The Role of Emotion in Practical Rationality

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy (Ph.D)

In the Faculty of Humanities

2014

Beccy Simpson

The School of Social Sciences
## Contents

Abstract..............................................................................................................................5  
Declaration.........................................................................................................................6  
Copyright Statement .......................................................................................................6  
Acknowledgements.........................................................................................................7  
Dedication.........................................................................................................................8  
Introduction......................................................................................................................9  

Chapter One  
Practical Rationality Consists in Doing What One Has Most Reason to Do......13  
1 Practical rationality is concerned with *normative* reasons.........................14  
2 The justificatory force of normative reasons is derived from their promoting a valued state of affairs..........................................................18  
   2.1 The ‘having’ relation is motivational, not epistemic.........................23  
   2.2 The same consideration can be a reason both for and against the same action..........................................................27  
3 The reasons that an agent has are determined by her values.................32  
   3.1 Practical rationality consists in doing whatever it is that one has most reason to do ..........................................................38  
      3.1.1 Reason’s role is as judge in both theoretical and practical rationality 39  
      3.1.2 The ‘having’ relation for both theoretical and practical reasons is motivational.........................................................40  
      3.1.3 Theoretical and practical agency are both expressed through rational guidance ..........................................................42  

Chapter Two  
An Agent is Practically Rational to the Extent that Her Action is Guided by the Norm of Doing What She Has Most Reason to Do ................48  
1 The relevant norm when assessing the rationality of an agent is that of the subjective ought..........................................................49  
2 The norm of practical agency is regulative..........................................................57  
3 The capacity to engage in practical reasoning is a requirement of practical agency............................................................................61  
4 Behaviour does not have to be the direct result of practical reasoning in order to be a candidate for rational action..........................67  
5 Reason-tracking mechanisms must be under the agent’s *regulative* rational control: this is in what rational agency consists.................................................76
Chapter Three
Practical Reasoning

1. Practical reasoning is conducted from the deliberative stance.
2. Practical reasoning has the distinct phenomenology of being under our control.
3. Practical reasoning is deliberation about reasons for action, not desire: the role of desires in practical deliberation is not as reasons but as tracking reasons.
4. The fundamental role of desires in practical reasoning.

Chapter Four
Emotions

Chapter Five
Emotions Provide the Access to Normative Reasons and thus Enable Practically Rational Action

1. Emotions as Explanatory Reasons
2. Emotions as Normative Reasons
   2.1 Emotion as a normative reason per se
   2.2 Emotion-proper properties as normative reasons
3. Emotions provide ‘quick and dirty’ access to normative reasons
   3.1 Emotions focus an agent’s attention on her reasons for action
   3.2 Preparing the body for action
4. Emotions provide us with access to concepts
5. Emotions as the only path of access to particular reasons

Chapter Six
The Roles of Emotion in Practical Reasoning

1. Emotions instigate practical reasoning
2. Emotions enable the agent to reach the right decision
   2.1 Emotions rule out some options from the beginning
   2.2 Emotions provide the necessary access to all of the reasons for action which an agent has
   2.3 Emotions determine the normative weight an agent assigns to a reason for action
3. Emotions Close Reasoning
4. Conclusion

Chapter Seven
Weak-Willed Action is Necessarily Irrational; Akratic Action is Not
Two types of ‘incontinence’: akrasia and weakness of will. Weakness of will as failure to act upon one’s intention when one ought to

An agent’s intention ought to be immune to reconsideration unless there are good grounds

Huckleberry Finn is a weak agent, not an akratic agent

Inverse weakness of will is, by definition, not responsive to reasons and thus, by definition, not rational

Inverse akrasia can be responsive to reasons and thus can be rational

Chapter Eight
Loss of Self-Control
Loss of Control as a defence to murder
The new defence
The concept of loss of control in case law
The type of control lost
Loss of control does not imply not knowing what one is doing
The agent is still an intentional agent
Loss of control rather than lack of control
The provocation gives rise to an intention, not just an inclination
The intention is out of reach of reason
Skewing the epistemic landscape
Conclusion

Bibliography
Abstract

In this thesis I argue that emotion is integral to practical rationality, contrary to the dominant tradition that has held that emotions are irrational and dangerous disruptive influences that we’d be better off without. In Chapter 1 I argue that practical rationality consists in doing what one has most normative reason to do, and in Chapter 2 that an agent is practically rational to the extent that she responds to her reasons; this is how she guides her actions in line with the norm of doing what she has most reason to do. This can be done in ways other than by the employment of practical reasoning. In Chapter 3 I argue for a picture of practical reasoning that stands against the division of emotion and rationality. This account makes room for the overwhelming evidence that challenges the traditional view of emotions as the enemy of practical rationality. Chapter 4 gives a brief overview of the philosophical literature of emotions, and their place in practical rationality. In Chapter 5 I argue that emotions provide us with the necessary access to our reasons for action which we need in order to be able to respond to them, and thereby to be practically rational. Further, as I argue in Chapter 6, emotions play vital roles in the process of practical reasoning itself. Thus practical rationality would not be better off without emotion. In Chapter 7 I argue that we should distinguish between two types of incontinent action (acting against ones all things considered judgement about what one has most reason to do) and that one of these – weakness of will – is necessarily irrational, but the other – akrasia – is not. In Chapter 8 I apply my thesis to the question in the practical domain of what it means to ‘lose self-control’ in the context of killing in response to a provocation, which is a defence to murder. I argue that the ‘control’ that is lost is the regulative guiding control characteristic of the reason-responder. Understanding practical agency as reason-responsiveness, and understanding the role that emotions play within it as per my thesis, enables this coherent understanding. Thus I am arguing for neither a pro-emotion nor anti-emotion view of the role of emotion in practical rationality. Emotions should not be seen as either ‘for’ rationality nor ‘against’ rationality: they are simply part of rationality.
Declaration
No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement
The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and she has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectualproperty pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on presentation of Theses.
Acknowledgements

My first thanks go Joel Smith, Tom Smith, and the late Peter Goldie, all three of whom have been my first supervisor at some point throughout the genesis of this thesis. My transition from MRes to Ph.D is largely creditable to Peter and Joel’s unceasing encouragement, and my continuation in the light of Peter’s death to Tom.

Secondly, I thank the Department of Philosophy at the University of Manchester as a whole for the provision of such a supportive and inclusive community. Most members of the department have at some point had input into my education, but in particular I thank Helen Beebee and Julian Dodd, who have taught me since I was a first year undergraduate and to whom credit is certainly due for my continued enthusiasm for the subject.

My fellow Ph.D candidates, in particular Lucy Tomlinson and Moritz Mueller, have created a lively, enthusiastic and encouraging graduate community, for which I have been most grateful and appreciative.

I also wish to thank those numerous other members of the Philosophical community who have encouraged me; in particular Marion Godman, Richard Holton, Derek Matravers, Timothy Schroeder, Barry Smith, and Ronnie de Sousa.

There are numerous people who have one way or another provided me with support, including Christopher Warne, Noel Stewart, Tony & Barbara Simpson, past and present MRes students, Martha Evans, Christopher Beach, Sara Kallock, Tom Loghran and Maeve MacPherson.
Dedication

For Peter,

and for Kath
Introduction

In my young days, a fellow with moral fiber had to avoid the emotions. Emotion was for sissies. A little moist fervor might be allowed when the school orchestra cranked the national anthem up to speed, but generally emotion was a furtive business, best concealed. Life was to be lived dry of eye and stiff of lip. More or less unconsciously, but following a dominant tradition of Western thought, people held that emotions were disruptive and irrational, dangerous interferences with the smooth operation of reason and deliberation. But such a negative assessment is very much on the retreat in philosophy and psychology now … “Emotional intelligence” has become a catchphrase, quite apart from its association with the soggier parts of the therapy industry (Blackburn, 2001).

The negative assessment of emotions may certainly be ‘on the retreat’ in both philosophy and psychology, but still they are seen as something separate to reason. The term ‘emotional intelligence’ may be proving to have more substance than is seen in self-help books, but still it is seen as an addition to ‘regular’ intelligence. We have reason, rationality and intellectual capacities on one side, and an emerging acceptance that emotions, on the other side, may have a positive contribution to make; this is still a picture where emotion is something separate to reason. Emotions might have a positive contribution, but still they are the base, animal, irrational side of our nature. It is this division of emotion and rationality that this thesis challenges. Our emotional capacities are not just residues of our evolutionary heritage, things that do some useful things sometimes, but that, ideally, we would transcend and be capable of being fully and perfectly rational without. They are not things with which, regretfully, we have been saddled and need to make the most of. On the contrary, practical rationality would be worse off, not better off, without emotion. This is the central claim of this thesis.

The roots of the view of emotions as being ‘disruptive and irrational, dangerous inferences with the smooth operation of reason and deliberation’ lie in the dominant philosophical approach to practical rationality. Practically rationality is identified with practical reasoning; conscious deliberation about what one ought to do, and since emotions are held to be dangerous influences to reasoning, they
are held to be barriers to rational action. Both the Humean and Kantian pictures of practical reasoning set up reason and desire, or emotion, against each other. The Kantian picture sees reason's job as controlling desire and emotion so as to guide action in line with reason; the Humean picture sees reason's job as the servant of desire and emotion, so as to aid them in achieving the ends that they set for the agent. Emotions either disrupt reasoning, or else they direct it, but either way we'd be better off without them when it comes to being practically rational, even if being so is impossible for us mere humans.

I thus begin by challenging this picture. In my first chapter I argue that practical rationality consists in doing what one has most reason to do, and that this is not the exclusive domain of practical reasoning. As I argue in my second chapter, although the capacity to engage in practical reasoning is vital for an agent to be able to be practically rational, she can be practically rational by roots other than conscious deliberation. Thus practical rationality is not identified with practical reasoning. I argue that an agent is practically rational to the extent that she guides her behaviour in line with the norm of doing what she has most reason to do, and that she does this to the extent that she is responsive to her reasons for action. This picture stands in opposition to the extreme of seeing practical agency as consisting in being responsive to practical reasoning, which follows the Kantian tradition of seeing reason's role as controlling emotion, and also in opposition to the picture of practical agency as consisting in something that is separate from practical reasoning, which follows the Human tradition as seeing reason's role as being the servant of emotion. On my picture of practical agency consisting in being responsive to reasons, the capacities for practical reasoning and the capacities for emotion and desire work together to achieve the aim of guiding an agent's action in line with her reasons for action.

In establishing a picture of practical rationality that does not already assume a separation of reason and emotion, I have made room for the mounting evidence from psychology that emotions help us to be practically rational. This is a controversial claim. It is becoming increasingly accepted that emotions promote our “ecological rationality”, our ability to cope successfully with our environments
(Mameli, 2004) - no one doubts, for example, that fear plays an important part in mammals’ successfully avoiding predators. But that they contribute to our *practical rationality* - our ability to make the right *choices* – is, at least philosophically, highly contentious. The dominant philosophical positions, as I have already said, see emotion and reason as being conflicting forces when it comes to our making decisions about how to act; as desire either leading reason astray, or as reason needing to control desire. That emotions contribute to our *ecological* rationality is consistent with either of these positions, for their being evolutionarily adaptive does not mean that they are beneficial to us in our current lives (Goldie, 2008, p. 154ff). Although emotions still have a function in enabling us to behave in ways that coincide with what we ought to do in those situations that are sufficiently similar to those in which they developed (for example, staying a safe distance from a lion), they can still be a disabling influence in our being able to do what we have most reason to do in situations where that is not what emotions prime us to do, and these situations make up the majority of those in which we find ourselves in the modern world. In particular, emotions are held to be a disruptive influence on our practical reasoning, for they prime us for automatic behaviours, and lead us astray with their motivational pulls (*ibid.*, p. 149). They are thus seen as something we’d be better off without, from the perspective of practical rationality. We may not be *able* to escape their effects, but we should *try* to do so to the best of our abilities so that we can act in accordance with what we have *most* reason to, mitigate their misleading us into acting in the way they have primed us for. Emotions, on the traditional view, might occasionally be useful, but for the main we’d be more practically rational without their influence.

My proposal is that we would *not* be more practically rational without the influence of emotions; we would be *less* so. Despite their ability to mislead us, they are something without which we could not guide our behaviour in line with the norm of practical rationality. This is because, as I argue in my fifth chapter, emotions provide us with the necessary *access* to our reasons for action which we need in order to be able to respond to them, and thereby to be practically rational. Further, as I argue in my sixth chapter, emotions play vital roles in the process of practical
reasoning itself. Not only would our ability to be practical rational be worse off without emotion, practical reasoning - the paradigmatic way by which human agents can be practically rational - would also be worse off without emotion. Thus practical rationality would not be better off without emotion.

This is not to deny that emotions can be the cause of irrational behaviour, and in my seventh and eighth chapters I argue that, in the light of my thesis of the role of emotion in practical rationality, we can make better sense of two sorts of irrationality in particular. In Chapter 7 I argue that we should distinguish between two types of incontinent action (acting against ones all things considered judgement about what one has most reason to do) and that one of these – weakness of will – is necessarily irrational, but the other – akrasia – is not. In Chapter 8 I apply my thesis to the question in the practical domain of what it means to ‘lose self-control’ in the context of killing in response to a provocation, which is a defence to murder. I argue that the ‘control’ that is lost is the regulative guiding control characteristic of the reason-responder. Understanding practical agency as reason-responsiveness, and understanding the role that emotions play within it as per my thesis, enables this coherent understanding.

Thus I am arguing for neither a pro-emotion nor anti-emotion view of the role of emotion in practical rationality. Emotions should not be seen as either ‘for’ rationality nor ‘against’ rationality: they are simply part of rationality.
Chapter One

Practical Rationality Consists in Doing What One Has Most Reason to Do.

In this thesis I argue that emotion is integral to practical rationality, contrary to the dominant tradition that has held that emotions are irrational and dangerous disruptive influences that we’d be better off without. In this first chapter I argue that a practical rationality consists in doing what one has most reason to do. In Chapter 2 I argue an agent is practically rational to the extent that she responds to her reasons; this is how she guides her actions in line with the norm of doing what she has most reason to do. This can be done in ways other than by the employment of practical reasoning; an agent can be practically rational – can do what she has most reason to do, and do so for those reasons - by other routes. In Chapter 3 I argue for a picture of practical reasoning that does not posit the existence of a ‘rational self’ that is separate from emotion; even this paradigmatic way of aiming to do what we have most reason to do for those reasons should not be seen as promoting the division of emotion and rationality. This account of practical rationality makes room for the overwhelming evidence that challenges the traditional view of emotions as the enemy of practical rationality, for which I shall argue in Chapter 4 onwards. In Chapters 7 and 8 I argue that my thesis of the role of emotions within practical rationality has applications in to concepts which have been seen as involving rational conflicts between emotion and reason, specifically akrasia and loss of self-control.

To begin, then, in this first chapter I must argue for my picture of practical rationality as consisting in doing what one has most reason to do. To understand this claim we need to be clear on a number of issues. We need to know with what sort of ‘reasons’ we are concerned, what it is for an agent to ‘have’ a reason, and why having ‘most’ reason is what makes an action rational. In section 1 I argue that we should draw a distinction between three sorts of reasons for action - normative, operative and explanatory – and it is the first of these sorts of reasons – normative reasons – that are relevant to the statement ‘what one has most
reason to do'. This is because normative reasons are the only sort of reasons that can transmit justification. Their justification, I shall argue in section 2, is derived from the fact that they promote something of value. Thus something is a normative reason to \( \Phi \) if it counts of favour of \( \Phi \)-ing by virtue of the fact that \( \Phi \)-ing will promote the thing of value from which it derives its normative force. In section 3 I argue that what makes it the case that an agent 'has' a particular normative reason is that it promotes something which she values, thus the reasons for action which an agent 'has' are determined by her values. The having relation is motivational, in that a reason for action is a reason that an agent has because it can motivate her by virtue of the fact that it promotes something which she values. In section 4 I argue that this entails that practical rationality consists in doing whatever it is that one has most reason to do, since what one has most reason to do will be that which is most justified, and practical agents aim to do that which is most justified.

1 Practical rationality is concerned with normative reasons

The type of ‘reasons’ relevant to the statement “what one has most reason to do” are so-called ‘normative’ reasons, which are considerations that count in favour of an action (Dancy 2000; Parfit 2001; Raz 1975, 1999; Scanlon 1998). This class of reasons contrasts with what are often called ‘explanatory’ reasons, which are those which explain why things are as they are, or were. Explanatory reasons explain the states of all sorts of things, for example the reason for my car’s being dirty is that it has not been washed in many months. As well explaining the states of inanimate objects, explanatory reasons account for the behaviour of non-human animals, for example the reason my cat hisses at the vacuum cleaner is that he is frightened; his being frightened explains his hissing. It can also account for states of affairs involving human agents. For example, we might say that Jennifer Ennis won the competition because her new training programme gave her an advantage over the other contestants, or that I went through the red light because I wasn’t paying due attention to the traffic lights. Neither Ennis’ training programme nor my lack of attention are considerations that count in favour of the states of affairs that resulted – lack of attention cannot justify going through a red
light - but they do explain why those states of affairs, including agents’ actions, came about.

Normative reasons, on the other hand, by counting in favour of an action, justify performing that action. My car’s being dirty might be a reason that justifies my washing it. Ennis’ success with her new training programme would count in favour of her continuing with that programme. Of course there may be competing normative reasons that detract from the course of action being that which the agent ought to pursue (perhaps, despite her success, Ennis has been miserable in her new programme, and thus the best thing for her to do is to change it), so that a consideration counts in favour of a particular action does not necessarily provide enough normative force to make performing it justified – i.e. to make it what one has most reason to do - but it does count in favour of that action.

Normative reasons need also to be distinguished from ‘operative’ reasons (Scanlon, 1998), which are purported facts that an agent treats as normative reasons – i.e. as being considerations that count in favour of a certain action - but which might or might not be normative reasons - propositions that in fact are true and in fact count in favour of the action for which they are treated as being so.¹ To borrow Pamela Hieronymi’s (2011, pp. 410-11) example, an agent might leave a room because she thinks a meeting is over (the purported fact which she thinks is true and which she thinks counts in favour of her leaving the room), but in fact it is not over, thus the reason for which she left the room was, although operative, was not normative (since there was no true proposition on the basis of which she left the room). The operative reason for which she acted was ‘the meeting is over’, but not only was this not a normative reason, it can also not be an explanatory reason either: we cannot use ‘the meeting is over’ to explain her action by saying ‘she left the room because the meeting was over’, since the meeting was not in fact over. The natural way of explaining the agent’s departure in this case is to appeal to her belief that the meeting was over, but this is a fact

¹ I will use the term ‘operative’ reasons, but the term ‘motivating’ reasons is also used by, for example, Dancy (2000), Parfit (2001), Schroeder (2008).
about her psychology, and not the reason for which the agent left the room. The explanatory reason, then, is a fact about her psychology, but the reason for which she acted (‘the meeting is over’) is not a fact about her psychology. Operative reasons and explanatory reasons are therefore two different things. An agent’s operative reasons themselves do not explain her action: her beliefs and desires, which had as their contents her operative reasons (her belief that ‘the meeting was over’, her desire to leave the room once ‘the meeting was over’) constitute, as Michael Smith (1994) puts it, the “rationalizing explanation” of her action.

