M., 2007. Excuses, Excuses. *Criminal Law and Philosophy*, 1(1), pp.21–39.

Basu, R. & Schroeder, M., 2019. Doxastic Wronging. In B. Kim & M. McGrath, eds. *Pragmatic Encroachment in Epistemology*. Routledge, pp. 181–205.

Besson, C., 2010. Propositions, Dispositions and Logical Knowledge. In M. Bonelli & A. Longo, eds. *Quid Est Veritas? Essays in Honour of Jonathan Barnes*. Bibliopolis, pp. 233–268.

Björnsson, G., 2017. Explaining (Away) the Epistemic Condition on Moral Responsibility. In P. Robichaud & J. W. Wieland, eds. *Responsibility - The Epistemic Condition*. Oxford University Press, pp. 141–164.

Björnsson, G. & Finlay, S., 2010. Metaethical Contextualism Defended. *Ethics*, 121(1), pp.7–36.

Blair, R.J.R., Mitchell, D.G.V. & Blair, K., 2005. *The Psychopath: Emotion and the Brain*, Blackwell.

Boghossian, P., 2003. Blind Reasoning. *Aristotelian Society Supplementary Volume*, 77(1), pp.225–248.

Bolinger, R.J., The Rational Impermissibility of Accepting Racial Generalizations. *Synthese*, pp.1–17.

BonJour, L., 1998. *In Defense of Pure Reason*, Cambridge University Press.

BonJour, L., 1985. *The Structure of Empirical Knowledge*, Harvard University Press.

Bradley, D., 2019. Are There Indefeasible Epistemic Rules? *Philosophers' Imprint*, 19(3), pp.1–19.

Broome, J., 1999. Normative Requirements. *Ratio*, 12(4), pp.398–419.

Broome, J., 2013. *Rationality Through Reasoning*, Wiley-Blackwell.

Brown, J., 2018. *Fallibilism: Evidence and Knowledge*, Oxford University Press.

Brown, J., 2008. Subject-Sensitive Invariantism and the Knowledge Norm for Practical Reasoning. *Noûs*, 42(2), pp.167–189.

Brownstein, M., 2014. Rationalizing Flow: Agency in Skilled Unreflective Action. *Philosophical Studies*, 168(2), pp.545–568.

Buss, S., 2014. Experiments In Vivo, In Vitro, and In Cathedra. *Ethics*, 124(4), pp.860–881.

Bykvist, K., 2014. Evaluative Uncertainty, Environmental Ethics, and Consequentialism. In A. Hiller, I. Ramona, & L. Kahn, eds. *Consequentialism and Environmental Ethics*. Routledge, pp. 122–136.

Calhoun, C., 1989. Responsibility and Reproach. *Ethics*, 99(2), pp.389–406.

Carroll, L., 1895. What the Tortoise Said to Achilles. *Mind*, 4(14), pp.278–280.

Casullo, A., 1988. Revisability, Reliabilism, and A Priori Knowledge. *Philosophy and Phenomenological Research*, 49(2), pp.187–213.

Cholbi, M., 2009. The Murderer at the Door: What Kant Should Have Said. *Philosophy and Phenomenological Research*, 79(1), pp.17–46.

Christensen, D., 2009. Disagreement as Evidence: The Epistemology of Controversy. *Philosophy Compass*, 4(5), pp.756–767.

Christensen, D., 2007. Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals. *Oxford Studies in Epistemology*, 2, pp.3–31.

Christensen, D., 2010a. Higher-Order Evidence. *Philosophy and Phenomenological*

Research, 81(1), pp.185–215.

Christensen, D., 2004. *Putting Logic in its Place: Formal Constraints on Rational Belief*, Oxford University Press.

Christensen, D., 2010b. Rational Reflection. *Philosophical Perspectives*, 24(1), pp.212–140.

Clarke, D.S., 1984. Ignoring Available Evidence. *Southern Journal of Philosophy*, 22(4), pp.452–467.