In distinguishing three types of reasons – normative, operative and explanatory – I am departing from the majority of the literature, which distinguishes just two: ‘normative’ and ‘motivating’. On this normative/motivating distinction, normative reasons are defined as those things that in fact provide justification for a certain action, while motivating reasons are defined as those things which motivate the agent’s action and thereby explain her action. However, the normative/motivating distinction does not neatly map on to a justifying/explanatory distinction, not least because, as I argued above, what explains an agent’s action are facts about her psychology, whereas what motivates her are facts (as she sees them) about the world: the reason that motivated Hieronymi’s agent to leave the room was ‘the meeting is over’, not her inward consideration of her belief that the meeting was over. We should therefore distinguish between ‘operative’ and ‘explanatory’ reasons. Operative reasons are whatever an agent treats as counting in favour of an action, explanatory reasons are facts about the world which explain her action; such will include her beliefs and desires, the contents of which will include her operative reasons. Thus the agent who leaves the room because she thinks

---

2 For Hieronymi’s argument that explanatory reasons are facts about the agent’s psychology, whereas the reasons for which an agent acts are facts about the world, see her Reasons for Action (2011). Hieronymi argues that Donald Davidson’s seminal Actions, Reasons, and Causes (1963) has led to thinking of reasons for action as psychological states that cause action, and that this tension with the idea of reasons for action as facts about the world has resulted in difficulty in understanding the place for the agent’s own reasons for acting in the explanation of her action. I depart from Hieronymi’s solution to this tension in proposing a tripartite account of types of reasons, but have been influenced by the considerations which motivated her own account of the explanation of action. See also Jonathan Dancy’s rejection of equating the reasons that motivate an agent’s action with her psychological states, which he calls ‘psychologism’ (2000, p. 15).

3 There is a lack of consistency in the literature over the terminology, but talk of a distinction between normative and motivating reasons is now the dominant practice, for example by Dancy (2000), Smith (1994) and Parfit (2001).
the meeting is over has no normative reason to do so (since the meeting is not over) but has an operative reason to do so (“the meeting is over”) and the explanatory reasons for her acting have as their content this operative reason (she believed that the meeting was over). The ideal of practical rationality would have normative, operative and explanatory reasons matched up, so that the agent’s operative reasons are normative, and the contents of the explanatory reasons of beliefs and desires would then contain these normative reasons. For example, if I wash my car because it is dirty, and it really is dirty and its being dirty counts in favour of me washing it then I have a normative reason for action (‘my car is dirty’) for which I act (my operative reason is ‘my car is dirty’) and the contents of the explanatory reasons for my action have as their contents this reason (my belief that ‘my car is dirty’).

My tripartite division of types of reasons also covers the distinction that is sometimes made between so-called ‘objective’ and ‘subjective’ justifying reasons. Such a distinction claims that there are two sorts of justifying reasons: those which the agent takes to justify her action (subjective reasons) and those that really do justify her action (objective reasons) (Schroeder, 2008). I am not adopting this objective/subjective terminology, partly because I think it misleading (as it can lead to confusion between normative justification provided by ‘objective’ reasons and rationalization provided by ‘subjective’ reasons, which, although is often referred to, especially in everyday language, as ‘justification’, is something quite different from normative justification), but in any case I do not need to. On my tripartite division of types of reasons, what would be called an agent’s ‘subjective’ reasons are her operative reasons, and what would be called her ‘objective’ reasons are her normative reasons, with only these normative reasons providing justification for action. They derive their justification from the values in virtue of which they are reasons.
2 The justificatory force of normative reasons is derived from their promoting a valued state of affairs

A normative reason provides justification for performing a certain action by counting in favour of that action. Thus:

There is a reason to Φ = there is a true proposition R which counts in favour of action Φ

This may be the extent of the analysis we can make. As T. M. Scanlon says,

I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favour of it. “Counts in favour how?” one might ask. “By providing a reason for it” seems to be the only answer (1998, p. 17).

Thus the question of whether or not R is a reason to Φ is answered by asking whether or not R counts in favour of Φ-ing. To the extent that R counts in favour of Φ-ing, R provides justification for Φ, but it might not do so decisively: if there are other reasons which count against Φ-ing then R might not provide enough normative force to decisively justify Φ. Even though my car really is dirty, and its being so counts in favour of my washing it today, there might be reasons that count against my washing it today, for example that tomorrow I am going on a long drive in the countryside, which will make it instantly dirty again, so to wash it today would be a waste of time.

I propose that what makes R a reason to Φ is that by Φ-ing some value will be promoted: R counts in favour of Φ-ing by virtue of the fact that it provides evidence that by Φ-ing some valuable state of affairs will obtain. The normative force of reasons is supposed to motivate action: an agent ought to do what she has most reason to do, i.e. that for which there is most normative force. Thus, that R counts in favour of Φ should motivate an agent’s Φ-ing. However, not all
reasons have the capacity to motivate all agents. \( R \) might count in favour of agent \( S \) \( \Phi \)-ing, but not in favour of agent \( T \) \( \Phi \)-ing. I propose that this is because the state of affairs by virtue of which \( R \) is a reason to \( \Phi \) is one that agent \( S \) values, but not one which agent \( T \) values. If the state of affairs by virtue of which \( R \) is a reason to \( \Phi \) is one that an agent values, then that agent has reason \( R \) to \( \Phi \). Thus:

\[
S \text{ has a reason to } \Phi = \text{ there is a true proposition } R \text{ which counts in favour of action } \Phi, \text{ which has the capacity to move } S \text{ to } \Phi \text{ by virtue of the fact that it promotes some state of affairs } A \text{ which } S \text{ values.}
\]

This is a version of Counterfactual Motivation Internalism about normative reasons. An agent ‘has’ a reason for action only by virtue of its being capable of moving her to act (which is the thesis of Motivational Internalism) in circumstances of a particular kind (which makes it Counterfactual). \(^4\) The particular kind of circumstances which need to obtain are that the agent has access to the reason; but that she does have access to the reason will not necessarily lead her \( \Phi \)-ing. By ‘access’ I mean that the agent needs to be able to treat the reason as a reason, so she must have some awareness of it, whether that be conscious or non-conscious. My thesis is that if an agent ‘has’ a reason then reason will move her if the right circumstances obtain, not that she ‘has’ the reason if the right circumstances obtain, thus agents ‘have’ reasons whether or not the right circumstances obtain, those circumstances being that the agent has access to the reason, and also that it would be practically rational for her to be moved by that reason.

The reason \( R \) might move an agent in the direction of \( \Phi \)-ing, but she might have reasons against \( \Phi \)-ing which outweigh \( R \), and thus she does not (and ought not

\(^4\) Motivational Internalism is, in Stephen Darwall’s terminology, an ‘existence’ form of internalism: a consideration is a reason for an agent only if some motivational fact about that agent obtains (1983, p. 52). This stands in contrast to externalism, which holds that something can be a reason for action even if it is not capable of motivating the agent. I am not concerned to defend motivational internalism here; there is amply argument for it elsewhere, for example in Darwall, also Smith (1994), Korsgaard (1986), McDowell (1995), Williams (1980). These authors hold various positions on the nature of the particular kind of circumstances that would satisfy the counterfactual, for example that the agent be in possession of full information (Smith, 1994), or that she be ideally virtuous (McDowell, 1995). Some proponents of Motivational Internalism hold an Actual version, whereby if an agent has a reason to \( \Phi \) then it follows by necessity that she actually is somewhat motivated to \( \Phi \).
to) $\Phi$. She might have access to $R$ and it might be that she *ought* to $\Phi$ on the basis of $R$, but because she is practically irrational in some way (for example she reasons incorrectly – see Chapter 3 – or she is weak-willed – see Chapter 7) she fails to $\Phi$. That she fails to $\Phi$ does not undermine the fact that she has reason $R$ to $\Phi$, for, as Christine Korsgaard puts it, there are many things that could interfere with the motivational influence of a reason:

Rage, passion, depression, distraction, grief, physical or mental illness: all these things could cause us to act irrationally, that is, to fail to be motivationally responsive to the rational considerations available to us … when they do move us … they move us with the force of necessity. But it will still not be the case that they necessarily move us. So a person may be irrational, not merely by failing to observe rational connections – say, failing to see that the sufficient means are at hand – but also by being “wilfully” blind to them, or even being indifferent to them when they are pointed out” (1986, pp. 13-14).

Thus an agent $S$ has normative reason $R$ to $\Phi$ if $R$ is true, counts in favour of $\Phi$-ing, and could, if $S$ was aware of $R$ and that it counts in favour of $\Phi$-ing, motivate her to $\Phi$ if it were practically rational (that is, if on the balance of all her reasons she has most reason to $\Phi$) for her to $\Phi$, even if she in fact is not so moved because she is practically *irrational*. The reason can *move* her towards $\Phi$-ing, even if it is not enough to *motivate* her $\Phi$-ing: if on balance she has most reason to not-$\Phi$, $R$ can still move her in the direction of $\Phi$-ing, even if it doesn’t have enough normative force to move her so much as to outweigh the normative force of her competing reasons.

If an agent is motivated to $\Phi$ on the basis of reason $R$, then she will have acted *for* reason $R$. Thus:

$$S \Phi s \text{ for a reason } = \text{ there is a true proposition } R \text{ which counts in favour of } \Phi, \text{ which motivates } S \text{ to } \Phi \text{ by virtue of the fact that it promotes some state of affairs } A \text{ which } S \text{ values.}$$

To illustrate the distinction between there being a reason, an agent having a reason, and an agent acting for a reason, and that these should be defined in terms of something being a reason for action by virtue of the fact that it indicates that performing that action will promote some valued state of affairs, consider the
following example. I’ve gone out to dinner with two companions – Aliyah and Cedric. On the wine list is an aged red Burgundy, which has particularly well-developed gamey flavours. Its having well-developed gamey flavours is something which counts in favour of drinking the Burgundy, by virtue of the fact that such flavours in a Burgundy are valued. There is a reason, then, to choose the Burgundy: that reason being ‘it has well-developed gamey flavours’. There is a true proposition $R$, ‘it has well-developed gamey flavours’, which counts in favour of action $\Phi$, choosing the Burgundy.

The wine list does not include tasting notes, but it does include the growers’ names, and the vintages. Neither Aliyah nor I know enough about the growers and vintages of Burgundy to be able to infer from the list that the wine will have particularly well-developed gamey flavours. As it happens, I adore the taste of such a wine, whereas Aliyah does not. Because of my preference such a taste in wine, I have a reason to choose this particular wine; its being gamey could move me to choose it by virtue of the fact that it promotes some state of affairs which I value; something along the lines of experiencing pleasure or enjoyment. Because Aliyah does not have my preference, she does not have a reason to suggest that we choose this particular wine; its being gamey could not motivate her to choose it, for doing so would not promote any state of affairs which she values.

I have a reason to choose the Burgundy, even though I am unaware of such. Cedric, however, has much more knowledge about wine that either myself or Aliyah, and he therefore knows, when reading the name of the grower and the vintage of the Burgundy, that it will have particularly well-developed gamey flavours. Furthermore, he shares my preference for such tastes, and thus as well having a reason to choose the Burgundy, Cedric is also aware of this. Cedric’s relation to the reason he has, that ‘it has well-developed gamey flavours’, is relevantly different to my relation to the same reason, because we can expect the reason to move Cedric to choose the Burgundy, but we cannot expect that it move

---

5 Generally, gamey-flavours are valued in Pinot Noir: it doesn’t matter that only some people may find such valuable, for my argument, since the reason will be one that is ‘had’ only by those who do find such flavours valuable.
me to the same choice. If Cedric is not moved to suggest such, we can rationally criticize him for this, but I could not be criticized on the same grounds. If Cedric does choose the Burgundy on the grounds that he expects it to have well-developed gamey flavours, we can say that he has acted for the reason. I, however, am unable to act for this reason, since I am unaware that I have this reason. Thus Cedric and I stand in a different epistemic relation to the reason, and because of this we expect the reason to play a part in Cedric’s motivations which we do not expect it to play in mine. Aliyah, of course, is unable to act for this reason because she does not have this reason, her not sharing Cedric and my preference for the particular taste. It is not her epistemic relation to the reason that entails that we do not expect it to move her, but rather that she does not stand in the ‘having’ relation to it.

We have, then, our distinction between there being a reason to choose the Burgundy (there is the true proposition ‘it has well-developed gamey flavours’ which counts in favour of choosing the Burgundy), an agent having a reason to choose the Burgundy (there is the true proposition ‘it has well-developed gamey flavours’ which could move Cedric or me to choose the Burgundy), and an agent choosing the Burgundy for that reason (there is a proposition ‘it has well-developed gamey flavours’, which moves Cedric to choose the Burgundy). Whether or not any of us at dinner had the reason ‘it has well developed gamey flavours’ to choose the Burgundy, the reason would still exist. It would still be the case that ‘it has well developed gamey flavours’ was a true proposition which counted in favour of choosing the Burgundy, because it would still be true that it having gamey flavours was a valuable aspect in the wine. Reasons thus exist independently of any agent ‘having’ them.

---

6 Whether or not it would be true that the reason existed if no-one in the world in fact valued this feature is a question on which I shall remain open, for I am not concerned here to argue for the existence of objective values. My thesis is that an agent has a reason if it could move her, which is compatible with Thomas Nagel’s (1997, pp. 105-106) argument that objective values need not actually motivate us: there could be a reason by virtue of an objective value which no-one actually holds but which we should hold (i.e. could, under the right conditions of, for example, virtue). However, unlike Nagel’s position, mine would not thereby hold that any agent had such a reason; that is, my account would not entail that agents thereby have external reasons for action. There might be such reasons, but no agent would have such reasons. Equally, my position would be compatible with those which, following J. L. Mackie (1977), maintain that objective values would be “queer” properties (e.g. Goldman 2007): R would then simply be a reason to Φ that no agent
The ‘having’ relation is motivational: even though I did not have epistemic access to the reason I still had it, and Aliyah did not have it, not because she did not have epistemic access to it, but because she could not be motivated by it. This is a different sort of ‘having’ relation to the way in which we talk about ‘having’ reasons for belief. We would usually say that an agent ‘has’ a reason to believe \( P \) iff she has epistemic access to that reason. Sometimes, especially in everyday language, we might talk about reasons for action in this same epistemically-relational way, meaning, when we say that there ‘is’ a reason to \( \Phi \), that there is a reason for a certain agent to \( \Phi \), and that the agent ‘has’ this reason only if she knows of it. I am going to call this alternative way of talking about having reasons the ‘epistemic relation account’ of ‘having’, since it holds that an agent ‘has’ a reason to \( \Phi \) iff (if and only if) she stands in the right epistemic relation to it. On an epistemic relation account of the ‘having’ relation, there ‘being’ a reason would equate to my account’s definition of an agent ‘having’ a reason, and an agent ‘having’ a reason would equate to my account’s definition of an agent having a reason to which she has access. I am going to reject this alternative epistemic relation account. This leaves my position with the prima facie problem of entailing that a reason \( R \) can be, simultaneously, both a reason to \( \Phi \) and a reason to \( \neg \Phi \), but I do not think this is a problem at all, as I shall now show.

2.1 The ‘having’ relation is motivational, not epistemic

On my account the ‘having’ relation is one of potential motivation, but the ‘having’ relation is sometimes talked about as if it is one of epistemic access; rather than saying that I have a normative reason to choose the Burgundy and Aliyah does not have a normative reason to choose the Burgundy, such would say that there is a reason for me to choose the Burgundy (and there is not a reason for Aliyah to) but it is not a reason that I have because I do not have knowledge of it. The actually had, on the basis that the value by virtue of which it is a reason does not exist. Because the value does not exist, the reason would have no justificatory force to transmit.

7 This alternative way of talking about ‘having’ a reason for action is mentioned briefly by Bernard Williams: “A may be ignorant of some fact such that if he did know it he would, in virtue of some element in [his subjective motivational set], be disposed to \( \varphi \): we can say that he has a reason to
only person in my example that would have the reason, on the epistemic relation account of ‘having’, would be Cedric, since he is the only person who knows that the wine is gamey; he is the only one with epistemic access to the proposition ‘it is gamey’. The epistemic relation account maintains that one has a reason to Φ only if there is a reason for one to Φ, which is in one’s possession, by virtue of one’s knowing it. However, my account of the having relation is preferable to an epistemic relation account because the latter runs in to trouble when one considers cases of agents who have false beliefs, as I shall now argue.

An epistemic relation account of what it is to have a reason struggle to give intelligible explanations of the actions of agents who have false beliefs, for example Bernard Williams’ man who drinks petrol thinking that it is gin (1980, p. 102) or Hieronymi’s agent who leaves the room because she thinks the meeting is over. This is because, since one can only know true propositions, the epistemic relation account entails that such agents do not ‘have’ any reasons at all, and thus their actions look unintelligible. To illustrate this point, I shall extend my Burgundy example to incorporate an agent who has false beliefs, showing how an epistemic relation account of ‘having’ cannot adequately account for the intelligibility of such an agent’s action.

Imagine that Danny is also joining us for dinner. Danny thinks that he has the same level of knowledge about wine as does Cedric, but in fact he has lots of false beliefs about wine, one of which is the erroneous belief that aging wines φ, though he does not know it. For it to be the case that he actually has such a reason, however, it seems that the relevance of the unknown fact to his actions has to be fairly close and immediate: otherwise one merely says that A would have a reason to φ if he knew the fact. I shall not pursue the question of the conditions for saying the one thing or the other, but it must be closely connected with the question of when the ignorance forms part of the explanation of what A actually does” (1980, p. 61; italics added). Mark Schroeder (2008) argues that to talk of an agent ‘having’ a reason in the terms of the epistemic relation account (which he calls the ‘Factoring Account’) is to make the mistake of confusing the subjective/objective reasons distinction. I have already rejected using this distinction, but his argument is similar to mine: Schroeder’s argument, as I understand it, is that since agents can act for reasons only if they ‘have’ them, because agents act for reasons that do not exist (in the terms of my account, they act for operative reasons that are not normative), it cannot be that by ‘having’ we mean something that exists and that is in the agent’s possession, since such would imply that sometimes agents possess things that do not exist. My argument, similarly, is that the epistemic relation account has trouble when it comes to agents who have false beliefs about their normative reasons.
brings out their fruit and floral notes, and suppresses any gamey qualities. Imagine too that, like Aliyah, he prefers fruity pinot noirs over more-developed, gamey pinot noirs. He thus believes, after looking at the wine list, that he ought to choose the aged Burgundy. Moreover, he then does choose the Burgundy (to my delight and his disappointment), for his operative reason, ‘it will be fruity’. On my account, ‘it will be fruity’ is not a normative reason to choose the Burgundy, since the Burgundy not, in fact, fruity, and thus Danny does not have a normative reason to choose it. There is, therefore, no normative reason for which Danny can choose the Burgundy, but there is an operative reason for which he does. Because the having relation, on my account, is one of motivation, all operative reasons can satisfy the counterfactual by the agent treating them as if they were normative, whereas on the epistemic relation account of having they cannot, since an agent can know only that which is true, and thus cannot ‘have’ an operative reason to act unless it is a normative reason, i.e. unless it is true both in the sense of being a true proposition, and in the sense of truly counting in favour of the action (an agent could mistakenly treat a true proposition as a normative reason by treating it as if it counted in favour of an action which it does not).

Now, were the ‘having’ relation an epistemic relation, such that I do not have a reason to choose the Burgundy, but Cedric does, we would run into difficulties when thinking of Danny. On the epistemic relation account Danny has neither a reason to choose the Burgundy, nor a reason to not choose the Burgundy, since he is not aware of any normative reason which counts in favour of ordering it, nor of any normative reason which counts in favour of not ordering it. And yet, clearly when Danny chooses the Burgundy, he does so for a reason – he does not do so on a mere whim.

What could be the reason for which he acts? If asked why Danny chose the Burgundy, the natural answer would be to say that it was because he believed that the wine would be fruity, and if we asked Danny why he chose the wine after he had done so, tasted it, and pulled a puckered-face at the taste of the gamey flavours, he would say ‘because I believed that it was going to be fruity’. The
explanatory reason for his action is his belief, but this cannot be the reason for which he chose the wine. I have already argued that agents do not act for psychological states, but for facts (as they see them) about the world, but even if I am wrong to do so, the epistemic relation account cannot appeal to this.