Cleckley, H., 1955. *The Mask of Sanity: an attempt to clarify some issues about the so-called psychopathic personality*, St. Louis: Mosby.

Clifford, W., 1877. The Ethics of Belief. *Contemporary Review*, 29, pp.289–310.

Coates, A., 2012. Rational Epistemic Akrasia. *American Philosophical Quarterly*, 49(2), pp.113–124.

Cohen, L.J., 1981. Can Human Irrationality be Experimentally Demonstrated? *Behavioral and Brain Sciences*, 4, pp.317–370.

Cohen, S., 1984. Justification and Truth. *Philosophical Studies*, 46(3), pp.279–295.

Cohen, S. & Comesaña, J., 2013. Williamson on Gettier Cases and Epistemic Logic. *Inquiry*, 56(1), pp.15–29.

Condorcet, M.J.A.N. de Caritat., M. de, 1995. An Essay on the Application of Analysis to the Probability of Decisions Rendered by a Plurality of Votes. In I. McLean & A. Urken, eds. *Classics of Social Choice*. Ann Arbor: University of Michigan Press, pp. 91–113.

Conee, E. & Feldman, R., 2004. *Evidentialism: Essays in Epistemology*, Oxford University Press.

Dancy, J., 2004. *Ethics Without Principles*, Oxford University Press.

Dancy, J., 2000. *Practical Reality*, Oxford University Press.

Davidson, D., 2004. Deception and Division. In *Problems of Rationality*. Oxford: Oxford University Press, pp. 199–213.

Davidson, D., 2001. How Is Weakness of the Will Possible? In *Essays on Actions and Events*. Oxford: Oxford University Press, pp. 21–43.

DePoe, J.M., 2011. Defeating the Self-Defeat Argument for Phenomenal Conservativism. *Philosophical Studies*, 152(3), pp.347–359.

Dolan, M.C. & Fullam, R.S., 2010. Moral/Conventional Transgression Distinction and Psychopathy in Conduct Disordered Adolescent Offenders. *Personality and Individual Differences*, 49(8), pp.995–1000.

Donagan, A., 1977. *The Theory of Morality*, University of Chicago Press.

Dummett, M.A.E., 1974. The Justification of Deduction. In M. Dummett, ed. *Truth and Other Enigmas*. Oxford University Press, pp. 290–318.

Elga, A., 2013. The Puzzle of the Unmarked clock and the New Rational Reflection Principle. *Philosophical Studies*, 164(1), pp.127–139.

Fantl, J. & McGrath, M., 2002. Evidence, Pragmatics, and Justification. *Philosophical Review*, 111, pp.67–94.

Fantl, J. & McGrath, M., 2009. *Knowledge in an Uncertain World*, Oxford University Press.

Feldman, R., 1988a. Epistemic Obligations. *Philosophical Perspectives*, 2, pp.235–256.

Feldman, R., 2007. Reasonable Religious Disagreements. In L. Antony, ed. *Philosophers Without Gods: Meditations on Atheism and the Secular Life*. Oxford University Press, pp. 194–214.

Feldman, R., 2005. Respecting the Evidence. *Philosophical Perspectives*, 19(1), pp.95–119.

Feldman, R., 1988b. Subjective and Objective Justification in Ethics and Epistemology. *The Monist*, 71(3), pp.405–419.

Feldman, R. & Conee, E., 1985. Evidentialism. *Philosophical Studies*, 48(1), pp.15–34.

Field, C., 2019a. It's OK to Make Mistakes: Against the Fixed Point Thesis. *Episteme*, 16(2), pp.175–185.

Field, C., 2019b. Recklessness and Uncertainty: Jackson Cases and Merely Apparent Assymmetry. *Journal of Moral Philosophy*, 16(4), pp.391–419.

Field, H., 2000. A Prioricity as an Evaluative Notion. In P. Boghossian & C. Peacocke, eds. *New Essays on the a Priori*. Oxford University Press, pp. 117–149.

Field, H., 2009. Epistemology without Metaphysics. *Philosophical Studies*, 143(2).

Fitelson, B., 2012. Evidence of Evidence is Not (Necessarily) Evidence. *Analysis*, 72(1), pp.85–88.

FitzPatrick, W.J., 2008. Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge. *Ethics*, 118(4), pp.589–613.

Frege, G., 1897. Logik. In H. Hermes, F. Kambartel, & F. Kaulbach, eds. *Nachgelassene Schriften und Wissenschaftlicher Briefwechseln*. Meine (1969), pp. 137–163.

Fricker, M., 2007. *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press.

Gibbard, A., 2005. Truth and Correct Belief. *Philosophical Issues*, 15(1), pp.338–

350.

Gibbons, J., 2013. *The Norm of Belief*, Oxford University Press.

Goldberg, S.C., 2017. Should Have Known. *Synthese*, 194(8), pp.2863–2894.

Goldberg, S.C., 2018. *To the Best of Our Knowledge: Social Expectations and Epistemic Normativity*, Oxford University Press.

Goldman, A.I., 1986. *Epistemology and Cognition*, Harvard University Press.

Graham, P.A., 2010. In Defense of Objectivism About Moral Obligation. *Ethics*, 121(1), pp.88–115.

Greco, D., 2014. A Puzzle About Epistemic Akrasia. *Philosophical Studies*, 167(2), pp.201–219.

Greenspan, P., 2016. Responsible Psychopaths Revisited. *The Journal of Ethics*, 20(1), pp.265–278.

Guerrero, A.A., 2007. Don't Know, Don't Kill: Moral Ignorance, Culpability, and Caution. *Philosophical Studies*, 136(1), pp.59–97.

Gustafsson, J.E. & Torpman, O., 2014. In Defence of My Favourite Theory. *Pacific Philosophical Quarterly*, 95(2), pp.159–174.

Harman, E., 2011. Does Moral Ignorance Exculpate? *Ratio*, 24(4), pp.443–468.

Harman, E., 2016. Morally Permissible Moral Mistakes. *Ethics*, 126(2), pp.366–393.

Harman, E., 2015. The Irrelevance of Moral Uncertainty. In *Oxford Studies in Metaethics, Volume 10*. Oxford University Press, pp. 53–79.

Harman, G., 1986. *Change in View*, MIT Press.

Hasan, A., 2013. Phenomenal Conservatism, Classical Foundationalism, and

Internalist Justification. *Philosophical Studies*, 162(2), pp.119–141.

Hawthorne, J. & Bovens, L., 1999. The Preface, the Lottery, and the Logic of Belief. *Mind*, 108(430), pp.241–264.

Hawthorne, J. & Srinivasan, A., 2013. Disagreement Without Transparency: Some Bleak Thoughts. In *The Epistemology of Disagreement*. Oxford: Oxford University Press, pp. 9–30.

Helm, P., 1994. *Belief Policies*, Cambridge University Press.

Hieronymi, P., 2004. The Force and Fairness of Blame. *Philosophical Perspectives*, 18(1), pp.115–148.

Holroyd, J., 2012. Responsibility for Implicit Bias. *Journal of Social Philosophy*, 43(3), pp.274–306.

Horowitz, S., 2014. Epistemic Akrasia. *Noûs*, 48(4), pp.718–744.

Horty, J.F., 2003. Reasoning with Moral Conflicts. *Noûs*, 37(4), pp.557–605.

Huemer, M., 2007. Compassionate Phenomenal Conservatism. *Philosophy and Phenomenological Research*, 74(1), pp.30–55.

Huemer, M., 2011a. Does Probability Theory Refute Coherentism? *Journal of Philosophy*, 108(1), pp.35–54.

Huemer, M., 2011b. Phenomenal Conservatism and Self-Defeat: a Reply to DePoe. *Philosophical Studies*, 156(1), pp.1–13.

Hughes, N., 2019. Dilemmic Epistemology. *Synthese*, 196(10), pp.4059–4090.