This is because on the epistemic relation account an agent has a reason only if she is aware of the reason, and can act for this reason only if she stands in the epistemic relation to the reason. If the reason for which Danny chose the Burgundy was his belief, then he would need to stand in the epistemic relation to this belief in order to have it and act for it. That means that he would need to have a second-order belief about his belief, and this sounds very odd indeed. We would still expect Danny to choose the Burgundy, whether or not he believed that he believed that it would be fruity. We would still say that he chose the Burgundy for a reason, whether or not he had such a second-order belief. If we asked him before the wine arrived why he has chosen it, he would say, ‘because it is fruity’, and couch it in terms of ‘because I believed it was fruity’ only once he realized that it wasn’t. His change in language does not mean that the reason for which he acted had changed; all that had changed is the way in which Danny could explain his action. Without couching it in terms of a belief, he would not be able to express his operative reason, that the wine was fruity, without thereby implying that he still believed that it was true, because of the relationship between asserting the truth of a proposition P and one’s belief in the truth of proposition P. Unless one draws attention to their having come apart – for example by presenting P as the content of a former belief - it is assumed that in asserting the truth of P, one is asserting that one believes in the truth of P.

In my example, I claimed that ‘it is gamey’ was a reason to choose the Burgundy. The epistemic relation account would say that ‘it is gamey’ was a reason for Cedric and for me to choose the Burgundy, but was not a reason for Aliyah or for Danny to choose the Burgundy. It would say for Aliyah and for Danny that ‘it is gamey’ was a reason to not choose the Burgundy. How does my account handle the relationship of Aliyah and Danny to the reason ‘it is gamey’? I would need to either say ‘it is gamey’ was both a reason to choose the Burgundy and to not
choose the Burgundy, or else deny that it was a reason to not choose the Burgundy. I am certainly not going to deny that it was a reason to not choose the Burgundy: ‘it is gamey’ was a consideration which counted in favour of not ordering the Burgundy (because it would thwart the valued state of affairs of pleasure or enjoyment), and it was a reason that Aliyah and Danny had (though neither of them knew it), but it was not a reason that Cedric or I had. And so instead I am left with the slightly odd-sounding outcome of ‘it is gamey’ being both a reason for ordering the Burgundy, and a reason for not ordering the Burgundy. I don’t think that this is a problem, as I shall show.

2.2 The same consideration can be a reason both for and against the same action

In ordinary language there ‘being’ a reason for $S$ to $\Phi$ and $S$ ‘having’ a reason to $\Phi$ are used interchangeably, but this can cause confusion when then talking about the reasons for which $S$ acted. In the first sense – there being a reason ‘for’ $S$ to $\Phi$ - we are talking of normative reasons, and in the second sense – $S$ acting ‘for’ a reason - of operative reasons which may or may not be normative, and saying of someone that $R$ is a reason ‘for’ them is often meant, in everyday language, to mean that $R$ was a reason which they took to be a reason ‘for’ them – i.e. that it was an operative reason which they took to be a normative reason. Thus rather than saying that $R$ is a reason ‘for’ one agent to $\Phi$ and a reason ‘for’ another agent to not-$\Phi$, my position is that $R$ is simultaneously a reason to $\Phi$ and to not-$\Phi$, independently of ‘being’ a reason by virtue of any particular agent. ‘It is gamey’ is a reason to choose the Burgundy, and it is also a reason to not choose the Burgundy.

I do not think that it is a problem that on my account a feature of a situation can be both a reason to $\Phi$ and a reason to not-$\Phi$. Even for an individual agent, a reason can be one they have that counts both for and against the same action. ‘This philosophical problem is hard’ could be simultaneously a reason that I have to tackle it (because it promotes a state of affairs that I value; contributing meaningfully to philosophical debate) and a reason that I have to not tackle it.
(because it thwarts another state of affairs that I value; feeling confident that I can accomplish tasks that I set myself). Reasons for action ultimately bottom out in values, and an agent has a plurality of values, thus it should not be surprising that when two of an agent’s values conflict in some way the same thing can count as a reason for and against the same action. Since something can be a reason an agent has both for and against an action, it is not at all surprising that a reason can be a reason one agent has for an action and simultaneously a reason another agent has against the same action. Since agents can have normative reasons only that exist (since normative reasons are, by definition, true propositions), this means that there can be a reason to \( \Phi \) which is simultaneously a reason to not-\( \Phi \). Thus the same reason can be a reason both for and against the same action.

I suspect that it sounds odd to say of a reason that it counts both for and against a particular action, because when talking of theoretical reasons we might be loath to say the same. Although we might say that \( R \) could be an operative reason both for which \( S \) believed \( P \), and a reason for which \( Q \) believed not-\( P \), normatively it would be a reason only for believing \( P \) or for believing not-\( P \); it couldn’t be both at the same time. I think that also is it might sound as though, if a reason can count both for and against the same action, that in that case just \( \text{anything} \) could count in favour of anything; that there is no objective fact-of-the-matter about whether something is, or is not, a reason to do something, whereas with theoretical reasons there is an objective fact-of-the-matter. Some exploration of the comparison with theoretical reasons, I think, will put these concerns to rest.

I have already said that normative reasons are things that count in favour of a certain action. Normative reasons can also count in favour of a certain attitude (Dancy, 2000), such as a belief. In the theoretical domain, it is clear that people often believe things for ‘bad’ reasons, by which I mean operative reasons for belief that are not normative reasons for belief. Imagine Allison, who has an undergraduate degree and lives in 21\(^{st}\) Century Britain, and who is aware of all the facts that are reasons to believe that Earth is spherical, yet does not believe Earth is spherical, having been persuaded by a cult that she has recently joined that the Earth being spherical is a conspiracy, and that is actually flat. The
testimony of the cult is the operative reason for which Allison believes that Earth is flat, but it is not a normative reason which Allison can have, since it is not a true proposition that counts in favour of believing Earth to be flat.

In the case of action, take Eliza, who joins Aliyah, Cedric, Danny and me for dinner. Eliza, like Cedric, knows that the Burgundy will be gamey and also likes gamey-tasting pinot noir, but she fails to be moved to choose the Burgundy because earlier she saw a black cat cross her path, and superstitiously believes a black cat crossing one’s path is a reason to do precisely the opposite of what one wants to do. The appearance of the cat is the operative reason for which Eliza does not choose the Burgundy, but it is not a normative reason to not choose the Burgundy, since it does not in fact count in favour of not choosing the Burgundy.

Why is it that the cult's testimony is not a normative reason to believe in a conspiracy about Earth’s shape, and that Eliza’s black cat crossing her path is not a normative reason to not choose the Burgundy? These facts do not count in favour, or justify, the belief or action that the agents take them to. What does make a fact count in favour, or justify, a belief or action? In the case of theoretical reasons, it looks as though facts count in favour of beliefs - are reasons for belief - when they are evidence of the truth of the proposition that forms the content of the belief. If $R$ is a reason for $P$, then the truth of $R$ supports, though need not necessitate, the truth of $P$; what makes $R$ count in favour of believing that $P$, is that $R$ is evidence for $P$. My evidence for my belief that my British Shorthair Cream cat is biscuit-coloured is my perceptual experience of biscuit-colour when looking at my cat. My cat’s colouring is in some way related to my perceptual experience of biscuit-colour when looking at him, such that his appearance to me is evidence of his actual colouring. This evidential relation appears to stem from the fact that the reasons that there are for asserting the truth of the proposition ‘my cat is the same colour as a biscuit’ are also the reasons for believing ‘my cat is the same colour as a biscuit’ – that my cat appears to be biscuit-coloured is both evidence for the truth of the proposition that he is biscuit-coloured, and also a reason – a fact that counts in favour of – the belief that he is biscuit-coloured.
Can something be both a reason to believe $P$ and also a reason to believe not-$P$? In the case of Allison we saw that just because someone thinks that something is a reason for belief does not make it the case that it really is a reason for belief. Thus a case in which we have two agents with the same piece of evidence, and one takes it as a reason to believe $P$ and the other takes it as a reason to believe not-$P$ is not going to help us out here. It looks as though in the theoretical domain something cannot in fact be a reason to both believe that $P$ and to believe that not-$P$. I think this is because reasons in the theoretical domain are evidential, such that it cannot be that both $P$ and not-$P$ make it the case that $R$ obtains, for, if they did, $R$ would not actually be a reason to believe one way or another – it failing to provide evidence one way or another, because it could be the result of either state of affairs. Thus it may be that a theoretical reason cannot count both for and against the same belief, and this might be why it sounds odd to say that a practical reason can count both for and against the same action.\(^8\)

But reasons for action are not evidence of the action, nor related to the action in the same way. In particular, what we might loosely call, without making any metaphysical claims, the ‘quasi-causal relationship’ seems to run in the wrong direction. Whereas we can say, for example, that the world being spherical in some way ‘makes’ it the case that the top of the mast of a ship is its last part to disappear over the horizon, and thus that the mast is the last part to disappear is evidence for the truth of the proposition that the world is spherical, and thereby is a reason to believe that the world is spherical, we cannot say that Cedric’s action of ordering the Burgundy makes it the case that the Burgundy is gamey. The truth

\(^8\) I am not ruling out the possibility that something can simultaneously be a reason to believe $P$ and a reason to believe not-$P$, merely suggesting that the intuition that it cannot may account for the intuition that a reason for action cannot be simultaneously a reason for and against the same thing. One might think, for example, that the UKIP party is gaining popularity might be a reason that counts in favour of believing that the Liberal Democrats will do well at the next election, by splitting the opposing voters, or alternatively a reason that counts against the same belief, by splitting the Lib Dem voters. We might say that in this case the reason is not sufficient to support either belief, in which case it is not actually a counter-example, or we might say that it is in fact an example of a reason for belief that counts for and against the same belief, in which case we have grounds for questioning the intuition about theoretical reasons too. Either way, it does not undermine my argument that reasons for action can simultaneously be a reason for and against the same action. (Thanks to Tom Smith for this point and example.)
of the proposition that is the content of a belief causes (in some sense, at least) the existence of the reasons which count in favour of the belief, but an action cannot cause the reasons which count in favour of the action; it’s the wrong way around.

However, I don’t think this is as asymmetrical as it first looks. What makes a reason count in favour of a belief is that it is evidence that the content of the belief is true. The aim of belief formation is to latch on to truths, so what makes \( R \) count in favour of believing that \( P \) is that \( R \) is evidence that believing \( P \) will bring one closer to truth. Analogously, what makes a fact count in favour of a particular action is that it is evidence that the action will promote some value that the agent has. Cedric values enjoying wine, gamey-tasting wine is enjoyable for him, so the fact that the wine will taste gamey is evidence that ordering the Burgundy will promote a state of affairs he values, that of enjoying wine.

Thus both practical reasons and theoretical reasons can be seen as providing evidence that responding to them as reasons will fulfil the aims of the agent. The difference between the theoretical and the practical domains lies in the fact that in belief formation all agents have the same goal, and there is only one goal - to form true beliefs, whereas in the practical domain an agent has lots of different goals - a multitude of values to be promoted - and what will enable her to fulfil these goals will very much depend upon how she is constituted. Cedric and Danny share the same value of enjoying wine, but choosing the Burgundy is going to promote this valuable state of affairs only for Cedric, since only Cedric is so constituted as to enjoy gamey wines. Thus the wine being gamey is a reason to choose the Burgundy that Cedric has, but Danny does not. And the difficulty of the philosophical problem is a reason I have to tackle it, and also a reason I have to not tackle it, because I have two different values: that of contributing meaningfully to philosophical debate, and that of feeling confident that I can achieve a task that I set myself. The difficulty of the philosophical problem is ‘picked out’ as a reason to tackle it by my concern to contribute meaningfully, and picked out as a reason to not tackle it by my concern to feel confident. Reasons for action could still described as evidential, but rather than providing evidence
for the truth of a proposition, which is the sole theoretical goal, they provide evidence that an action will achieve a practical goal of promoting a valuable state of affairs, and agents have a plurality of values, and thus a plurality of practical goals.

This does not mean that anything can count as a practical reason for anything. There needs to be the right kind of connection between a practical reason and the action for which it is a reason. Eliza’s cat crossing her path was not a reason for not ordering the Burgundy, because not ordering the Burgundy would not promote any state of affairs which Eliza values. Were it true, say, that if a black cat crosses one’s path and one then drank red wine one would fall into a deep sickness, then it could have been a reason to not choose the Burgundy, if physical health is something Eliza values. But there was, in fact, no connection between seeing a black cat and any of Eliza’s values, thus seeing a black cat was not a reason for Eliza to not choose the Burgundy. An agent has a reason \( R \) to \( \Phi \) only if by \( \Phi \)-ing she would promote some state of affairs which she values.

3 The reasons that an agent has are determined by her values

I have argued that there is a reason \( R \) to \( \Phi \) iff there is a true proposition \( R \) which counts in favour of action \( \Phi \), and that reasons count in favour of actions by virtue of the fact that they provide evidence that by \( \Phi \)-ing a value, \( A \), will be promoted. I have also argued that the ‘having’ relation between an agent and normative reasons for action is motivational, not epistemic: a reason \( R \) is a reason that agent \( S \) has to \( \Phi \) iff \( R \) could move \( S \) to \( \Phi \). A reason \( R \) will move an agent \( S \) to \( \Phi \) because it provides justification for her \( \Phi \)-ing. The justificatory force of a normative reason is derived from its promoting a valuable state of affairs \( A \), thus \( R \) will move \( S \) to \( \Phi \) iff it provides evidence that by her \( \Phi \)-ing \( A \) will be promoted, and \( S \) values \( A \).
Thus, $R$ is a reason that $S$ has to $\Phi$ by virtue of the fact that $S$ values $A$: the reasons that an agent has are determined by her values.\(^9\)

An agent has a plurality of values so she will have a plurality of reasons for action; features of her current situation will be reason-giving considerations that count for and against different actions. For example, say I am trying to decide whether to attend a yoga class this afternoon. When reflecting, I consider the facts of my current circumstances, for example that I am going out for dinner this evening, that I am feeling quite bright and motivated this morning, to decide whether these features of my circumstances count in favour of going to a yoga class. I don’t just pick one feature whose basis upon which I make my decision; I am trying to work out what course of action best responds to all of the reason-giving features of my circumstances, not just those which promote one particular value. For example, that yoga will suppress my appetite is a reason to not go to yoga by virtue of the fact that it promotes the state of affairs of my gaining pleasure from dining out, which is a state of affairs that I value. That I’m feeling bright and motivated is also a reason to not go to yoga, but because it promotes a different valued state of affairs, that of using the time to write productively. My reasons are determined by my values, but they are ‘picked out’, that is brought to my awareness, by my desires and concerns.

My desires and concerns help me to act for what really are reason-giving features of my circumstances. I want to be aware of all the relevant reason-giving

---

\(^9\) I maintain that an agent’s reasons are determined by her values, which coincides with the internalist picture of normative reasons, denying the ‘external interpretation’ of reasons statements, as Bernard Williams put it, that an agent has a reason to $\Phi$ even if she lacks “any motive which will be served or furthered by [her] $\Phi$-ing” (Williams, 1980). However, I do not agree with Williams’ argument that a “sound deliberative route” is the only way in which an agent can come to have a reason which she did not previously. I agree with John McDowell that it is plausible that in certain cases there are no rational routes that would effect an agent’s transition to seeing matters aright by coming to see what really are reasons for her, and their real normative force (1995, p. 102), and that a non-rational route, such as conversion or rhetoric, might be needed to effect that transition. In Chapter 8, for example, I argue that in cases in which a person is provoked into being so angry that they see their intention to kill their provoker as being justified, there is no rational route by which they can come to see matters aright and realize that they in fact ought not to kill their potential victim, but that what is needed is a non-rational effect from an emotion such as fear, to enable them to see and respond to the reasons they have to not act upon their intention to kill.
considerations and give them their due weight, and not waste my time deliberating about considerations that have no or little bearing on my decision. Which considerations of my circumstances are reason-giving for me at this time will be picked out by my desires and concerns. For example if getting the most pleasure from dining out is of importance to me, then yoga’s potentially appetite-suppressing effect will seem like a strong reason-giving consideration for me to not go to yoga; if eating less is important to me (perhaps I am suffering from digestive troubles of late), the same consideration would count in favour of my going to yoga. If tonight I am going out for dinner purely for the pleasure of the company, and will eat only for the purpose of satisfying whatever appetite I happen to have, yoga’s potential appetite-suppressing effect is not going to have much normative force in either direction.

My reasons include not just those picked out by my occurrent, or manifested desires, but also by standing, or background, desires. Occurrent desires are those that are playing a role in one’s psyche at the current time. For example, if I am consciously aware that I want to eat less, I can reason that yoga supressing my appetite gives me a reason to go to yoga, on the basis that it will realize this desire to eat less. My desire to eat less is in this case an occurrent desire, because it is playing a role in my psyche, but desires of which I am not consciously aware can also play this role.\(^\text{10}\) Perhaps I have a non-conscious desire to avoid a person who is likely to be attending the yoga class, which guides me towards choosing to not go. This desire is occurrent, because it is playing a role in my current psyche, though I am not aware that it is doing so. Occurrent desires need not be in control of my actions: my desires to have a good appetite for dinner and to avoid seeing the person I don’t want to may be occurrent even as I am gathering together my yoga mat and setting off for the class.

\(^\text{10}\) Traditional models of mind stemming from Descartes involved an identification of the psychological aspects of a person with consciousness, and thus would exclude the possibility of non-conscious mental states, such as desires and beliefs. However, it is no longer controversial to assume that such are not just possible, but conceptually entirely unproblematic. For example, see Thomas Smythe’s ‘Unconscious Desires and the Meaning of ‘Desire‘” (1972).
Conscious and non-conscious desires will also be playing a background role as standing desires. Standing desires are those which I have but might not be playing a role in my current psyche. For example, I might have a standing desire to have a good sleeping pattern, which will only sometimes be manifested as an occurrent desire that guides my behaviour. This standing desire might not be relevant to my decision about whether or not to go to yoga, but some standing desires will be relevant towards the decision at hand, even though they are not occurrent, and I might become aware of such if I overlook them in making my decision. If I decided to not go to yoga because I was panicking about a deadline, and let down my friend to whom I’d promised my attendance, I might regret that I failed to appreciate the importance of that relationship when making my decision. Although at the time of deciding to not go to yoga my desire to be a good friend was not occurrent, not playing a role in my decision making, it was a standing desire that I had. This means that I had a reason to go to yoga that was picked out by my standing desire to be a good friend, though the desire was not occurrent. Thus an agent’s reasons for action are picked out not just by her occurrent desires.

Standing desires lie, as we might say, “in the back of our minds” most of the time, and occasionally generate thoughts, feelings and actions of the relevant sorts. Some of my standing desires will be ‘intrinsic’ desires. Intrinsic desires are those that are for states of affairs that are desired for their own sakes. It is widely agreed that pleasure is desired for its own sake, but anything that we value for itself, rather than for some further end, is a candidate for an intrinsic desire. The extent and acquisition of intrinsic desires are contentious topics in to which I will not

---

11 By talking of desires being in the ‘background’ or ‘back of our minds’, I am not making the same distinction as Philip Pettit and Michael Smith in their ‘Backgrounding Desire’ (1990). Pettit and Smith argue that desire is always in the ‘background’ of the genesis of human action, and such background desires are part of the motivating reasons for an agent’s action. A desire is in the ‘foreground’ iff the agent believes she has that desire and is “moved by the belief that a justifying reason for the decision was that the option chosen promised to satisfy that desire” (ibid., p.586). My point here is not that desires provide motivating reasons for action (I have already rejected adopting this terminology; on my account desires are usually explanatory reasons, though an agent might, mistakenly, think a desire per se is a normative reason for action, and thus adopt it as an operative reason – see Chapter 3), but that desires play a role in determining what are normative reasons for an agent. The normative reasons an agent has are determined by her values, via her desires, both occurrent and standing.
delve here, but I favour a non-hedonistic account, thus assume that we have many more intrinsic desires that just pleasure, though am very sympathetic to the idea that that all other intrinsic desires are in a sense generated by what looks likely to be an innate intrinsic desire for pleasure. For example, Timothy Schroeder (2004) proposes that new intrinsic desires are acquired by associating new states of affairs with ones that already satisfy existing intrinsic desires, according to the principles of reward-based learning. Such a view is very much consistent with the evidence from cognitive science that I will argue challenges the traditional view of emotions as being the enemy of practical rationality. For example, António Damásio’s Somatic Marker Hypothesis (1994), which I shall expound in Chapter 6, proposes that pleasure ‘marks’ certain states of affairs as being advantageous to us, and that via reward-based learning we then come to desire such states of affairs. However, I do not think that my picture of practical rationality depends upon any particular account of the extent or acquisition of intrinsic desires. The psychological hedonist might label what I call intrinsic desires as instrumental desires for pleasure, but in the context of my position this would be a mere matter of terminology.