Husak, D., 2005. On the Supposed Priority of Justification to Excuse. *Law and Philosophy*, 24(6), pp.557–594.

Ichikawa, J.J. & Jarvis, B.W., 2013. *The Rules of Thought*, Oxford University Press.

Jackson, F., 1991. Decision-Theoretic Consequentialism and the Nearest and Dearest Objection. *Ethics*, 101(3), pp.461–482.

Jeshion, R., 2000. On the Obvious. *Philosophical and Phenomenological Research*, 60(2), pp. 333-355.

Johnson-King, Z., 2018. Accidentally Doing the Right Thing. Forthcoming in *Philosophy and Phenomenological Research*, pp. 1-21.

Johnson-King, Z., Praiseworthy Motivations. Forthcoming in *Noûs*, pp. 1-23.

Joyce, J., 2009. Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In F. Huber & C. Schmidt-Petri, eds. *Degrees of Belief*, pp. 263-297. Synthese.

Kahneman, D., 2011. *Thinking, Fast and Slow*, London: Penguin.

Kahneman, D., Slovic, P. & Tversky, A., 1982. *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press.

Kant, I., 1797. On a Supposed Right to Lie from Philanthropy. In *Practical Philosophy*. New York: Cambridge University Press, pp. 611–15.

Kelly, T., 2006. Evidence. *Philosophy Compass*, pp. 933-955.

Kelly, T., 2002. The Rationality of Belief and Other Propositional Attitudes. *Philosophical Studies*, 110(2), pp. 163-196.

Kelp, C., 2016. Justified Belief: Knowledge First-Style. *Philosophy and Phenomenological Research*, 92(3), pp. 79-100.

Kiesewetter, B., 2011. "Ought" and the Perspective of the Agent. *Journal of Ethics and Social Philosophy*, 5(3), pp. 1-24.

Kiesewetter, B., 2013. *The Normativity of Rationality*.

Kitcher, P., 1980. A Priori Knowledge. *Philosophical Review*, 89(1), pp. 3-23.

Kolodny, N., 2005. Why Be Rational? *Mind*, 114(455), pp. 509-563.

Kornblith, H., 1983. Justified Belief and Epistemically Responsible Action. *Philosophical Review*, 92, pp. 33-48.

Korsgaard, C., 1986. The Right to Lie: Kant on Dealing with Evil. *Philosophy and Public Affairs*, 15(4), pp.325–349.

Kvanvig, J., 2003. Propositionalism and the Perspectival Character of Justification. *American Philosophical Quarterly*, 40(1), pp. 3-18.

Kvanvig, J.L., 2014. *Rationality and Reflection: How to Think About What to Think*, Oxford University Press.

Kvanvig, J.L., 1984. Subjective Justification. *Mind*, 93(369), pp. 71-84.

Lackey, J., 2016. What Is Justified Group Belief? *Philosophical Review Recent Issues*, 125(3), pp. 341-396.

Lasonen-Aarnio, M., Enkrasia or Evidentialism? Forthcoming in *Philosophical Studies*, pp. 1-36.

Lasonen-Aarnio, M., 2014. Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research*, 88(2), pp.314–345.

Lasonen-Aarnio, M., 2010. Unreasonable Knowledge. *Philosophical Perspectives*, 24(1), pp. 1-21.

Lemmon, E.J., 1965. Deontic Logic and the Logic of Imperatives. *Logique Et Analyse*, 8(29), pp. 39-61.

Levy, N., 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*, Oxford University Press.

Lewis, D., 2004. Letters to Priest and Beall. In *The Law of Non-Contradiction*, pp. 176-177. Oxford University Press.

Lewis, D., 1982. Logic for Equivocators. *Noûs*, 16(3), pp. 431-441.

Littlejohn, C., A Plea for Epistemic Excuses. Forthcoming in F. D. Julien Dutant, ed. *The New Evil Demon Problem*. Oxford University Press. Available at: https://philpapers.org/archive/LITAPF.pdf.

Littlejohn, C., 2012. *Justification and the Truth-Connection*, Cambridge University Press.

Littlejohn, C., 2011. Reasons and Belief's Justification. In A. Reisner & A. Steglich-Petersen, eds. *Reasons for Belief*, pp. 111-130. Cambridge University Press.

Littlejohn, C., 2015. Stop Making Sense? On a Puzzle about Rationality. *Philosophy and Phenomenological Research*, 93(3), pp. 257-272.

Littlejohn, C., 2009. The Externalist's Demon. *Canadian Journal of Philosophy*, 39(3), pp. 399-434.

Lockhart, T., 2000. *Moral Uncertainty and its Consequences*, Oxford University Press.

Lord, E., 2018. *The Importance of Being Rational*, Oxford University Press.

Lycan, W.G., 2013. Phenomenal Conservatism and the Principle of Credulity. In C. Tucker, ed. *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism*, pp. 293-305. Oxford University Press.

Mantel, S., 2018. *Determined By Reasons*, Routledge.

Markovits, J., 2010. Acting for the Right Reasons. *Philosophical Review*, 119(2), pp. 201-242.

Mason, E., 2015. Moral Ignorance and Blameworthiness. *Philosophical Studies*, pp. 3037-3057.

Mathiesen, K., 2006. The Epistemic Features of Group Belief. *Episteme*, 2(3), p. 161-175.

Moller, D., 2011. Abortion and Moral Risk. *Philosophy*, 86(3), pp. 425–443.

Montero, B.G., 2016. *Thought in Action: Expertise and the Conscious Mind*, Oxford University Press.

Nelkin, D.K., 2009. Responsibility, Rational Abilities, and Two Kinds of Fairness Arguments. *Philosophical Explorations*, 12(2), pp. 151-165.

Nguyen, C.T., Cognitive Islands and Runaway Echo Chambers: Problems for Epistemic Dependence on Experts. Forthcoming in *Synthese*, pp. 1-19.

Nguyen, C.T., Echo Chambers and Epistemic Bubbles. Forthcoming in *Episteme*, pp. 1-21.

Owens, D., 2002. Epistemic Akrasia. *The Monist*, 85(3), pp. 381-397.

Parfit, D., 2011. *On What Matters: Two-Volume Set*, Oxford University Press.

Pereboom, D., 2005. *Living Without Free Will*, Cambridge University Press.

Peregrin, J., 2014. *Inferentialism: Why Rules Matter*, Palgrave Macmillian.

Pigozzi, G., Hansen, J. & van der Torre, L., 2007. Ten Philosophical Problems in Deontic Logic. In *Normative multi-agent systems. Dagstuhl Seminar Proceedings*, pp. 1-26.

Pittard, J. & Worsnip, A., 2017. Metanormative Contextualism and Normative Uncertainty. *Mind*, 126(501), pp. 155-193.

Pollock, J., 1979. A Plethora of Epistemological Theories. In G. Pappas, ed.

Justification and Knowledge, pp. 93-115. D. Reidel.

Pollock, J.L., 1986. *Contemporary Theories of Knowledge*, Hutchinson.

Priest, G., 1985. Contradiction, Belief and Rationality. *Proceedings of the Aristotelian Society*, 86(1), pp. 99-116.

Priest, G., 2005. *Doubt Truth to Be a Liar*, Oxford University Press.

Priest, G., 2006. *In Contradiction: A Study of the Transconsistent*, Oxford University Press.

Putnam, H., 1978. There is at Least One A Priori Truth. *Erkenntnis*, 13(1), pp. 153-170.

Railton, P., 2014. The Affective dog and its Rational Tale: Intuition and Attunement. *Ethics*, 124(4), pp.813-159.

Raz, J., 1975. *Practical Reason and Norms*, Hutchinson.

Raz, J., 2005. The Myth of Instrumental Rationality. *Journal of Ethics and Social Philosophy*, 1(1), pp. 2-28.