My position assumes that there is some relationship between an agent’s values and her intrinsic desires such that if she values state of affairs A she has an intrinsic desire that A be the case. That she has intrinsic desire that A pertains gives her reason for the desire that A be brought about, or maintained. This then gives her reason to desire to act so as to bring it about that A obtains, which gives her reason to desire to do whatever it is that will bring it about that A obtains.

---

12 In terms of the origin of intrinsic desires, the psychological hedonist holds that there is only one intrinsic desire – for pleasure – which is innate (e.g. Pollock, 2006), whilst of those who hold that we have more, some propose that it is possible to acquire such by reasoning (e.g. Smith, 1994), against Hume’s insistence that reasoning could not play such a role (Hume, [1739/40] 1978).

13 I will talk as if an agent’s values determine her intrinsic desires, but it does not actually matter to my overall argument whether in fact valuing state of affairs A simply is to have the intrinsic desire that A pertains, or even whether that she already has an intrinsic desire that A pertains that gives rise to her valuing A. If either of these two alternatives is preferable to the position that values precede intrinsic desires, then my argument would be simply that an agent’s reasons are determined by her intrinsic desires, but this would not entail that desires are reasons, or that an agent’s reasons for action are determined by whatever she happens to want to do. It would still only be intrinsic desires that would determine reasons for action, not her occurred or manifested desires, or any non-intrinsic standing desires. However, there are very good reasons to believe that reasons for action are in fact determined by values, not desires, for example see Dancy (2000), Raz (1999), Scanlon (1998), Quinn (1993).
These desires that follow from the intrinsic desire are ‘instrumental’ desires; they are means to an end, the end being that the state of affairs $A$ pertain, for which the agent has an intrinsic desire; there is no further end for which she desires that $A$ pertains. The normative force of these desires is transitional, such that the intrinsic desire that $A$ be the case provides the justification for the instrumental desire to do whatever it is that will bring it about that $A$ is the case.

From this instrumental desire to do whatever it is that will bring it about that $A$ is the case is generated the realizer desire to perform a specific action. If the agent believes that her $\Phi$-ing would bring it about that $A$, or would make it more likely that $A$, then she has reason, and, transitively, justification for desiring to $\Phi$. This desire for action is a ‘realizer’ desire because $\Phi$-ing would realize state of affairs $A$. The desire to $\Phi$ is justified, and transmits its normative force to the action of $\Phi$-ing, but that the desire is justified does not entail that the action will be justified, since there will be likely be reasons other than those transmitted by the realizer desire that will count for and against the action of $\Phi$-ing. Instrumental desires, as I shall argue in Chapter 3, ‘provide’ reasons for action only in the sense that they transfer those reasons for which they are justified to the action, but they do not ‘provide’ reasons in the sense of generating them. It is the agent’s values that provide the normative force, and the instrumental desires merely transmit this force via the reasons for which they are justified. Thus the normative force of the reasons that justify the agent’s desire to $\Phi$ will be up against all of the reasons that count against her $\Phi$-ing, which will derive their normative force from the other of her values which provide her with her reasons against $\Phi$-ing. Whether or not she ought to $\Phi$ is thus dependent upon the normative force of all of her reasons, not just those transmitted by her desire to $\Phi$, and thus the question of whether or not she ought to $\Phi$ is thereby entirely dependent upon her values, since the normative force of her reasons is derived from their instrumental value in promoting those states of affairs.\(^\text{14}\) What an agent ought to do is whatever has

\(^{14}\) My thesis is that all an agent’s reasons are dependent upon her values. The special status of moral reasons, the so-called “Central Problem” for internalism, can be accounted for within this. As I will argue in the next section, an agent has a reason if it could motivate her to act, whether or not it does as a matter of fact motivate her to act. Whether or not a reason can motivate an agent to act is dependent upon her valuing the state of affairs from which the reason derives its motivational force, by virtue of the fact that it would promote the existence of that state of affairs.
most normative justification: practical rationality consists in doing what one has most normative reason to do.

3.1 Practical rationality consists in doing whatever it is that one has most reason to do

When one deliberates about what one ought to do, one is trying to latch on to all those, and only those, considerations that are reason-giving for one in one’s situation. One is trying to work out what one ought to do, all things considered, that is, one what has most reason to do, taking in to account not just those reasons for which one has whatever realizer desires are occurrent, but all of the reasons one has that are relevant to the situation. Practical reasoning is therefore reason-tracking; it should track what one has most (normative) reason to do. From the fact that practical reasoning is a way of working out what one has most reason to do, and that practical reasoning aims at practical rationality, we can

My proposal is simply that all humans value the states of affairs from which moral reasons derive their normative force, thus all agents have these reasons. That agents do not always do what they morally ought to do, or that there can be severe disagreement about what is the morally right thing to do, is not down to people lacking the relevant moral reasons, but is down to various mistakes or disagreements as to how to bring it about that the relevant valued state of affairs pertains. All humans value justice, for example, but it can be very hard for us to appreciate what actually is just and how to bring it about, especially when oneself is going to be affected as states of affairs in which oneself or one’s loved ones are at a slight disadvantage might seem to one to be those states of affairs that would be most just. It could be that the fact that all humans value certain things such as justice, is precisely what makes moral reasons special: they are reasons that every single human has, whereas other reasons are not. The anomaly of psychopaths could be explained either as agents who don’t, for some reason, have the relevant values, or else that they do not have the relevant intrinsic desire that is associated with moral values, or it could be that for some reasons the relevant intrinsic desires do not operate as they would do in moral agents, such that the appropriate instrumental desires, and thus realizer desires, to promote the relevant state of affairs, do not arise. As I argue in Chapter 6, there is evidence that such could be the result of physical damage to the brain which prevents the right feelings to be generated that are needed to respond to those reasons. The way in which psychopaths can be ‘taught’ to behave morally is to respond to different reasons, such as fear of incarceration which, although can result in performing the morally right actions, is not performing the morally right actions for the moral reasons which make them right. Even moral agents might sometimes be doing the right thing for such non-moral reasons, for example in Chapter 8 I argue that the presence of a police officer can enable an agent to restrain herself from killing by enabling her to respond to the reason of her own welfare. The virtuous person would not need such to help her to do the right thing, but the rest of us might, at least some of the time. Laws can be seen as ‘placeholders’ for moral reasons: if an agent endorses a policy of avoiding illegal action then the reasons for which an action are illegal, many of which will be moral, will be implicitly endorsed by her when she decides against an action on the basis of its illegality, in the same way as, as I shall argue later in this thesis, particularly in Chapter 7, an agent implicitly endorses the reasons for which she is experiencing a certain emotion when she decides to ‘trust’ that emotion and acts on it.
deduce that *practical rationality consists in doing what one has most reason to do.*

One might be sceptical that practical rationality *can* be rationality, properly construed (Korsgaard, 1986). Such scepticism about practical rationality arises from the observation that the method by which we are able to be theoretically rational – that of reasoning – does not seem to properly apply to practical matters. Such an argument is often traced back to David Hume, and has something like the following structure:

1. Reason is a faculty that judges of the truth and falsity of ideas (Hume, 1888, p. II.i.1/458);
2. Actions are not ideas;
3. Therefore actions cannot properly be said to conform or be contrary to reason.

That practical reason does not judge the truth and falsity of ideas in the way that theoretical reason does, does not entail that we ought to be sceptics about practical rationality. The differences between theoretical reason and practical reason, far from leading to such scepticism, lead us, by analogous reasoning, to a better understanding of practical rationality. In both practical and theoretical rationality the role of reason is to judge, the ‘having’ relation is motivational, and an agent’s autonomy in both theoretical and practical reasoning is expressed via rational guidance.

### 3.1.1 Reason’s role is as judge in both theoretical and practical rationality

Theoretical reason judges the truth and falsity of ideas because an agent’s *sole* theoretical aim is to believe a proposition if and only if it is true. However, as we have already seen, agents have a plurality of *practical* goals, and so reason simply has a different task to play in practical rationality. This does not mean it is
not ‘proper’. As I argued in section 2, practical reasons provide evidence that performing the action for which they are a reason will fulfil the practical goals of the agent; that of promoting the states of affairs that she values. As I will argue in Chapter 3, practical reasoning is not just means-ends reasoning, but also reasoning about ends: the question is not just whether Φ-ing will bring it about that a desired state of affairs, but whether the values that it will promote are more important or less important than those by virtue of which the reasons against Φ-ing have the normative force that they do. Theoretical reasoning judges whether a proposition $P$ is true or false; practical reasoning judges whether a true proposition $R$ really does count in favour of action $\Phi$, and whether, on the balance of reasons, action $\Phi$ is or is not justified.

3.1.2 The ‘having’ relation for both theoretical and practical reasons is motivational

It is because agents have only one theoretical end, that of believing a proposition iff it is true, that it seems natural to speak of the ‘having’ relation of theoretical reasons as being simply one of epistemic access. An agent ought to be moved by a theoretical reason $R$ if she has epistemic access to it, therefore there is no need to distinguish between the epistemic and motivational relation of epistemic reasons, however it is not the case that an agent ought to be motivated by a practical reason just because she has epistemic access to it. Practical reasons are dependent upon an agent’s values, thus reason $R$ will be one which agent $S$ has only if she has the relevant value, whereas theoretical reasons are not dependent upon an agent’s values, thus theoretical reason $TR$ is a reason that all agent’s ‘have’ in the motivational sense. We therefore talk of ‘having’ a theoretical reason to distinguish between agents who do and do not have access to that reason, for in the theoretical domain such is what distinguishes whether or not an agent can respond to a reason as a reason. What makes it the case that an agent ‘has’ either sort of reason is that she can respond to it as a reason, but for theoretical reasons this seems to be just a question of epistemic access, whereas in the practical domain it is not. However, it is not actually so straightforward. What makes it the case that an agent ‘has’ a reason – practical
or theoretical – is that she can be moved to the corresponding action or belief by it. Just knowing the truth of a proposition that is in fact a theoretical reason to believe $P$ is not necessarily to be in a position where one is capable of being moved to form the corresponding belief that $P$.

Recall Aliyah from my earlier example. She was unable to respond to the reason ‘it is gamey’ as a reason to choose the Burgundy, not because she didn’t know that it was gamey, but because she couldn’t be motivated by this reason because its being gamey would not promote any state of affairs which she valued: even had she known that the Burgundy would be gamey she would not have been motivated to choose it on the basis of such a reason. An agent needs to have the right values to be able to be motivated by a practical reason, but also needs the corresponding desires and concerns which will ‘pick out’ the reason: without such, comprehension of the truth of the proposition of $R$ cannot motivate action. Analogously, to be able to respond to a theoretical reason as a reason, an agent needs to have the right prior beliefs and faculties. The fact that the very top of the mast is the last thing to disappear of a ship as it sails away from shore is a reason to believe that Earth is spherical. This is a reason that you have, whether or not you were aware of the fact before you read this paragraph. It was a reason that you had because contemplation of the fact from the perspective of an intelligent and well-educated person in the 21st century can motivate, or reaffirm, the belief that Earth is spherical. From the perspective of an intelligent and well-educated person in Plato’s Academy, however, contemplation of the disappearing mast of a ship could not motivate the belief that Earth was spherical. The ability of an agent to be moved to form a belief from awareness of a fact depends upon their current set of beliefs since a potential belief’s lack of coherence with one’s current set is, of course, likely to result in rejecting the belief. Thus although in the time of Plato there existed the theoretical reason - that the mast of the ship disappeared last over the horizon – that counted in favour of the belief that Earth is spherical, it was not a reason that Plato’s students had as it could not move them to form the corresponding belief. For you to have a reason - theoretical or practical – entails that you are capable of responding to that reason as a reason: that it can move you to the corresponding action or belief.
The ‘having’ relation of theoretical reasons is thus also motivational. To be able to judge the truth and falsity of ideas, reason is reliant upon the agent having the right prior beliefs and faculties. The difference between theoretical and practical rationality in this respect is that practical rationality is much more dependent upon the subjective facts of the agent. Nearly all agents of a particular class (for example all humans of normal intelligence who have had a basic education) have the same reasons for belief, because to have a reason for belief is not dependent upon the sort of differences that hold between agents in relation to their reasons for action. We should all believe the same thing: there is only one goal of belief formation – to believe a proposition iff it is true – thus we should believe whatever is true, and since what is theoretically true is independent of agents, what we should and shouldn’t believe is universal. What we should and shouldn’t do, on the other hand, is not universal. What we should and shouldn’t do is dependent upon our individual values.\(^{15}\) This does not mean that there is no objective fact-of-the-matter about what an agent ought to do; an agent ought to do whatever she has most reason to do, and what she has most reason to do, as we have seen, depends on her values, not on her opinion. This will be spelled out more explicitly in Chapter 2.

### 3.1.3 Theoretical and practical agency are both expressed through rational guidance

There is an apparent asymmetry between theoretical and practical reasoning in that it appears, \textit{prima facie}, as though belief formation is involuntary, whereas intention-to-act formation is voluntary. However, this distinction is not so clear; as I shall argue, both practical reasoning and theoretical reasoning rely on voluntary and involuntary mechanisms, and rational agents guide both their theoretical and

\(^{15}\) This does not entail that there are can be no universal truths about moral oughts. If all agents have the relevant moral values, as I suggested in the previous footnote, then the question of what agents morally ought to do would be akin to the question of what agent ought to believe. What agents morally ought to do would still be dependent upon their values, but this would not lead to a lack of universalizability: what agents ought to believe is still dependent upon their theoretical goals, it’s just that the goal (to believe that which is true) is universal. If the goals of morality (to promote moral values) are universal then what an agent ought to do, morally, will be universal.
practical reasoning via such and in Chapter 2 I argue that an agent is practically rational to the extent that she guides her actions in this way. In this section I show why the apparent asymmetry between the way in which theoretical and practical reasoning is under voluntary control is not real.

Many philosophers and psychologists maintain that belief is an involuntary disposition formed in response to perceived evidence; comprehension of theoretical reasons causes one to form the corresponding belief, without one having to make a decision as to what to believe. This is not to say that one does not sometimes have to do something that looks like deciding what it is that one ought to believe – for example whether to conclude from perusal of the evidence that one should believe that there will be a Conservative government or a Labour government after the next general election – but such cases can be described as acceptance of the proposition $P$, rather than belief in the truth of proposition $P$.

Acceptance is a state of mind distinct from belief. For example, van Fraassen (1980) argues that scientists should accept their theories, rather than believe them; literal belief is reserved for a theory’s claims about observables, while the remaining theoretical apparatus is to be used rather than believed, for the purposes of hypothesis formation, experimental design and so forth. Acceptance doesn’t just ask whether $P$, but also considers the cost and benefits of accepting or failing to accept $P$, which are often independent from the truth of $P$. For example, we might decide to cease inquiry and to act as though the matter is settled, because it is more important to have an answer than it is to have the answer. Thus in deciding whether or not one ought to believe that the next government will be Conservative, the pressing nature of coming to a conclusion influences one’s deciding to accept the answer ‘yes’. Acceptance tolerates a number of aims; it does not just aim at truth, as does belief. It also is more open to being under the voluntary control of the agent that is belief. In these two ways practical intention-formation is more analogous to acceptance than it is to belief.
Practical reasoning is more similar to acceptance of propositions than belief formation, because at some point the agent comes to a decision to cease inquiry into what she should do, and settle the matter by the formation of an intention to act, or by acting. Beliefs are cumulative and constantly open to revision, and often one does not need to make up one’s mind about whether or not \( P \), but can suspend belief until sufficient evidence accumulates. But this is rarely, if ever, the case in practical matters, and by the time all the evidence has rolled in the time for action may have passed. An agent must make up her mind one way or the other, and sometimes this must be done on unsatisfactory consideration of her reasons.

However, that practical reasoning is more analogous to acceptance than belief does not mean that practical rationality is always and completely the result of mechanisms under the voluntary control of the agent. Sometimes the agent does not make a decision to cease inquiry and to settle the matter by the formation of an intention to act; she either does not enter into practical deliberation about what she should do before acting, or else she does enter into practical deliberation, but acts before she reaches a cognitive conclusion. But in these cases we would say that the agent necessarily acted *involuntarily*. Usually, when saying that someone did something involuntarily, we mean that they were compelled in some way, *against* their will, rather than that they didn’t form a judgement or intention about what to do before they did it. In the same way, if we said that someone believes something ‘involuntarily’ we would mean that it was in some way against their will. Thus an agent’s *autonomy* can be expressed in belief-formation and intention-formation, even though the mechanisms by which beliefs and intentions to act are formed are not under their voluntary control, if the mechanisms are under the agent’s *guidance*.

For example, some philosophers suggest that there *are* cases of agents voluntarily forming beliefs. A person diagnosed with cancer might be aware of the research that suggests that people are much more likely to recover from cancer.

---

16 See Chapter 7 for argument about the role of intentions.
if they believe that they will do, thus it would be prudent for the person to believe
that he will survive, even if he knows that there is sufficient evidence for holding
that belief – an example of what William James calls cases, “where faith in a fact
can help create the fact” (1979, p. 25). This would suggest that there are
pragmatic reasons for belief as well as evidential reasons for belief, something
like those involved in the acceptance of Pascal’s Wager. However, Pascal’s
advice to the non-believer’s objection that “[I] am made that I cannot believe.
What, then, would you have me do?” (Pascal, 1941, p. 66; §233), is that she
should take steps so as to believe – steps such as attending mass and
“lesseni[ng] the passions” that prevent her from believing. The advice is to get
herself to the perspective of a believer, whereby their operative reasons for belief
in the existence of God can then become her operative reasons for believing in
the existence of God. It is not that the conclusion of the wager is an operative
reason for her belief; rather the conclusion of the wager is an operative reason
for her action of going to mass et cetera which will put her in a position where
what she will take to be evidential reasons will move her to form of a belief in the
existence of God. She is not believing at will, but acting at will: her belief is still
an involuntary disposition formed in response to perceived evidence, but she is
guiding her beliefs, and will be responsible for her beliefs, by virtue of the way in
which she influences the evidence to which her involuntary belief formation
mechanism responds.

Similarly with the case of our cancer patient, what she can do, in order to get
herself to the perspective of one who truly believes that she will survive, is to do
things like disallowing herself from reading research into likely survival rates,
researching her symptoms and entertaining thoughts of death, and engaging
instead in the sorts of activities in which those who truly believe they will survive
engage, such as making plans to go on holiday the following year and starting
long-term projects. In this way, she will be doing precisely the opposite of what is
argued, by Evidentialists, are the ethical norms of belief formation: that one ought
to believe that, and only that for which one has sufficient evidence, and that one
ought to do all that one reasonably can to seek out all available epistemic
evidence for a belief, since she will be wilfully ignoring certain evidence and
opportunities to seek new evidence. She will not directly believe at will, but will instead act voluntarily, in ways that will make it more likely that she will form the belief she wishes to. It is not, therefore, that she has prudential reasons to believe – in the sense in which I have defined having reasons, in that there is a true proposition \( R \) which could move \( S \) to believe that \( P \) – since \( R \) (that people are more likely to survive if they believe that they will) is not evidence for the truth of \( P \) (that she is going to survive). Rather, \( R \) is a reason to behave as if she believed that \( P \), with the hope that such behaviour will lead to the state of affairs where she will adopt the same operational reasons to believe that she will survive as those who already believe they will survive have for their belief. These ‘prudential reasons’, then, are not theoretical reasons for belief, but practical reasons for action: given our patient’s desire to survive, it is practically rational for her to behave as if she believed she would. Her end – to survive – is a practical end. The end of theoretical rationality is a theoretical end – to believe that which is true. Our cancer patient will be purposefully avoiding fulfilling this end, but that does not entail that she will be forming a belief at will, but that she will be voluntarily exerting influence over the reasons on which her belief-formation mechanisms operate.