Reed, N., McLeod, P. & Dienes, Z., 2010. Implicit Knowledge and Motor Skill: What People Who Know How to Catch Don't Know. *Consciousness and Cognition*, 19(1), pp. 63-76.

Reisner, A., 2013. Is the Enkratic Principle a Requirement of Rationality? *Organon F*, 20(4), pp. 436-462.

Richman, K.A. & Bidshahri, R., 2018. Autism, Theory of Mind, and the Reactive Attitudes. *Bioethics*, 32(1), pp. 43-49.

Robin, Z., 2016. Attributability, Accountability, and Implicit Bias. In J. Saul & M. Brownstein, eds. *Implicit Bias and Philosophy, Volume 2: Moral Responsibility, Structural Injustice, and Ethics*, pp. 62-89. Oxford University Press.

Roche, W., 2018. Is Evidence of Evidence Evidence? Screening-Off vs. No-Defeaters. *Episteme*, 15(4), pp. 119-124.

Roche, W. & Shogenji, T., 2013. Confirmation, Transitivity, and Moore: the Screening-Off Approach. *Philosophical Studies*, (3), pp. 1-21.

Rosen, G., 2002. Culpability and Ignorance. *Proceedings of the Aristotelian Society*, 103(1), pp. 61-84.

Rosen, G., 2004. Skepticism About Moral Responsibility. *Philosophical Perspectives*, 18(1), pp. 295-313.

Ross, J. et al., 2006. Rejecting Ethical Deflationism*. *Ethics*, 116, pp.742–768.

Ross, W.D., 1939. *Foundations of Ethics*, Oxford University Press.

Russell, G., 2017. Logic Isn't Normative. *Inquiry: An Interdisciplinary Journal of Philosophy*, pp. 1-18.

Sayre-McCord, G., 1986. Deontic Logic and the Priority of Moral Theory. *Noûs*, 20(2), pp. 179-197.

Scanlon, T., 1998. *What We Owe to Each Other*, Belknap Press of Harvard University Press.

Schapiro, T., 2006. Kantian Rigorism and Mitigating Circumstances. *Ethics*, 117(1), pp.32–57.

Schroeder, M., 2009. Means-End Coherence, Stringency, and Subjective Reasons. *Philosophical Studies*, 143(2), pp. 223-248.

Schroeder, M., 2004. The Scope of Instrumental Reason. *Philosophical Perspectives*, 18(1), pp. 337-364.

Schwarz, W., 1970. Kant's Refutation of Charitable Lies. *Ethics*, 81(1), pp.62–67.

Schwitzgebel, E., 2006. The Unreliability of Naive Introspection. *Philosophical Review*, 117(2), pp. 245-273.

Sepielli, A., 2009. What to Do When You Don't Know What to Do. *Oxford Studies in Metaethics*, 4, pp. 5-28.

Sher, G., 2006. Out of Control. *Ethics*, 116(2), pp. 285-301.

Sher, G., 2009. *Who Knew?: Responsibility Without Awareness*, Oxford University Press.

Shoemaker, D.W., 2011. Psychopathy, Responsibility, and the Moral/Conventional Distinction. *Southern Journal of Philosophy*, 49(1), pp. 99-124.

Sie, M., 2005. *Justifying Blame: Why Free Will Matters and Why It Does Not*, Rodopi.

Skipper, M. & Steglich-Petersen, A., 2019. Group Disagreement: A Belief Aggregation Perspective. *Synthese*, 196(10), pp. 4033-4058.

Skorupski, J., 2010. *The Domain of Reasons*, Oxford University Press.

Sliwa, P., 2016. Moral Worth and Moral Knowledge. *Philosophy and Phenomenological Research*, 93(2), pp. 393-418.

Sliwa, P. & Horowitz, S., 2015. Respecting All the Evidence. *Philosophical Studies*, 172(11), pp.2835–2858.

Slote, M., 1982. Is Virtue Possible? *Analysis*, 42(2), pp. 70-76.