Just as one can put oneself in positions where one is more or less likely to form certain beliefs, one can put oneself in positions where one is more or less likely to form certain intentions to act. We put ourselves – deliberately – into situations where we are more likely to make one decision over the other, just as our cancer patient deliberately puts herself into a situation where she is more likely to form the belief that she will survive. For example, a dieter would be well advised to keep his kitchen cupboards free from the sorts of foods that might thwart his appetite.

\[\text{For example, Descartes maintains that when forming a judgement, "it is clear by the natural light that perception of the intellect should always precede the determination of the will" (1984, p. 41; §60) and Locke that, "[h]e that believes without having any Reason for believing, may be in love with his own Fancies; but neither seeks Truth as he ought, nor pays the Obedience due to his Maker, who would have him use those discerning Faculties he has given him, to keep him out of Mistake and Errour (1824, pp. 219; Bk IV, Ch XVII, §24). William Kingdon Clifford argues that we are obligated to seek evidence, be open to new evidence, and to consider the evidence of others. He criticised a hypothetical ship-owner for pushing aside worries he had as to the fitness of his ship for the prudential reason that repairs would be costly, and concluded with the so-called ‘Clifford’s Principle’, that "It is wrong always, everywhere, and for anyone to believe anything on insufficient evidence" (Clifford, 1877 [1999]).}\]
efforts to consume fewer calories, in recognition of the fact that when hungry he is liable to decide to eat high-calorie foods if they are available. Ideally, our dieter would decide to consume low calorie foods whatever was available, since, given his end of weight-loss, he would decide to do whatever it was that best met that end. However, he has multiple ends some of them will conflict with his end of losing weight, and in the moment of deciding what to eat the instrumental and realizer desires to meet other ends such as deriving maximum pleasure from food, might lead him away from doing what he in fact has most reason to do.

In the theoretical domain, we can choose to read reliable sources of information over unreliable sources (reading a broadsheet rather than a tabloid, for example), to spend our time with those we recognize as being more likely to reach rational conclusions (a respected colleague, rather than the village busybody), to discuss our beliefs with friends who will not shy away from challenging us if they think us wrong. In the practical domain we may avoid situations where we are liable to decide to do things not in our best interests, such as dining with the wife of a friend to whom we are very attracted, becoming drunk at tense family gatherings, or filling our kitchen cupboards with crisps when we are trying to eat healthy snacks. These methods are part and parcel of human rationality. As finite beings we do not have the time or the capacities to assess every single fact against every other fact when forming a belief or coming to a decision about what to do. We have many shortcuts which we need to, and do, use in enabling ourselves to be rational; to respond to reasons\textsuperscript{18}. In the next Chapter I will argue that the most obvious way in which we can respond to reasons – by forming judgement as to what we have most reason to do and then forming an intention to act on such – is not the only way in which we can be reason-responsive. An agent is practically rational to the extent that she responds to those reasons which she has for action, a capacity that is expressed by her being a self-regulative reason-responder.

\textsuperscript{18} For example, Louise Antony argues that since recent empirical studies show that humans rely on ‘shortcuts’ and ‘tricks’ in knowledge acquisition, so having such cannot be disparaged as irrational (2000, p. 115). By the same approach, the goal of practical rationality is to do things that we have most reason to do, and so relying on the sort of short-cuts that we need to in order to do so cannot be disparaged as irrational.
Chapter Two

An Agent is Practically Rational to the Extent that Her Action is Guided by the Norm of Doing What She Has Most Reason to Do

In this second Chapter I argue that an agent is practically rational to the extent that her action is guided by the norm of doing what one has most reason to do, and that the way in which an agent does this is by correctly responding to the reasons for action she has to which she has access.

In Chapter 1 I argued that practical rationality consists in doing what one has most reason to do, where by ‘reason’ is meant ‘normative reason’. I argued that the normative force of a reason is derived from whichever values it is by virtue of which it counts in favour of the action for which it is a reason, and that the ‘having’ relation in which an agent stands to reasons is motivational, such that what it means for a reason to be had by an agent is for it to be capable of moving her in the direction of that action. However, as I will argue in this second chapter, whether or not the reason does move her towards the action depends (as well as upon her being practically rational) on whether she has access to the reason. Without access, the agent cannot act for the reason, even though she has the reason. Thus, as I argue in section 1, the standard against which practical agency is to be measured is not that of doing what one has most reason to do, but of doing what one has most accessible reason to do, which I call the ‘subjective ought’. In section 2 I argue that this norm is a regulative norm; that is that it is because her action is guided in line with the norm, rather than because it meets the norm as a ‘target’ that an agent’s action counts as rational. In section 3 I argue that in order for an agent’s actions to be regulated by the norm of doing what she has most reason to do, the capacity to engage in practical reasoning is vital, but, as I argue in section 4, actions do not have to be the direct result of practical reasoning in order to be so regulated; that is, engaging in conscious guidance of one’s action via practical reasoning is not necessary for an action to be rational,
but the agent’s ability to have done so, is. An agent is practically rational, I conclude in section 5, to the extent that her action is guided by the norm of doing what she has most reason to do, which she achieves by correctly responding to her reasons for action. This picture of practical agency makes room for the role for emotion within practical rationality, which I argue for in Chapters 4 onwards.

1 The relevant norm when assessing the rationality of an agent is that of the subjective ought

In the previous chapter I argued that S has a reason to Φ iff there is a true proposition R which counts in favour of Φ-ing and could move S to Φ. My argument was not that a true proposition in itself can motivate action; S has to be in some way aware of the reason for it to move her towards the corresponding action, whether that is consciously or non-consciously. Thus S could have a reason and yet not in fact be moved by it because she is unaware of it. In the example of Chapter 1, I had the reason ‘it is gamey’ to choose the Burgundy, and yet I was not moved to choose the Burgundy, because I was not aware of this reason. Aliyah was not moved to choose the Burgundy because she did not have the reason, since she did not value gamey-flavours in wine. The difference between Aliyah and me was that I could be moved to choose the Burgundy for this reason were I aware of it, whereas Aliyah could not. It is the capacity to be moved, rather than the being moved, which determines whether or not an agent has a reason, but as well as having a reason an agent needs to be aware of the reason in order to be able to act for it. Because an agent can act for a reason only if she is aware of it, an agent can only be expected to act for those reasons of which she could be aware, and the class of reasons of which an agent could be aware is what I mean by those reasons to which she has ‘access’. Practical rationality consists in doing what one has most reason to do, but this is the ‘objective’ ought; what an agent objectively ought to do is that what she has most reason to do. But this is not the standard against which practical agency is to be measured. The standard against which rational agency is to be measured is that of the ‘subjective’ ought, which is whatever it is that the agent has most accessible reason to do.
An agent can only act for a reason if she both \textit{has} the reason and is \textit{aware} of the reason. Only if these two conditions are fulfilled can the agent be moved by the normative force of the reason towards the action in favour of which it counts. However, having a reason and being aware of that reason are not sufficient conditions for being moved by a reason. Imagine that Felicity also joins our dinner party. Felicity both \textit{has} the reason ‘it is gamey’ that counts in favour of her ordering the Burgundy, and she is \textit{aware} that she has this reason, yet she fails to choose the Burgundy. She does not fail to choose the Burgundy on the basis that she judges this reason to be ‘outweighed’ by other reasons that count in favour of not ordering the Burgundy; she just fails to be moved by her reasons. There are a number of ways in which she could fail in this respect, including failing to form the \textit{intention} to choose the Burgundy despite having formed the \textit{judgement} that she had most reason to do so, or correctly forming the intention but failing to act upon it for some reason, or simply failing to consider the reason when deciding what to do. In failing to be \textit{moved} by the normative force of the reasons she has of which she is aware, Felicity would be failing to \textit{respond} to her reasons as reasons; she has failed to be moved by them as she ought to be. Thus it is not the case that all an agent needs in order to be practically rational is to have awareness of the reasons which she has. She needs also to be moved by their normative force; she needs to \textit{respond} to the reasons she has for action \textit{as} reasons for action.

The correct way for an agent to respond to the reasons she has for action as reasons for action is to be moved by their normative force. If $R$ is a reason $S$ has to $\Phi$, she should be motivated towards $\Phi$-ing to the extent of $R$’s normative force. If $R$ has little normative force then it may be that $S$ should have very little motivation to $\Phi$ on the basis of $R$, though of course she may have a number of other reasons to $\Phi$ whose combined normative force is strong, and if she responds correctly to these reasons she will have substantial motivation to $\Phi$. Even if she does, she might have even stronger reasons to not- $\Phi$, thus despite feeling a significant motivation in the direction of $\Phi$-ing, she will have greater motivation to not- $\Phi$. In this case, if she correctly responds to her reasons for
action she will, despite being moved towards $\Phi$-ing, not be moved far enough in its direction to actually $\Phi$. The reasons she has to not- $\Phi$ will be the reasons for which she acts. To respond to a reason is to be correctly moved by the normative force of that reason, and to act for a reason is to act on the basis of having been moved to act by the normative force of that reason. Thus in order to act for a reason an agent needs first to be responding to that reason. This means that she could fail to do what she ought to do despite responding to reasons and acting for reasons. If she correctly responds to all of her reasons then she will do what she ought to do, for she will be moved appropriately by all of the normative weights of her reasons and will be motivated to do that which she has most reason to do. But she might correctly respond to only some of her reasons, either completely ignoring others, or mistaking their normative weights, and thus be moved too far or not far enough towards the action in favour of which they count. If what she in fact has most reason to do is to not- $\Phi$, but she does have reasons that count in favour of $\Phi$-ing, and she responds to these reasons and on the basis of them $\Phi$s, then she will be acting for normative reasons, despite having not done what she has most reason to do.$^{19}$

$^{19}$ Some philosophers construe rationality as correctly responding to reasons, for example Joseph Raz argues that, “proper functioning consists in proper responsiveness to reasons” (1999, p. 19). For a detailed objection to construing rationality as such, see John Broome’s “Does Rationality Consist in Responding Correctly to Reasons” (2007). Broome’s objection to the thesis is that although rationality entails that one will respond correctly to reasons, it is not in what it consists. He concludes that rationality must be an independent source of requirements in its own right. I don’t think his argument is a problem for my thesis: ‘correctly’ responding to reasons, for human agents, involves complying with what he calls the ‘requirements of rationality’. For example, Broome argues that it is a requirement of rationality that one does not have any contradictory intentions, but that correctly responding to reasons does not necessarily rule this out, thus there must be something over and above the requirement that agents correctly respond to their reasons. But, unless there was in fact equal normative weight for two contradictory actions, for example to $\Phi$ and to not- $\Phi$, correctly responding to reasons would rule out the possibility of holding contradictory intentions, since if the agent correctly responded to her reasons she would form the intention to do that for which there was most normative force and not form the intention to do that for which there was less normative force. And if there was indeed equal normative weight behind $\Phi$-ing and not $\Phi$-ing, this, if she correctly responded to her reasons, would not lead the agent to forming an intention to do both, anyway. Correctly responding to reasons would lead her to form the intention to do whatever she had most reason to do, and in the case of there being nothing for which she has most reason, she wouldn’t form an intention to do either. She might have to make a decision, of course, but then, as I argue in Chapter 3, what she should do is re-deliberate, perhaps taking in to account additional reasons or examining those already considered to decide whether they do really have the normative force she first accorded them. So correctly responding to reasons rules out the possibility of an agent holding contradictory intentions, and thus we do not need to posit the existence of an independent source of the requirements of rationality. Instead, we can see rationality as being, as does Niko Kolodny, not a disposition to act upon a special class of reasons, but as “a kind of executive virtue” (Kolodny, 2005, p. 544) Kolodny argues that some virtues are dispositions to respond correctly to certain kinds of reasons,
Since, as I argued in Chapter 1, the ‘having’ relation in which an agent stands to normative reasons is that of the capacity to motivate, it is clear that an agent can only respond to those reasons for action which she has. Additionally, she can only respond to those reasons of which she is aware. I could not be moved by the reason ‘it was gamey’ because I did not know of the reason’s existence, thus an agent can only respond to those reasons she has of which she is aware, and can thus act for only those reasons of which she is aware. There may be reasons an agent has of which she is not aware, but of which she could be aware, and these reasons – those to which she has ‘access’ – because she could respond to them were she to do whatever was needed in order for her to become aware of them, we can include in the class of reasons to which an agent can respond and act for. Thus we can expect an agent to respond to all of those reasons she has for action to which she has access. Because this is as much as we can expect of an agent, this is the standard by which rational agency should be judged. Thus although the ideal of practical rationality – doing whatever it is that one has most normative reason to do – is what an agent ‘objectively’ ought to do, the standard by which we should measure practical agency is that which the agent ‘subjectively’ ought to do – that which she has most accessible reason to do. These uses of ‘objective’ and ‘subjective’ do not map on to the objective/subjective reasons distinction that I mentioned in Chapter 1 (this fact

“[k]indness is a disposition to respond to the needs of others, for example, and justice is a disposition to respond to considerations of fairness” (ibid., p.553), but rationality “is a disposition to execute one’s beliefs about one’s reasons for and against one’s attitudes” (ibid., p. 554; italics added). I will likewise argue that the disposition to reflect upon one’s reasons and to have one’s judgements about such guide one’s actions is necessary for practical agency. For a human agent to be able to correctly respond to reasons she needs to have such a disposition, thus part of what it is to correctly respond to reasons is to be guided by the norms of rationality. 20 I am not concerned to present an account of the reasons to which an agent has access, but suffice to say, the class of reasons to which an agent has access will often be broader than the class of reasons of which she is aware. The class of reasons to which an agent has access is going to be something like ‘those reasons of which we can reasonably expect her to be aware’. For example, we might say of Bernard Williams’ famous example of the agent who drinks petrol thinking that it is gin (1980, p. 102), that not only was he not aware of the reason he had to not drink the contents of the glass (i.e. that it was petrol), but that he did not have access to this reason either, for it would not be reasonable of us to expect him to check whether or not the glass contained petrol. However, we might say that once he brought the glass to his lips he then did have access to the reason as the smell ought to have alerted him to it, or that had there been other grounds, such as knowing his host likely to play such a prank, we could reasonably expect that he have some awareness of the reason. This is not to say that the class of reasons to which an agent has access encompasses all reasons of which she could, if the conditions were right and she tried hard enough, become aware; this would be not be a reasonable expectation.
being another reason to avoid adopting that terminology), since both oughts refer to the agent’s normative, rather than operative, reasons. What an agent ‘subjectively’ ought to do is that which she has most accessible normative reason to do.

When I did not choose the Burgundy, in the example from Chapter 1, if I was doing that for which I had most accessible reason, I was doing what I subjectively ought to do. Thus, despite not managing to do what I objectively ought to have done (that is, choose the Burgundy), I was practically rational. I don’t think this should be thought of as a controversial claim: analogously, in the theoretical domain, we would not call an agent irrational if she had a reason to believe $P$, yet was unaware (and reasonably so) of this reason and thus did not believe $P$. If I had a false belief that it would snow tomorrow on the basis of the newspaper’s weather forecast, not realizing that the newspaper was last week’s, my belief would not be irrational, but merely false. I just got it wrong, but I did not fail to be guided by the norm for belief formation (that of believing what I ought to); my belief was grounded in the best evidence available to me.

We may say, with hindsight, that we ought to have done something because of certain reasons that counted in favour of doing it, even though at the time of forming the intention to act we did not have access to these reasons. For example, imagine I have chosen to order a glass of Pinotage in the hope that it will exhibit some gaminess while Cedric orders the Burgundy. I try a sip of Cedric’s Burgundy, and then exclaim, “Oh, I ought to have ordered the Burgundy! I didn’t realise it would taste like that.” I might then claim that I ought to have ordered the Burgundy, but this is a different claim to saying that at the time I ought to have ordered it. At the time of making my wine choice I did not have access to the reason for choosing the Burgundy that ‘it tastes gamey’, thus to level a charge of irrationality against me for my decision would be misplaced. It is true that I could have asked Cedric whether or not the Burgundy would be to my liking, but unless it even occurred to me that the pinot might be gamey, this looks like a
rather steep requirement.\textsuperscript{21} We should instead say that I \textit{objectively ought} to have ordered the Burgundy – something I can appreciate with hindsight – but that I \textit{subjectively ought} – the subject being me at the time I made my choice – to have ordered the Pinotage. I can only act \textit{for} reasons of which I am (not necessarily consciously) aware, and given that I was not consciously aware that the Burgundy would be gamey unless I had some non-conscious awareness that it would be gamey, I could not have acted for this reason.

One might object to my claim that I was not irrational. One might want to insist that in not choosing the Burgundy I \textit{was} irrational, on the basis that it was not what I \textit{objectively} ought to do. But this would have a very odd consequence. If one argued the rational thing for me to do, on the basis that it was what I objectively ought to do, was to choose the Burgundy, and I did do, \textit{for} what reason could I have done so? I could not have done so for the reason ‘it is gamey’ given that I had no access to it. Assuming that I did not have non-conscious access to this knowledge, it would seem very strange had I chosen the Burgundy, given that I did not believe I would like it. It would look as though I was not acting for reasons at all, in which case I wouldn’t be acting rationally; I would have \textit{accidentally} done that which I objectively ought to have done, and, as I shall argue in section 4, accidentally doing what one ought to is not an expression of rational agency.

What an agent \textit{subjectively} ought to do is dependent upon the \textit{first-person perspective}: as I shall argue in Chapter 3, only \textit{the agent herself} can decide what she subjectively ought to do, for only she has the necessary access to the reasons she has and only she can be moved by their relative normative forces. When we say of another that they ‘ought’ to do something, we are making an \textit{objective} claim; though this is not to say that from the third person perspective we are actually in a better position to judge what someone else in fact has most reason to do. We are rarely, if ever, in the necessary position of omniscience that

\textsuperscript{21} There would need to be good \textit{grounds} for me to question my decision to justify my doing so, just as Williams’ petrol drinker in the footnote above could be rationally criticized for taking a sip if he had good grounds to suspect his glass contained petrol.
would give us the required knowledge of all the reasons another agent has. We can be put into this position as readers of novels, of course. For example, it is clear to us that Dorothea Brooke in George Eliot's *Middlemarch* objectively ought to not enter into her fatal marriage with Edward Casaubon: her sense of fulfilment in the marriage is vested in her hope to become educated (“what a lake compared with my little pool!”), to have her curiosity nurtured and to be of constant usefulness to a man of sixty who really needs her nineteen-year-old eyes for reading. That she should not marry Casaubon is clear to the reader, for the reader can appreciate that these reasons for which Dorothea acts are not adequate reasons for marriage. But Dorothea,

with all her eagerness to know the truths of life, retain[s] very childish ideas about marriage. She [feels] sure that she would have accepted the judicious Hooker, if she had been born in time to save him from that wretched mistake he made in matrimony; or John Milton when his blindness had come on; or any of the other great men whose odd habits it would have been glorious piety to endure (Eliot, 1967, p. 12)

*We* can say that Dorothea objectively ought not to have married Casaubon, but *we* have the advantage of Eliot's third-person objective narration of the events. Practical agents do *not* have this: they deliberate from the *first-person perspective*. They are very unlikely to have access to all of their reasons for action, thus it is a very different claim to say that Dorothea – the actual agent that she is at the time she accepts Casaubon’s proposal – *subjectively ought* not to have done: we should level a charge of irrationality against Dorothea for her decision. Perhaps Dorothea was irrational, but if so, not by virtue of her failing to see her reasons from that of the objective third-person. To say that she was for such would be to say that we humans are *incapable* of being practically rational, since we *cannot* take the objective third-person perspective in response to our own reasons (a position for which I shall argue in Chapter 3), and by analogy that we could not even be theoretically rational, since omniscience eludes us.

It is important to stress that my claim that the standard against which rational agency is to be measured is that of the subjective ought does *not* mean that *whatever* an agent decides she should do actually *is* what she subjectively ought to do. We often fail to do what we *subjectively* ought to do, for example because
we overlook reasons to which she have (conscious or non-conscious) access, or because we incorrectly assess the normative weights of our reasons, and thus we fail to be guided by the regulative norm as we ought to be. But it does mean that the existence of reasons to which an agent did not have access can entail that there can be a difference between what she objectively ought to do and what she subjectively ought to do, and the charge of irrationality cannot be levelled against an agent who fails to do what she objectively ought to do, but succeeds in doing that which she subjectively ought to, if as I shall argue, she does so because she is guiding her behaviour in line with the norm of doing whatever it is that she has most reason to do. Bernard Williams’ example of man who drinks petrol thinking that it is gin (1980, p. 102) would, on my account, have a reason to refrain from drinking the contents of the glass (that being that it is petrol), but this reason is one to which he does not have access, and, therefore, one for which he cannot act. Thus Williams’ agent objectively ought not to drink the contents of the glass, because he has the reason not to, but given that he does not have access to this reason, we cannot say that he subjectively ought not to drink the contents of the glass on the basis of this reason, and it is this subjective ought against which his rational agency is to be measured.

Since what an agent has most accessible reason to do is still ‘a matter of fact’, by ‘subjective’ is not meant ‘as according to the subject’s judgement; what an agent subjectively ought to do is not a question to which she has privileged access, such that she can say something to the effect of, “What I ought to do is whatever I think I ought to do”. The question of what she subjectively ought to do is ‘objective’ in that it is not determined by her opinion, but is ‘subjective’ in that it is dependent upon certain facts about her, as the subject. In Williams’ gin-drinker example, the objective fact that the man doesn’t know that the glass contains petrol means that what he subjectively ought to do is not the same as what he objectively ought to do, but that, for example, he believes that he ought to drink the contents of the glass does not make it the case that this is what he subjectively ought to do. It is not the case that whatever an agent thinks they ought to do is, in fact, what they ought to do in my subjective sense of ought.
Since an agent can act only for those reasons to which she has access, and, as she does not have access to all of her reasons, she cannot be expected to be always able to do that which she has most reason to do. She can only be expected to be able to do that for which she has most accessible reason to do, thus the correct standard against which the extent of her rational agency should be measured is that of the subjective ought. The charge of irrationality should not be levelled against the agent who fails to do what she objective ought to do, but succeeds in doing what she subjectively ought to do, thus from now on when I talk of an agent doing what she 'ought' to do, I am referring to the subjective (that for which she has most accessible reason to do), unless I specifically state otherwise. Further, the charge of irrationality should not even necessarily be levelled against the agent who fails to do what she subjectively ought to do, for, as I shall now argue, the norm is regulative rather than a target; it is by virtue of the fact that an action is guided by the agent in line with this norm that it is rational action, not by virtue of the fact that it achieves satisfaction of the norm.

2 The norm of practical agency is regulative

An action can be rational even if it is not in fact what an agent subjectively ought to do, because rational agency is not an all-or-nothing state, but a matter of degree. At one end of the spectrum we have the ‘ideally’ rational agent, who does what she subjectively (and even objectively) ought: she does precisely that which, as a matter of fact, she has most accessible reason to do, for precisely those reasons which make what she does precisely what she ought to do. If we held this to be a ‘target’ ideal of practical rationality, holding that agents are practically rational only when they hit this target, we would be conceding that we, as humans, are hardly, if ever, practically rational. We are limited creatures, in many ways, and as such we are unlikely to meet such a target ideal, and are also unlikely to be able to know when we have or have not done so. And so, to avoid this unwelcome entailment, my picture of human practical agency holds that agents are practically rational to degrees, with the norm of doing whatever it is one has most accessible reason to do being a ‘regulative’ ideal rather than a ‘target’ one.
As a regulative ideal, doing whatever it is that one has most accessible reason to do is a norm that guides an agent’s behaviour by exerting normative guidance over her practices, and is thereby useful even if not attainable. An agent is practically rational to the extent that her behaviour is guided by this norm, such that if she is perfectly guided by it this will entail that she does what she subjectively ought to do, and the more guidance the norm exerts over her behaviour the closer she will get to doing what she has most accessible reason to do. Of course, an agent might coincidentally do what she has most accessible reason to do without the norm having exerted any regulative guidance over her behaviour at all, but such an agent, although she has done what she in fact has most reason to do, will not have done so for those reasons, and thus her action does not merit rational praise. It is a ‘lucky accident’ that she happens to have done the right thing; her doing what she ought to do was not the result of her exercising her practical agency. In section 5 I shall argue for this claim, that practical agency required that the agent acted for those reasons which made what she did what she ought to have done, but in this section I shall explain what it is for a norm to be a regulative norm, rather than a target norm.

A regulative norm works analogously to the way in which a thermostat regulates the temperature of a room. In a thermostat, one ‘component’ continuously monitors the state of the temperature in relation to an externally set value, say of 22 degrees centigrade. If the system departs from this set-point value, the

---

22 It will not necessarily bring her closer to what she objectively ought to do. It may be that what agent S has most reason to do is to Φ, but the reasons for which she ought to Φ are not all accessible to her. According to the reasons she has to which she has access, she ought to Ψ; thus if she is perfectly guided by the norm of doing what she has most accessible reason to do, she will Ψ. It could be that had the norm exerted less guidance over her behaviour she would have actually been more likely to Φ, for she may have had access to some of the reasons she had to Φ, and been swayed by the normative force of these, even though these reasons on their own did not justify her Φ-ing in the light of the normative force of her reasons to Ψ: the reasons she had which in fact made it the case that she ought to Φ – those with the decisive normative force – were not accessible to her, and so were not those to which she was responding, and so were not those for which she acted. Thus although the norm of practical agency (the subjective ought of doing what one has most accessible reason to do), aims to track the norm of practical agency (the objective ought of doing what one has most reason to do), being guided by it will not necessarily bring one closer to meeting the objective ought. It will however, bring one closer to meeting the subjective ought, which is why the norm exerting guidance is the subjective, not the objective.
monitoring module sends a signal to second component which modulates the inputs into the system – by switching the radiators on– until the set-point value is restored. The signal then ceases, and the modulation of inputs stops. A norm that specifies regulative ideals likewise involves some form of self-monitoring for conformity to a standard or aim, departures from which cause the agent to make corresponding alterations in her course of action (Railton, 2006, p. 10). One ‘component’ continuously monitors the state of the agent’s behaviour in relation to the standard of doing what she has most reason to do. If her behaviour departs from the norm, a ‘signal’ is sent to a second ‘component’ which modulates the ‘inputs’ which will bring her behaviour back in line with the norm of doing what she has most reason to do. Let us see if we can cash out this analogy in a way that sheds light on how the norm of the subjective ought of practical rationality can regulate an agent’s behaviour.23

The agent’s behaviour is analogous to the room’s temperature, with the regulative ideal of the temperature being 22 degrees and the regulative ideal of the agent’s behaviour being doing what she has most reason to do. The thermostat is composed of two ‘modules’, one monitoring, one modulating, and while I certainly do now want to claim that human behaviour is controlled by an analogously modular system, that there are two sorts of ‘processing’ involved is an idea that is consistent with the currently popular ‘Dual Processing’ accounts of cognitive psychology.24 Dual Processing theories maintain that humans have two distinct mechanisms for thought processing. One is fast, automatic and non-conscious (System 1), the other slow, controlled and conscious (System 2). Conscious reflection about what one has most reason to do, i.e. practical reasoning, is a

23 I am not here claiming that, as Karen Jones puts it, we are “merely [systems] – however well functioning – of subsystems, that passively register and respond to environmental stimuli much as a thermostat registers and responds to changes in temperature” (2003, pp. 188-189). I agree with Jones that to think of oneself in this way is to “give up thinking of [oneself] asrationally guiding [one’s] actions via [one’s] reasons”, for “reasons [are] not merely registered, but understood as reasons, that is, understood as justifying the performance of an action” (ibid., p.189). As will become clear, my argument is that our capacity for practical reasoning, which is not a passively-registering-and-responding mechanism, entails that we are rationally guiding our actions via reasons understood as reasons, even when our behaviour is being guided by the subsystems which are under its control and not directly by it itself.

24 I am not arguing for a modular account of human behaviour. For a related apprehension about assuming modular accounts of human behaviour in the field of emotions and rationality, see Jones (2006).
process of System 2. Those routes we might refer to as ‘non-rational’, for example emotion, are processes of System 1.

The way in which the systems work together is relevantly analogous to the way in which the two modules of the thermostat work together, and can thus provide some illumination as to how the subjective ought might work as a regulative norm. According to Dual Processing theories, most of the time we rely on System 1 thought processing. System 2 is activated only when it detects certain ‘cues’ from System 1. System 1 will generate such cues when System 2 is needed because there is some sort of error in the deliverances of one or more of the mechanisms of System 1, for example if a question arises for which System 1 does not have an answer, or when there is a conflict between the deliverances of System 1 mechanisms. Psychologist Daniel Kahneman explains it thus:

System 1 continuously generates suggestions for System 2: impressions, intuitions and feelings ... When all goes smoothly, which is most of the time, System 2 adopts the suggestions of System 1 with little or no modification ... System 2 is mobilized when … an event is detected that violates the model of the world that System 1 maintains (Kahneman, 2011, p. 24).

I propose that the regulative norm of doing whatever it is one has most reason to do exerts normative guidance over the practices of rational agents in this way. System 1 mechanisms, such as emotions, track reasons for action, and if there are grounds to suspect that the agent would not be doing what she has most reason to do by adopting these deliverances of System 1, a signal is generated which activates System 2; conscious reflection. The agent then engages in practical reasoning to ‘rule’ on the potential conflict or error of the deliverances of System 1; she consciously reflects upon the reasons to which she has access in order to decide what she in fact has most reason to do. Once she has ‘ruled’ on the matter the ‘signal’ ceases and thus she has ‘recalibrated’ her System 1 mechanisms to better track her reasons for action, which means she will only need to again engage in practical reasoning if something else happens which this ‘ruling’ did not cover, when another ‘error signal’ will be sent. In Chapter 3 I will present this picture of practical reasoning in more detail, and in Chapter 6 I will
argue that emotions play a number of vital roles in the process, but now I will argue that because agents have this capacity to engage in practical reasoning – because System 2 ought to be activated when needed - their behaviour is guided by the regulative norm of doing what they have most reason to do. Thus the capacity to engage in practical reasoning is a requirement of practical agency.

3 The capacity to engage in practical reasoning is a requirement of practical agency

The least controversial claim about how we manage to guide our actions in line with the norm of doing what we have most reason to do is that we do so by engaging in conscious deliberation about what we have reason to do; i.e. we use practical reasoning. This seems the most obvious way in which we can act for reasons, and indeed it may seem at first glance that this is the only way by which we can be practically rational. But that practical reasoning aims to achieve this end does not of course entail that it is necessarily the only means of achieving this end, and in section 4 I shall argue that it is not at all the only means of achieving this end, nor even is it always the best mechanism by which to guide our action in line with the regulative norm of doing what we have most accessible reason to do. Our System 1 mechanisms can achieve this end too, but can do so only because we have the capacity to engage in practical reasoning and thus ensure that these systems are in fact accurately reason-tracking. Therefore, as I argue in this section, the capacity to engage in practical reasoning is a prior requirement of practical agency.

I am making two claims. The first is that only those who have the capacity to engage in practical reasoning count as rational agents in the first place. The

25 In the wine choice example from Chapter 1, my not choosing the Burgundy was being guided by the regulative norm even though I did not in fact do what I had most reason to do, because the ‘system’ was working as it ought to. There was no ‘signal’ that my choice was incorrect, so I was rational to act upon my choice. Had there been grounds to question whether my judgement about what I had most reason to do was in line with what I in fact had most reason to do, then, if I were rational, I ought to have questioned it.
capacity to engage in practical reasoning is what distinguishes rational agents from creatures that merely exhibit behaviours, thus the capacity to engage in practical reasoning is a prior requirement of rational agency. This I take to be uncontroversial. The second claim is that it is only if the capacity to engage in practical reasoning is active that one is currently functioning as a practically rational agent, although, as I argue in section 4, one does not have to be currently engaged in practical reasoning in order to be functioning as a practically rational agent. Thus not all behaviour of those with the capacity to engage in practical reasoning is a candidate for rational action, yet behaviour which is not the result of *occurrent* practical reasoning can be. Both parts of this second claim are controversial, for some hold that practical reasoning has no special normative status, and others that it is the *only* process which has normative status. Views which entail either of these extremes should be rejected, as I shall now argue.

Being an *agent*, rather than just a creature that exhibits behaviours, involves being *self-governing*, so that when an agent acts it is *she* who acts. For an agent’s behaviour to be an *action*, it cannot be that it is a mere accident that she does what she does, because she must be *responsible* for having done what she did; her action is then an expression of her autonomous agency. An agent’s autonomy as an agent is grounded in the fact that she alone can initiate her actions. This means that she must regard her intentions to act as authoritative, for were she to challenge her own authority, her decisions about how to act would no longer be *her* practical conclusions, and their power to move her would not be a manifestation of *her* power to move herself; it would not be the power of her own agency. Thus if an agent fails to govern herself when she acts this must be because what she does is independent of her power to determine how she will act. When her actions are not under her control, she suffers from self-alienation: in one sense ‘she’ is doing something, but it is not an action of *hers*. Thus the products of practical reasoning – intentions – *must* have a special normative standing for if she behaves in a way that contradicts them, she has undermined the very conception of self-governing agency.
On the face of it then, the incontinent agent – the agent who acts against her all-things-considered judgement formed as the result of practical reasoning about what she ought to do – is not acting autonomously, even if she in fact ends up doing what she objectively ought to do. In Chapter 7 I will argue that this isn’t quite the case, for it is only agents who fail to act upon their intentions that are excluded from being considered self-governing, not agent who judge that they ought to Φ but who fail to form the intention to Φ, however for now let us consider the intuition that an agent whose action is not guided by the authority of her own judgement appears to be, at least on the face of it, to failing to act autonomously. This intuition makes it seem obvious that in performing an incontinent action, rather than the continent one, an agent necessarily does something irrational, no matter what the rationality of her judgement about what she ought to do, such that incontinence always adds an element of irrationality over and above any irrationality that might be involved in the formation of her errant judgement. This is a popular view, endorsed explicitly by Thomas Scanlon (1998, p. 25), Christine Korsgaard (1997, p. 222), Joseph Raz (1999, p. 16), and R. Jay Wallace, who expresses the appeal of the position when he argues that,

It is important to our conception of persons as rational agents ... that practical deliberation be correctly related to motivation. In particular, the motivations and actions of rational agents are guided by and responsive to their deliberative reflection about what they have reason to do. Unless this guidance condition (as we might call it) can be satisfied, we will not be able to make sense of the idea that persons are genuine agents, capable of determining what they shall do through the process of deliberation (1999, p. 219).

Wallace’s argument is that our conception of persons as rational agents includes that they guide their actions in response to the reasons that they have come to see as reasons as the result of deliberative reflection, i.e. practical reasoning. I will call such a view the ‘Responsiveness to Reasoning’ picture.

The Responsiveness to Reasoning picture gives the judgements about what we have most reason to do that are the results of our conscious deliberation special normative standing just in virtue of the kind of judgements that they are, and maintains that agent makes a rational failure if she fails to comply with this
requirement. The essence of practical agency, on such an account, is the capacity to *evaluate* one’s motives and to adjust these motives in response to one’s evaluations about what follows from one’s beliefs and desires. The rational ‘guidance condition’, as Wallace puts it, can be satisfied only by agents’ deliberative reflection about what they have most reason to do. But one can exercise this capacity despite having false beliefs about one’s reasons, thus the Responsiveness to Reasoning picture does not identify practical agency with being guided by *correct* evaluative and normative judgements, and it entails that an agent is irrational by virtue of being incontinent, no matter what the rational status of her all-things-considered judgement (atcj). Additionally, the commitment to rational guidance, on such a picture, has *unique* expression in acts of continence, which seems to entail that the vast majority of our behaviour, which is not the result of a conscious process of reasoning, is not practically rational. For these reasons I reject this view: the picture is far too narrow a view of practical rationality, because it fails, sometimes incoherently as I shall argue in section 4, to account for the rational status of actions that are not motivated by atcjs. However, I accept the intuition which makes the position so attractive; our conception of persons as rational agents includes that they guide their actions in response to the reasons that they have come to see as reasons as the result of practical reasoning. Thus we should reject the position at the other end of the scale which accords *no* special status to our practical judgements and holds that what it is for an agent to be governing her action is for it to simply cohere with some other attitude of hers. Such positions do not represent us as practical *agents*, and so cannot be correct.

I shall refer to these positions at the other end of the scale as ‘Coherentist’. The Coherentist has it that an agent governs her action if and only if her motives cohere with some other attitude that represents her point of view of the action. For example, Harry Frankfurt (1988) holds that this other attitude with which an agent’s motives need to cohere is her highest-order desires regarding which of her first-order desires moves her to act; Gary Watson (1982) argues that the relevant view is constituted by the agent’s evaluative judgements regarding which actions are worth performing; and Michael Bratman (1979) argues that in addition
there must be harmony between what the agent does and her more or less long-term plans. Such approaches have been shown to suffer from major problems. There are two main ones, as I see it, which mean we should reject such pictures on the grounds that they count as autonomous actions which quite clearly violate the requirement of self-governing rational agency. The first is that a person would be autonomous so long as she had no objection to her desires and their motivational effects, including those that were the result of an addiction. The second is that the origin and content of a person’s higher-order attitudes are irrelevant to whether she is an autonomous agent: she need have done nothing to bring it about that she has these attitudes and the attitudes need not be especially rational or well-informed, thus a third party could inculcate into the agent a first order desire and an endorsing second-order desire, and on such models the agent would still be called autonomous.26 But, as I said at the beginning of this section, being an agent, rather than just a creature that exhibits behaviours, involves being self-governing, such that when an agent acts it is she who is responsible for her action. Coherentist pictures allow behaviour for which the agent is not responsible to count as autonomous, and should thus be rejected.

Thus we are left with the intuition that the capacity to govern one’s behaviour via practical reasoning is vital to practical agency. It is what distinguishes us from non-rational creatures. Rational agents are not what Karen Jones (2003) calls, mere ‘reason-trackers’; systems that passively register and respond to environmental stimuli. The behaviour of non-rational agents, for example non-human animals, infants and those with severe mental incapacities can reflect the reason-giving considerations of their situations, yet such are not rational agents, because they are incapable of acting autonomously by guiding their own behaviour in line with their reasons. What separates us from non-rational agents is that we are able to engage in practical reasoning and form atcjs about what it is that we ought to do, which then lead to intentional action.27 This does not mean

26 See James Stacy Taylor’s introductory chapter to Personal Autonomy (Taylor, 2005, pp. 1-10) for a detailed discussion of the various problems faced by such hierarchical approaches.
27 As Timothy Scanlon argues, the capacities of the ‘rational creature include both the capacities to recognize, assess, and be moved by reasons … These reflective capacities set us apart from creatures who, although they can act purposefully, as my cat does when she tries to get into the cabinet where the cat food is kept, cannot raise or answer the question whether a given purpose
that practical reasoning is the only way in which we can be practically rational, but it does mean that the ability to engage in such is necessary to count as being a practical agent. If there were an organism that happened to perfectly track all of its reasons for action, and thus always happened to do what it had most reason to do, but which did not have the capacity to engage in practical reasoning, it would not count as a practical agent because it would not be responding to reasons, but merely tracking them. A thermostat might manage to keep a room at a constant 22 degrees but it would not thereby count as responding to reasons. Only those with the capacity to reflect upon their reasons for action count as responding, rather than just tracking, reasons. Organisms that reason-track can be moved to behave by reasons, but they are not being moved by the normative force of reasons. To be moved by the normative force of reasons an organism needs to be able to decide to act on the basis of recognizing that the reason to \( \Phi \) makes \( \Phi \)-ing what it ought to do. As Raz says, “it is people’s responsiveness to reason which accounts for their ability to act intentionally” (1999, p. 39; italics added).