Smith, A.M., 2008. Control, Responsibility, and Moral Assessment. *Philosophical Studies*, 138(3), pp.367–392.

Smith, A.M., 2005. Responsibility for Attitudes: Activity and Passivity in Mental Life. *Source: Ethics*, 115(2), pp.236–271.

Smith, H., 1983. Culpable Ignorance. *Philosophical Review*, 92(4), pp. 543-571.

Smith, H., 2011. Non-Tracing Cases of Culpable Ignorance. *Criminal Law and Philosophy*, 5(2), pp. 115-146.

Smith, M., 1994. *The Moral Problem*, Blackwell.

Smithies, D., 2015. Ideal Rationality and Logical Omniscience. *Synthese*, 192(9), pp. 2769-2793.

Smithies, D., 2012. Moore's Paradox and the Accessibility of Justification. *Philosophy and Phenomenological Research*, 85(2), pp. 273-300.

Sosa, E., 1980. The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge. *Midwest Studies in Philosophy*, 5(1), pp. 3-26.

Sperber, D., 2011. Why Do Humans Reason? Arguments for an Argumentative Theory. *Behavioral and Brain Sciences*, 34(2), pp. 57-74.

Srinivasan, A., 2015a. Are We Luminous? *Philosophy and Phenomenological Research*, 90(2), pp.294–319.

Srinivasan, A., 2015b. Normativity without Cartesian Privilege. *Philosophical Issues*, 25(1), pp.273–299.

Stalnaker, R., 1991. The Problem of Logical Omniscience, I. *Synthese*, 89(3).

Steinberger, F., 2019. Three Ways in Which Logic Might Be Normative. *Journal of Philosophy*, 116(1), pp. 425-440.

Steup, M., 1999. A Defense of Internalism. In L. Pojman, ed. *The Theory of Knowledge: Classical and Contemporary Readings, 2nd edition*, pp. 373-384. Wadsworth Publishing.

Strawson, G., 1986. *Freedom and Belief*, Oxford University Press.

Strawson, P.F., 1962. Freedom and Resentment. *Proceedings of the British Academy*, 48, pp. 1-25.

Sturgeon, S., 1998. Visual Experience. *Proceedings of the Aristotelian Society*, 72(2), pp. 179-200.

Summerfield, D.M., 1991. Modest A Priori Knowledge. *Philosophy and Phenomenological Research*, 51(1), pp. 39-66.

Sutton, J., 2005. Stick to What You Know. *Noûs*, 39(3), pp. 359-396.

Sutton, J., 2007. *Without Justification*, MIT Press.

Swain, M., 1979. Justification and the Basis of Belief. In G. Pappas, ed. *Justification and Knowledge*, pp. 25-49. D. Reidel.

Sylvan, K., 2015. What Apparent Reasons Appear to Be. *Philosophical Studies*, 172(3), p. 587-606.

Thagard, P., 1982. From the Descriptive to the Normative in Psychology and Logic. *Philosophy of Science*, 49, pp. 24–42.

Titelbaum, M.G., 2015a. How to Derive a Narrow-Scope Requirement from Wide-Scope Requirements. *Philosophical Studies*, 172(2), pp. 535-542.

Titelbaum, M.G., 2015b. Rationality's Fixed Point (Or: In Defense of Right Reason). In *Oxford Studies in Epistemology*. pp. 253–294.

Turri, J., 2010. On the Relationship Between Propositional and Doxastic Justification. *Philosophy and Phenomenological Research*, 80(2), pp. 312-326.

Tversky, A. & Kahneman, D., 1974. Judgment Under Uncertainty: Heuristics and Biases. *Science*, 185(4157), pp. 1124-1131.

Tversky, A. & Kahneman, D., 1993. Probabilistic Reasoning. In A. Goldman, ed. *Readings in Philosophy and Cognitive Science*, pp. 43-68. MIT Press.