For example, both humans and non-human animals have the fear response, which moves an organism to behave in certain ways, for example running away. The fear response is a reason-tracking mechanism in that it enables an organism to do what the reason-giving considerations of its situation support it doing. But these behaviours are non-rational, thus behaving in accordance with reasons is not enough to bestow rationality. The practical agent has the ability to reflect upon her reason-tracking mechanisms to decide whether or not to behave in accordance with their deliverances. Barbara is afraid of heights, but as a practical agent she can reflect upon her fear response, assessing whether the reason-giving considerations that her fear response is ‘latching on’ to really are good enough reasons to run away or not. When hiking around the cliffs on the south-east coast of Scotland, Barbara might experience the fear response at various

---

provides adequate reason for action” (1998, p. 23). Jones defines rational agents as “reason-responders” who “must possess and exercise a complex set of capacities if they are to respond to reasons understood as reasons. Among these capacities are the capacities to step back from any actional impulse and inquire whether the desire really reflects anything choiceworthy in the action” (2003, p. 189). See also Korsgaard (1999) whom I discuss in Chapter 3.
points during her hike: she experiences vertigo and wants to run away. But whereas an animal, infant or person with severe mental incapacity would do just that, Barbara can remind herself that there in fact is nothing to fear – she is perfectly safe – and carry on enjoying her day, despite her vertigo and urge to run. Practical agency includes the ability to reflect upon seeming reasons for action, in order to decide whether they do or not support the course of action which they purport to. In this way practical agents do not just track their reasons, but they guide their behaviour in line with their reasons.

Of course, just because Barbara can engage in practical reasoning does not entail that all of her actions will thereby be the product of self-governing agency (were this the case, rational agents would never be able to fail to be rational, which is of course not my argument). In section 5 I will argue that for actions to count as the products of rational agency they have to be under the guidance of the capacity for practical reasoning such that if she ought to consciously reflect upon the deliverances of her reason tracking mechanisms, the rational agent would. First, I will here argue that actions which are not the direct result of practical reasoning are nevertheless candidates for rational action, and then in Section 5 I will argue that what distinguishes such actions from those that are not candidates for rational action is that they are under this sort of guidance.

4 Behaviour does not have to be the direct result of practical reasoning in order to be a candidate for rational action

In the example above, Barbara judged that she ought not to run away, and intentionally acted in accordance with it, and thus was practically rational. However, we can imagine that Barbara does not need to do this. She has been hiking around these cliffs in south-east Scotland for some thirty years now, and although she still experiences vertigo, she no longer needs to deliberate about whether or not she should run away. After a number of occasions on which she did need to consciously deliberate, it has now become habitual for Barbara to respond differently to her fear; she no longer feels sufficiently motivated to run away. The absence of conscious deliberation does not make Barbara’s
perseverance with her hike any less rational: certainly, it does not make it irrational, but it does not render it non-rational, either. The infant’s automatically running away in response to fear is non-rational, for it is not subject to rational criticism; the infant is not in the business of being practically rational – he is not a practical agent. Barbara, on the other hand, is: by virtue of the fact that she has exercised her rational capacities to modify her response to fear she has exercised practical agency. Her prior deliberation is now guiding her current behaviour, even though it is not under the direct control of practical reasoning. She no longer needs to directly control her behaviour via conscious deliberation because her prior reasoning recalibrated her habitual responses to her fear. There is thus no need for System 2 to be engaged: System 1 is accurately reason-tracking. Her behaviour, although now being guided by System 1, is a result of the exercise of her rational capacities in engaging System 2 processing, and as such is subject to rational criticism. That it is subject to rational criticism does not entail that it will be rational, of course: if she acted against her intention to resist the urge to run away it would be irrational. But in as much as her behaviour reflects her previous judgement, despite it being non-conscious, it is a candidate for rational criticism, unlike the behaviour of infants and non-human animals.

I have argued that for an agent to be acting for a reason, she needs to be aware of it. It should be clear by now that I am not arguing that the agent needs to be consciously aware of it. If Barbara now hikes with no problem, she is acting for the reasons which support her not running away when she feels scared of heights, yet is not consciously aware of such, and does not need to be consciously aware of such for us to say that she is behaving rationally. I am thus agreeing with Arpaly that an agent needs to be acting for the reasons which make her action that which she ought to do, and that an agent can do so without conscious awareness of those reasons. But for such an account to work, we need a way of distinguishing when an agent acts for reasons of which they are not consciously aware, and when they act in accordance with their reasons, but

---

28 Arpaly further argues that for “an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons – that is, the reasons for which she acts are identical to the reasons for which the action is right” (2008, p. 72).
not for those reasons, otherwise acting for reasons collapses into simply happening to do whatever it is one has most reason to do, which would leave us with no way of distinguishing between the autonomous agent and the agent who does as she ought by accident. Thus we need a distinction between acting in accordance with reasons, and acting for reasons of which one is not consciously aware. To illustrate this distinction, let us look at how fast and habitual actions can be practically rational. The conditions under which they can be rational illuminates the conditions which separate actions which are done for reasons, from those that are not.

Consider Arpaly’s (2000) case of the inverse akrasia of Emily whose all-things-considered judgement is, and always has been, that she should pursue the PhD in chemistry for which she is currently reading. Emily is feeling a lack of motivation, sad and restless, and these feelings are in response to reasons such as not having the talents required. These are, Arpaly asks us to suppose, good reasons for Emily to not be pursuing the PhD, and Emily notices and registers these reasons, but ignores them when deliberating about whether or not to continue in her studies, and believes her feelings to be groundless. However, one day, she impulsively leaves the programme, and thinks herself irrational for doing so, yet experiences a sense of relief. Years down the line she realises the reasons that caused those feelings and cites them as the reasons for her not getting a PhD, and regards herself as irrational, not for quitting, but for not leaving the programme sooner than she did (2000, pp. 49-50). Arpaly argues that Emily is more rational for quitting her PhD than she would have been had she continued with it, not just because she has good reasons to do so, but because she acts for these reasons. She asks us to compare Emily with Alice, a fellow student, who has the same good reasons to quit as does Emily, but whose restlessness is the result of a deep lack of self-esteem, which would have caused her to feel exactly the same whether or not the good reasons for quitting were present. When Alice quits she is not acting for the good reasons that Emily is, but because of low self-esteem, which is not a good reason to leave the course: it is Alice’s operative reason, but it is not a normative reason (or at least not one with enough normative force) so does not justify her action. Arpaly argues that,
Alice is twice irrational – first, like Emily, for holding a false conviction in the face of evidence, and second for leaving the program due to psychologically powerful bad reasons. Emily, on the other hand, leaves the program for good reasons, and her only irrationality is in her failure to give up her errant conviction. (2000, p. 50)

Emily’s original judgement is not rational – it is not in accordance with her reasons – but her action of leaving the programme, is, maintains Arpaly, because it is what she ought to do. Arpaly thus appears to argue that these cases of, as Aristotle puts it, ‘foolishness combined with incontinence’ can be at least more rational than the same foolishness combined with continence. Emily is, Arpaly argues, more rational for her inversely incontinent action of leaving the programme than she would have been displaying continence and remaining in the programme; being incontinent does not add any irrationality to Emily’s action, over and above that which was involved in the formation of her poor atcj.

In order for Emily’s action to count as being rational, it is vital that she did not accidentally do what she ought to have done, but that she acted for her reasons; that her action was an expression of her antonymous agency. Clearly, since Emily is not consciously aware of these good reasons, Arpaly’s position is that one can act for reasons of which one is not consciously aware. If she is right – and I think she is in this regard – we can be practically rational by routes other than that of acting upon the best judgements produced as a result of practical reasoning. However, one can accept this without accepting the Coherentist picture and claim that there is nothing special about atc judgements. One can even accept that Emily is more rational for leaving the course than she would have been if she’d stayed, without thereby committing oneself to the claim that the practical judgements of practical reasoning have no special normative standing. And one can do this without having to claim, as does the Responsiveness to Reasoning

---

29 Aristotle warns against arguments that entail inversely weak-willed actions can be virtuous: “A certain argument, then, concludes that foolishness combined with incontinence is virtue. For incontinence makes someone act contrary to what he supposes [is right]; but since he supposes that good things are bad and that it is wrong to do them, he will do the good actions, not the bad” (1925, pp. 176; §1146a28-31). In Chapter 7 I argue that we should rightly be suspicious of such, but that inverse akatic action – that which is against one’s atcj but in line with one’s intentions – can be both rational and virtuous.
picture, that atc judgements have special normative standing just in virtue of the sorts of judgements that they are.

This is done by separating the claim that our conception of persons as rational agents includes that they guide their actions via their reasons, from the claim that they need to do this by recognizing their reasons as a result of conscious deliberation, that is by adopting what I call the Responsiveness to Reasons picture. The Responsiveness to Reasons picture allows that conscious guidance is not necessary for rational agency – thus embracing the benefits of the Coherentist picture – without abandoning the conception of practical agency – this avoiding the Coherentist’s problems. The Responsiveness to Reasons picture can allow a special normative role for atcjs, because it can recognise that guiding our actions via our best reasons by complying with our atcj is an expression of our practical agency. It can embrace the psychological naturalistic position that explains intentional human behaviour as basically continuous with the behaviour of non-rational animals, but retain recognition of our epistemic and practical agency. On the Responsiveness to Reasons picture, an action does not have to be the direct result of conscious deliberation in order to count as being under self-governing agency, but it must have been made with the active capacity to have been so.

If, as maintains the Responsiveness to Reasoning picture, the only way in which a person can exhibit practical rationality is through acting on consciously-formed atcjs as the result of, at least minimal, deliberation, then Barbara, who no longer needs to consciously deliberate about whether or not she should run away when she experiences vertigo, would not deserve rational praise for continuing with her hike. If this position were modified to allow Barbara’s action to be rational on the basis that stemmed from her previous judgement, we would be left with the strange outcome where Barbara was rational, but her husband Tony, who never had to consciously deliberate about whether or not he should run away when he experienced fear when walking on the cliffs, would not be due rational praise for he would not be subject to rational criticism at all. Why should Barbara’s action have a rational standing over and above that of Tony’s just because she at one
point had to consciously reflect on it? This seems contrary to the intuition that Tony might actually be the *more* rational for never having to ‘make’ himself continue in the face of fear, and would also be contrary to the intuitively plausible Aristotelian view of the superiority of the virtuous man over the continent, as the virtuous man does not even *need* to deliberate about what he should do. If the position was that Tony’s well-functioning reason-tracking mechanisms moving him to continue hiking in the face of fear was *arational*, then that would seem to need to extend to Barbara’s now better-functioning reason-tracking mechanisms moving her to continue, and then we’re back to the odd position of what was once a rational action no longer being subject to rational criticism.

But the picture at the other end of the spectrum, the Coherentist picture, leaves us with equally strange outcomes. It leaves us with *no* special normative standing for action as the result of practical deliberation. For example, Arpaly’s position is that Emily is more rational for leaving her course, but less rational than she would have been had she formed the judgement that she should leave her course based upon her good reasons, but this lessening of rationality is not due to Emily’s acting against her judgement, but is just because her beliefs are incoherent, with there being nothing special about the fact that it is her *atcj* with which they are incoherent. An atcj, on the Coherentist view, has no special normative standing, thus it’s not because it conflicts with *this* belief in particular that Emily is not fully rational. We are therefore left with there being nothing particularly irrational in acting incontinently, and nothing particularly rational in acting contently, or virtuously. Neither of these extremes, that of the Responsiveness to Reasoning Picture or the Coherentist picture, can encompass all of, and only, the times when we subject an agent’s actions to rational criticism. We rationally criticise the behaviours of only those who are capable of practical reasoning, but not only when they have engaged in practical reasoning. We rationally criticise those who *ought* to have engaged in practical reasoning and who fail to do as they ought to have done *because* they failed to engage in practical reasoning.
For example, had Barbara not examined her feeling of wanting to turn back on her hike, but had just ‘gone with whatever she was feeling’ and turned back, we would criticise her for this. This is because she should have examined the deliverances of her reason-tracking mechanisms, to decide whether they were really reflecting her reasons or not. System 2 should have been activated because it should have detected an error in the deliverances of her System 1 reason-tracking mechanisms. Barbara’s behaviour is irrational because she ought to have realized that she did not have good enough reason to turn back. Thus practical agents need not just the ability to engage in practical reasoning, but they need to be sensitive to when they ought to engage in it. If an agent does not need to consciously deliberate about whether or not to do what they are doing their non-deliberative actions can still be rational if they are such that they would deliberate if they needed to. Thus Barbara is irrational for failing to deliberate when she ought to have done, but Tony’s action of continuing with his hike is still capable of being a candidate for rational action because he didn’t need to deliberate. His action will count as rational if and only if he would have deliberated had he ought to have done. Tony’s self-monitoring capacities produce no error signal, since he is already acting in accordance with his reasons, but were he to think of running away, the error message would be produced, picked up by System 2, and his behaviour correspondingly affected, by engaging in practical deliberation about what he ought to do.

If Barbara has got to the point where she behaves just as Tony does, and does not need to engage in practical deliberation about whether or not she ought to run away, this is because as a result of her self-monitoring and behaviour modulation capacities, she has recalibrated her reason-tracking mechanism of fear. I have adopted the term ‘recalibration’ from Jones (2003). Jones’ account of practical agency maintains that because we reflectively accept our ‘sub-systems’ as generally-reliable reason-tracking mechanisms, the things tracked by these mechanisms qualify as putative reasons for action, but that because they are only putative agents must monitor their reason-tracking mechanisms, “stepping in where necessary to discount those mechanisms, and where possible, to recalibrate them into reason-tracking mechanisms through habituation” (2003, p. 196). Just as I argue that Tony’s action is rational iff he would have deliberated had he ought to have done because System 2 would have been activated in response to the error signal from System 1, Jones argues that rational agent, “would not rely on that first order sub-system were it reasonable for her to believe that it failed to reason-track” (ibid., p.195). That is, if there were good grounds.
to deliver different representations of the reasons that pertain. It may have become a much weaker response, or one that is triggered only by better reasons than it was before (perhaps only by sheer drops rather than craggy slopes), or the very beginning feelings of fear trigger the error message, so that behaviour is readjusted before the fear response can dominate her perspective in the way it used to. She has adjusted her habitual responses so that they better track her reasons, improving her conformity to the norm of practical rationality. Because she has recalibrated her response, she has guided her behaviour in line with the regulative norm of doing what she has most reason to do. This is why her actions then count as rational, even though they are not the direct result of practical reasoning. Tony’s action is rational because he was already guiding his responses in line with the norm of doing what he has most reason to do, for had he grounds to suspect that there was a deviation from the norm, he would have recalibrated his reason-tracking mechanisms to better comply with it. Thus habitual actions can be under an agent’s regulative self-governance, and count as practically rational expressions of her rational agency.

‘Fast’ action – action that occurs before an agent can deliberate – can also express the commitment to rational guidance and thus count as rational, if it is the result of rational guidance. Consider Arpaly’s example of a tennis player:

It is not a provocative view that an accomplished tennis player …does not have time to deliberate on all her moves during a fast-paced game. Not only that, but given the complex factors to which she responds, she is unlikely to be able to reconstruct all her reasons for action after the game. However, even after the ball is served, we can legitimately judge her moves as rational (“That was brilliant!”) or criticize her for irrationality (“What on earth were you thinking there?”). If actions not deliberated upon
to doubt that his fear was justified that were ‘detected’ by another of his System 1 reason-tracking mechanisms, he would not rely on his fear as providing the accurate representation of what action the reasons of his circumstances called for. He would have engaged in deliberation about whether or not he really ought to turn back. 31 System 2, of which conscious deliberation is a process, is not activated as quickly as is System 1, and so often we act as the result of System 1 processes before we can deliberate. As I explain in Chapter 5, this is thought to be an evolutionarily adaptive feature of System 1, and in fact is an aid to practical rationality as it enables us to respond to reasons quickly enough in situations where, were we to wait until System 2 could respond, we would no longer be able to respond appropriately. For example, the fear response enables organisms to run away immediately the reason to run away is perceived, which means they are more likely to manage to escape. Were they to wait around until they had deliberated about what to do, the threat to their life might have been already proved real.
are irrational by definition, it is hard to see how we can praise the player for the brilliance (rationality) of her action, and it is also hard to see how we can criticize her for a “crazy” (irrational) action, for we do not expect a person to be able to deliberate in the middle of a face-paced game. (Goldie, 2008, pp. 51-2)

One might object that the tennis player’s action is not rational or irrational at all; it is simply arational: that we do criticise tennis in rational terms doesn’t mean that we mean to, strictly speaking: after all, we criticise the behaviour of our pets and children with similar language (‘Oh Puss; why are you miaowing at the fridge? You know your food is not kept in there. Oh clever boy, that's right: it’s in the cupboard!’), yet we do not really subject their behaviour to rational criticism. But the tennis player’s actions are significantly different to this. She has trained her responses to comply with what she would do were she to deliberate about what she ought to do. She has calibrated them to better respond to her reasons, thus they are guided by the regulative norm of doing what one has most reason to do. Of course we do perform some fast or habitual actions which we can call arational – for example humming tunelessly, throwing an "uncooperative" tin opener on the ground or out of the window, jumping up and down in excitement (Hursthouse, 1991, p. 58) – but that does not mean that all fast or habitual actions are arational. Joseph Raz calls actions such as humming "expressive actions”, and argues that although they are intentional actions, they are not actions that are “done for reasons” (1999, p. 39). So long as actions are done for reasons, they are candidates for rational criticism. Some fast and habitual actions – such as Barbara’s habituation to continue in the face of her fear, Tony’s disposition to carry on in the face of his, and Arpaly’s tennis player’s fast responses in her game – are done for reasons because they are responding to reasons, and so are subject to rational criticism. Our reason-tracking mechanisms count as being responsive to reasons when they are under our regulative guidance. When our behaviour is under our regulative guidance it is being guided by the norm of doing what we have most accessible reason to do, and thus it counts as being the action of a practical agent and so as a candidate for rational action. Behaviour that is not under the agent’s regulative guidance will be arational, but that which is under such guidance will be practically rational to the extent that it is guided by the norm of practical agency.
5 Reason-tracking mechanisms must be under the agent’s regulative rational control: this is in what rational agency consists

The problem with the Coherentist picture is that it does not recognize our practical agency. As Jones puts it, to think of oneself as well-functioning subsystems that passively registers and respond to one’s environment is to give up thinking of oneself as guiding one’s actions via one’s reasons, for in guiding our actions via reasons we do not merely register reasons but understand them as justifying our actions (2003, pp. 188-189). Thus for our behaviours to count as being those of our agency, they must be under our guidance. Those behaviours that are not under our guidance are simply not candidates for rational action. Thus for the behaviour that is under the control of System 1 processing to count as being practically rational actions, those reason-tracking mechanisms responsible for them must be under the agent’s regulative control. To the extent that such actions are being guided by the norm of doing what one has most accessible reason to do, they will be practically rational. Since the way in which an agent guides her actions in line with the norm of doing what she has most accessible reason to do is by responding to her reasons, an action will be rational to the extent that it is responding to an agent’s accessible reasons for action. The agent who Φs for what are good reasons, yet, because she is limited, fails to respond to all of her reasons, which in fact entail that she ought to not-Φ, is practically rational to the extent that the reasons for which she does act actually justify what she does. The agent who, as a matter of coincidence, Ψs, which in fact is what she ought to do, but who does so for reasons that do not actually justify Ψ-ing, is less practically rational because, despite having done that which she ought to have done, the reasons for which she acted did not justify her action.

What does it mean for a reason-tracking mechanism to be under the agent’s regulative rational control? It means that the agent isn’t passively responding to her environment, but is actively responding to her environment. As Rowland Stout argues, “[p]art of what it is to be an agent with free will is to be capable of
determining what one should do according to the reasons one has” (Raz, 1999, p. 11), thus for an action to be an expression of rational agency it needs to be that the agent has determined that she should do what she does. This does not mean, as I have already argued, that she must have consciously decided to do what she does as the result of having engaged in practical reasoning, but it does entail that she is an active agent in the action. What does it mean for her to be active? It is for her capacity for practical reasoning to be sensitive to her reasons; that is, for it to be such that was System 2 processing – conscious deliberation – needed, it would be engaged. It is for her capacity for practical reasoning to be sensitive to those reasons which ought to activate it. As Raz argues,

We are active when our mental life displays sensitivity to reasons, and we are passive when such mental events occur in a way which is not sensitive to reasons (Raz, 1999, p. 11).