Unger, P., 1986. The Cone Model of Knowledge. *Philosophical Topics*, 14(1), pp. 125-178.

Velleman, D., 2000. On the Aim of Belief. In D. Velleman, ed. *The Possibility of Practical Reason*, pp. 244-281. Oxford University Press.

Watson, G., 1996. Two Faces of Responsibility. *Philosophical Topics*, 24(2), pp. 227-248.

Way, J., 2011. The Symmetry of Rational Requirements. *Philosophical Studies*, 155(2), pp. 227-239.

Way, J. & Whiting, D., 2016. If You Justifiably Believe that You Ought to Φ, You Ought to Φ. *Philosophical Studies*, 173(7), pp. 1873-1895.

Way, J. & Whiting, D., 2017. Perspectivism and the Argument from Guidance. *Ethical Theory and Moral Practice*, 20(2), pp. 361-374.

Weatherson, B., 2019. *Normative Externalism*, Oxford University Press.

Weatherson, B., 2014. Running Risks Morally. *Philosophical Studies*, 167(1), pp. 141-163.

Wedgwood, R., 2002. The Aim of Belief. *Philosophical Perspectives*, 36(16), pp. 267-297.

Wedgwood, R., 2017. *The Value of Rationality*, Oxford University Press.

Weinrib, J., 2008. The Juridical Significance of Kant's "Supposed Right to Lie." *Kantian Review*, 13(1), pp.141–170.

Whiting, D., 2014. Keep Things in Perspective: Reasons, Rationality, and the A Priori. *Journal of Ethics and Social Philosophy*, 8(1), pp. 1-22.

Williams, B., 1981. Persons, Character, and Morality. In J. Rachels, ed. *Moral Luck*, pp. 1-19. Cambridge University Press.

Williamson, T., 2017. Ambiguous Rationality. *Episteme*, 14(3), pp.263–274.

Williamson, T., 2013. Gettier Cases in Epistemic Logic. *Inquiry : An Interdisciplinary Journal of Philosophy*, 56(1), pp. 1-14.

Williamson, T., 2011. Improbable Knowing. In T. Dougherty, ed. *Evidentialism and its Discontents*, pp. 147-165. Oxford University Press.

Williamson, T., Justifications, Excuses, and Sceptical Scenarios. In J. Dutant & D. Dohrn, eds. *The New Evil Demon Problem*. Oxford University Press, pp. 1689–1699.

Williamson, T., 2002. *Knowledge and its Limits*, Oxford: Oxford University Press.

Williamson, T., 2014. Very Improbable Knowing. *Erkenntnis*, 79(5), pp. 971-999.

Wolf, S., 1987. Sanity and the Metaphysics of Responsibility. In F. D. Schoeman, ed. *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, pp. 46-62. Cambridge University Press.

Wood, A., 2008. *Kantian Ethics*, Cambridge University Press.

Worsnip, A., "Ought"-Contextualism Beyond the Parochial. *Philosophical Studies*, 176(11), pp. 3099-3119.

Worsnip, A., 2018. The Conflict of Evidence and Coherence. *Philosophy and Phenomenological Research*, 96(1), pp. 3-44.

Wright, C., 2004a. Intuition, Entitlement and the Epistemology of Logical Laws. *Dialectica: International Journal of Philosophy of Knowledge*, 58(1), pp.155-175.

Wright, C., 2004b. On Epistemic Entitlement: Warrant for Nothing? *Aristotelian Society Supplementary Volume*, 78, pp. 167-212.

Wright, C., 2014. On Epistemic Entitlement (II). In *Scepticism and Perceptual Justification*, pp. 213-248. Oxford University Press.

Zimmerman, M.J., 2014. *Ignorance and Moral Obligation*, Oxford University Press.

Zimmerman, M.J., 2008. *Living with Uncertainty: The Moral Significance of Ignorance*, Cambridge University Press.

Zimmerman, M.J., 1997. Moral Responsibility and Ignorance. *Ethics*, 107(3), pp. 410–426.