The agent’s mental life is sensitive to reasons even when not engaging in practical reasoning if it would involve conscious reflection if a reason was detected about which conscious reflection was required. Thus we are active as agents even when our reasons are not the contents of our mental lives if they would be if such was needed. A reason-tracking mechanism is thus under the agent’s regulative guidance if it’s deliverances would be subjected to conscious rational reflection were they to not be in accordance with what the agent has most accessible reason to do. If deviances from the norm would be detected and trigger System 2 processing, the agent’s behaviour is being guided by the norm, and thus is under her rational guidance control. So long as this is the case we are responding to our reasons.

That an action is under the agent’s regulative guidance control does not ensure that it is going to be a rational action. Actions are rational to the extent that they are a response to the agent’s accessible reasons for action, and just because a deviance from the norm would trigger practical reasoning does not mean that the resulting action will thereby be whatever it is that the agent ought to do. The agent might ignore the trigger, she might reason incorrectly, she might fail to respond to all of her reasons or their actual normative forces. There are numerous ways in which an agent’s actions might fall short of being that which she has most
accessible normative reason to do. She might not realize, as is often the case, that what she did was irrational until after she’s done it, or at least she might remain unsure as to whether she did or did not do the right thing. In such cases the irrational action might itself prompt recalibration, and thus contribute towards future rationality. This would be, as Jones puts it, an expression of an agent’s ongoing commitment to rational guidance. Guidance via reasons commits the agent to the on-going cultivation and improvement of her reason-tracking mechanisms so as to ensure that she is as responsive to her reasons as possible. Thus if an action is an expression of this on-going commitment to rational guidance it is a credit to the agent’s practical agency.

For example, Arpaly argues that Emily is not rational just by virtue of having accidentally done what it is that she ought to have done (remember that Alice, who acted for bad reasons although she had the same good reasons as Emily, did not deserve rational praise on Arpaly’s account), but at most I think we can say that Emily’s action is rational only to the extent of being less irrational than her continuing on her course would have been, for Emily’s leaving the course showed little evidence at the time of expressing her practical agency, though once she reflected upon the reasons for which she did so later in life she could recalibrate her behaviour as part of her commitment to her ongoing cultivation and improvement of her reason-tracking mechanisms.

To see this, imagine Jenny, who deliberates that she should consume three whole chocolate cakes, one after another (perhaps Jenny is going to be in a cake-eating contest, and she wishes to train). Jenny starts eating the cakes and then finds herself unable to follow through with her intention because her stomach is so distended that she does not feel able to swallow another bite. Jenny’s incorrect assessment that she should eat three whole chocolate cakes, in being incorrect, prevented her from being able to complete the task; it became impossible after a point. In the same way, if it was because Emily’s intention to override her emotional discomfort was faulty that it became impossible for her to do so after a point, it to Emily’s credit that she failed to follow through with her intention only so much as it is to Jenny’s credit that she failed to follow through with her
intention. Both Emily's and Jenny's failure to act on their intentions entailed that they did what they in fact ought to do, but their failures were not an expression of their rational agency. We credit agents with forming intentions and following through with those intentions, and although, as I have argued, we can credit agents who perform fast and habitual actions without having formed conscious intentions, the reason we praise agents such as Arpaly's tennis player is that she has trained her impulses and she therefore does not need to deliberate: her reason-tracking mechanisms are acting as they should be, because of how she has trained them, re-calibrating them as and when necessary. On Arplay's story, Emily has not trained herself to not override her emotional impulses (in the way that, perhaps, soldiers have to train themselves to not respond to certain reasons in order to be able to kill). The difference between Emily and the tennis player is that the tennis player has actively trained herself in the right way, whereas Emily has failed to train herself in the wrong way. She has failed to take herself as far away as she tried to from her standard rationality. This means that she is not as irrational as she might have been had she succeeded, but she is still less rational than she should have been. Thank goodness, we might say, that she was unable to be as irrational as she intended!

The tennis player, on the other hand, is better than standard: she has actively trained her impulses in the right way. It may have been a different story: she may have been worse than standard. We can imagine that she may erroneously deliberates that using a single-handed backhand would enhance her game, and then habituates herself to do so until, in the middle of one match she impulsively reaches for a shot with a double-handed grip, which is actually what she ought to do, and it leads to further action that she ought to do, as she then continues to use the double-handed grip. That first double-handed shot may have been what she ought to have, but it marked the end of walking down the path of irrationality, rather than displaying supreme rationality. Our reaction to this turn around, had we been watching the game, and following this player throughout this irrational phase, is not to say “Bravo!”, as we may have done when an already excellent player takes a shot so genius that we are amazed at their skill, but to say “Thank goodness! Now don’t go down that path again!” Coming to one’s senses is more
of a relief than an achievement. Thus although our tennis player and Emily eventually performed the right actions, they do not automatically attract praise for having done so. A fast or habitual action can be rational if and only if it is the result of an agent’s on-going commitment to rational guidance. Barbara’s continuing on her hike in the face of her fear many years after she last had to deliberated about whether to do so, being the result of her conscious efforts at re-calibration, is an expression of her autonomous agency and so is rational. And her husband Tony’s continuing in the face of feelings of fear having never had to consciously deliberate about it is also a rational action if he would not have relied upon the deliverances of his fear response had he reason to mistrust it.

An agent is practically rational to the extent that her action is guided by the regulative norm of doing what she has most accessible reason to do. This guidance requires that the agent’s capacity for practical reasoning is active, and the proper functioning of her agency consists, as Raz puts it, “in proper responsiveness to reasons” (1999, p. 19), such that practical reasoning will be engaged when it should be. Thus it is part of responding correctly to reasons that the agent will engage in practical reasoning when it is needed. In the next chapter I argue for a picture of practical reasoning that allows a substantial part to be played by System 1 reason-tracking mechanisms such as emotion.
Chapter Three
Practical Reasoning

In Chapter 1 I argued that practical rationality consists in doing what one has most reason to do, and that the normative force of one’s reasons is derived from the fact that they promote something one values, such that $R$ is a normative reason to $\Phi$ by virtue of the fact that it will promote state of affairs $A$ from which it derives its normative force. In Chapter 2 I argued that an agent is practically rational to the extent that she is guided by the norm of doing what she has most accessible reason to do, and that her action is so guided to the extent that it is responsive to her normative reasons for action. The capacity for practical reasoning, I argued, is what distinguishes practical agents from creatures who merely exhibit behaviour, and that although actions do not have to be the direct result of practical reasoning in order to be rational, they do need to be under the agent’s active regulative guidance control, which is manifested by the fact that she would engage in practical reasoning about the reasons for which she is acting if she needed to. Thus practical reasoning is at the core of practical rationality, but as a regulative capacity that is active even when not occurrent. In this Chapter I will argue for a particular picture of this core component of practical reasoning. The picture for which I argue does not posit the existence of a rational ‘self’ that is the ‘head’ separate from the ‘heart’. This picture of practical reasoning makes room for substantial roles for emotion, and thus for the argument that emotion is integral to human practical rationality, contrary to the traditional view of emotions as being irrational and disruptive influences to reason, for which I argue in Chapter 6.

The main claim I make about practical reasoning is that it is not something that is ‘better off’ without emotion, because it is not a process of trying to be free from our subjective desires and concerns. We do not ‘step back’ to position from which our reasons for action are viewed from an objective, third-person perspective such as that from which we judge the actions of characters in novels. We are always in the first-person position, such that the decision we make can only ever
be that of the subjective ought: what we have most accessible reason to do. Thus, to repeat an example from Chapter 2, although it is true that Dorothea from Middlemarch objectively ought not to marry Casaubon, Dorothea cannot engage in practical reasoning in order to determine this. All she can do is to deliberate about what she subjectively ought to do, for she can deliberate, respond to and act for only those reasons to which she has access. But, as I shall argue in this chapter, she cannot take an objective stance towards even these reasons, for not only do an agent’s values determine what reasons she has for action, they also determine how she deliberates about those reasons. They are employed in the reasoning process, and have to be so employed in order for us to reach decisions.

Thus, when we engage in practical reasoning, we do so not from the position of, what I will call, the ‘Kantian Captain’, a position that is ideally devoid of desire and emotion, independent of subjective value and thus from which we can make an objective, value-independent assessment of our reasons in order to discover the objective fact of what action the reasons for action we have to which we have access best support. However, the picture at the other end of the scale, the traditional Humean position of practical reasoning, is inadequate to the phenomenology of deliberation as it presents the agent as passive, being ‘blown around’ by the various strengths of her passions. I thus argue for a picture of practical deliberation that allows reason its proper place in practical thought, but does not posit the existence of a separate rational self that aims to reach an unemotional decision that reflects some independent standard of reason. This leaves open the possibility of accommodating the substantial body of evidence that emotion is integral to practical reason and therefore practical rationality, for which I shall argue in Part 2 of this thesis.

In this chapter I first argue that practical reasoning is conducted from what I will call the ‘deliberative stance’, which is distinguished from the ‘theoretical stance’. From the deliberative stance the agent is not making a theoretical enquiry but a deliberative enquiry, in that she is not trying to discover the pre-determined fact of the subjective ought, but is asking the normative question of what she ought to do. In section 2 I argue that practical reasoning has a distinct phenomenology of
being under our control, and that we need to acknowledge our sense that we ‘step-back’ to a position from which we reflect on our reasons for action and can decide whether or not we endorse the deliverances of our reason-tracking mechanisms and decide whether or not we will act upon our desires. We are not being ‘blown around’ by our passions, and, as I argue in section 3, we are deliberating about our reasons for action, not about our desires. In section 4 I argue because we cannot make an impartial assessment of our reasons, the deliberative stance is not that of the impartial ‘Kantian Captain’, and that there is therefore no separate rational self that conducts practical reasoning ‘upon’ our reasons.

1 Practical reasoning is conducted from the deliberative stance

Practical reasoning is the process of deliberating about what one ought to do, and concludes in the agent forming an intention to act. Practical reasoning differs from theoretical reasoning in that the theoretical question is answered by the discovery of a fact which already existed, whereas the practical question is answered, not by the discovery of an intention which already existed, but by the formation of an intention, such as ‘I will go to bed’; ‘We shall go to war’. This decision is not something that already exists and that the agent ‘discovers’, but something that she makes, and that does not exist until she makes it. In answering both practical and theoretical questions the agents considers features of her situation as potential reasons, but in the theoretical ‘stance’ towards such she is considering their potential to count in favour of believing that something is

32 That the conclusion of practical deliberation is an intention to act, rather than actually the action, is contested (see, for example, Tenenbaum, 2007), but, as Joseph Raz argues, the position that the conclusion of practical reasoning is an intention, “allows for failure to act in the way entailed by the premises which is not a failure of reasoning but is due to inability, forgetfulness, weakness of will etc” (1978, p. 5): if the conclusion of practical reasoning were action, then any such failures would have to be explained as incorrect reasoning, a position which would be rather difficult to defend. Indeed, it is not even definite that Aristotle, to whom credit is given for the thesis that the conclusion is in action, actually held this view (Charles, 1984, pp. 57-108). See also John Broome’s argument that “making a decision is as close to acting as reasoning can properly get you” (1999, p. 407).
true, whereas in the deliberative stance towards such she is considering their potential to count in favour of her performing a certain action.\textsuperscript{33}

The deliberative stance can be taken towards normative questions other than those about bodily action. For example, both stances could be taken towards one's emotional experience. In the theoretical stance, the question “what do I feel” would be answered by the formation of a belief about what it is in fact that one feels, where what one is discovering is an antecedent fact. One might think, “I am aware that I am having an emotional response to this situation, but I am not sure what it is”, and theoretical enquiry will aim to discover what the pre-existing feeling is. By contrast, in the deliberative stance the question “what do I feel?” is not aimed at discovering an antecedent fact, for one does not exist: the matter of what one feels has not yet been settled. It is settled in the coming to the decision of what one ought to feel. In this deliberative stance, one might think, “I don’t know what to feel about this situation: what should I feel?” and deliberation will aim to come to a decision about what one ought to feel.\textsuperscript{34} One might apply both the theoretical and practical question in the same situation; first asking the theoretical question of what it is that one does feel, and then the practical question of whether or not one ought to feel that emotion, that is, whether or not the emotion is appropriate. The answer to one of the questions may affect the answer to the other, but still, they are two different questions.

Likewise, the question of “what shall I do” can be asked from both the deliberative stance and the theoretical stance. If the question is asked from the theoretical stance then the matter is one of prediction, based upon evidence of one’s own past behaviour. If the question is asked from the deliberative stance then the decision is based upon reasons for acting in one way or another, and is a matter of decision, of being “up to [one] to decide what [one] will do” (Hampshire & Hart, 33 I am here adopting Richard Moran’s (2001) terminology for describing the difference in the perspectives from which one engages in practical reasoning (the deliberative stance) and theoretical reasoning (the theoretical stance).

34 See also Moran (2001): the question ‘I don’t know what I want’ may either “express a divided consciousness containing certain definite though repressed desires, or it may express someone deliberating about what’s desirable” (Moran, 2001, p. 57). If the former, the agent is in the theoretical stance; if the latter, she is in the deliberative stance.
The theoretical question is an empirical one; the deliberative question is normative. And whereas the answer to the theoretical question is a matter of a pre-determined fact which we are attempting to discover, the answer to the deliberative question has the distinct phenomenology of being up to us, in that the decision we reach as to what we will do is under our control.

2 Practical reasoning has the distinct phenomenology of being under our control

Practical deliberation has the distinct phenomenology of being under our control. Our experience of deliberating is that we examine what appear to us to be our normative reasons for action and decide whether or not we endorse them as such, whether we accept their putative normative forces, of deciding which action we will choose to then take. In the language of the Dual Processing theories of cognitive science which I discussed in the previous chapter, the System 2 capacity of conscious deliberation involves reflection upon the deliverances of System 1 reason-tracking mechanisms, deciding whether or not to endorse the putative reasons and their normative forces as they are presented. $R$ initially seems like a reason to $\Phi$, as presented by a mechanism of System 1, but when System 2 is activated the agent can question whether $R$ really is a reason to $\Phi$, and whether, if so, it actually justifies her $\Phi$-ing in the light of her other reasons. Thus the phenomenology of practical reasoning is that we can decide whether or not something is a reason, that we can decide how much normative force it has as a reason, and that we can decide what action to take on the basis of our decisions as to our reasons and their normative forces. It seems, as Marcia Baron puts it, that we deliberate with “mastery” (2010, p. 35), such that our decision is made under our control, rather than at random or by a pre-determined causal chain. We think of ourselves as decision makers as being self-controlling; we are aware of the unruly forces of our desires but with care we can usually control them so that we have chosen whether or not to act on a desire rather than our behaviour having been controlled by them. This classic dualism of heart and head, desire and reason, has reason in control so long as thing are going well. Such a view can be traced back to Plato, who divides the “soul into three parts,
two being like steeds and the third like a charioteer” (Phaedrus, 253c-d), one steed representing our honourable inclinations, the other our dishonourable (thus the steeds are our desires), and the charioteer steering the whole soul in the ‘right’ direction (the charioteer being ‘reason’). It is the role of practical reasoning to steer us in the right direction. Reasoning is responsible for guiding our behaviour in line with what we ought to do, that is, in line with what we have most reason to do. Thus acting against the norm of doing what we have most reason to do is acting against reason, and reason would have failed to properly control our behaviour if this happens. In the previous chapter I endorsed these intuitions; the role of practical reasoning is to guide an agent’s behaviour in line with the regulative norm of doing what she has most accessible normative reason to do.

David Hume turned this model upside down, arguing that, “[t]is not contrary to reason to prefer the destruction of the whole world to the scratching of my little finger” (1888, p. 416). On the Humean picture of practical reasoning, the course is set by the steeds, not the charioteer: the charioteer doesn’t steer the horses, but is rather carried in whichever direction the stronger of the horses takes him. The charioteer’s job is reduced to instrumental reasoning, to working out the means to the ends, which are set by the passions. If the horses steer one in the wrong direction, then it is their fault, and not the charioteer’s, for reason “can never oppose passion in the direction of the will” (ibid p 413; italics added). It may be wrong to prefer the destruction of the whole world to the pricking of one’s little finger, but, on Hume’s account, it is not a failure of reason, but rather of passion. This Humean picture of reason being the slave to the passions appears to present us as being ‘blown around’ by our various passions, with the only deliberation available to us being that about how best to satisfy whatever it is we desire. Rather than reason being the Platonic charioteer, steering the course of action, it is the platonic ship master, being bullied by his various crew members, his desires:

*Picture a ship master in height and strength surpassing all others on the ship, but is slightly deaf and of similarly impaired vision, and whose knowledge of navigation is on a par with his sight and hearing. Conceive of the sailors to be wrangling with one another for control of the helm, each claiming to that it is his right to steer though he has never learned*
the art … they are always clustered about the shipmaster importuning him and stopping at nothing to induce him to turn over the helm to them. And sometimes, if they fail and others get his ear, they put the others to death or cast them out from the ship, and then, after binding and stupifying the worthy shipmaster with mandragora, or intoxication or otherwise, they take command of the ship, consume its stores and, drinking and feasting, make such a voyage of it as is to be expected from such … With such goings on aboard a ship do you not think that the real pilot would indeed be called … a useless fellow, by the sailors in ships managed after this fashion? (Republic, 488-489a)

It is not difficult to see the unattractiveness of seeing ourselves as being akin to this sort of ship. There are a multitude of voices, each passion being represented by a crew member, with the direction of the voyage being determined by the resolution of conflicting pressures among them, not by any decision of what one ought to do. On this picture if one crew member really wanted to do something, then he would ‘overwhelm’ the others and do it, and thus it looks as though whatever we end up doing will just be whatever we most feel like doing at the time. This is not true to the phenomenology of our deliberating with mastery.

The phenomenology of deliberation is that we are able to ‘step back’ to a position from which we can decide whether or not we ought to act on a desire; to decide if we endorse the putative reasons presented by our reason-tracking mechanisms, to decide how much normative force we think these reasons have, and to ‘weigh up’ our various reasons for action rather than just acting on whatever impulse we happen to have at the time. We regard our ability to do so as one the capacities we have as rational agents (Jones, 2003, p. 189), which separates us from non-rational creatures. As Christine Korsgaard puts it,

A lower animal’s attention is fixed on the world. Its perceptions are its beliefs and its desires are its will. It is engaged in conscious activities, but it is not conscious of them … But we human animals turn our attention on to our perceptions and desires themselves, on to our own mental activities, and we are conscious of them. That is why we can think about them.

And this sets us a problem no other animal has. It is the problem of the normative. For our capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them, and to call them into question … I desire, and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn’t dominate me and now I have a

There are two sorts of ‘stepping back’ conflated in this description. The first is stepping back from a desire to act (a realizer desire) to enquire as to whether one ought to have that desire. We might desire to do something for no good reason, for example out of habit, or for ‘misplaced’ reasons: a smoker might desire a cigarette because they are in a situation which triggers a desire for one, which doesn’t justify desiring one as it does not reflect anything choiceworthy in the situation; a person may desire to engage in a heated argument with a colleague because they are in a bad mood, caused by earlier news that the builders they have employed have once again made a hash of the job, which does not justify their desire to argue with their colleague (and is in this sense ‘misplaced’ attribution of justification).

The second sort of stepping back is to ask whether one ought to act upon that realizer desire: even if the desire is itself justified (perhaps the person wants to argue with their colleague because the colleague has repeatedly insulted them) it may not be the case that she ought to act upon that desire (for there may be many reasons against doing so, such as that it will cause problems in the continued working relationship, or with other colleagues, or her superiors). In both cases the agent steps back to the deliberative stance, for she is asking a normative question (‘ought I to want to do this?’, ‘ought I to do this?’), considering reasons that count in favour of what ought to be the case (whether it is appropriate to have the desire to act, or whether it is appropriate to act in accordance with the desire). The agent cannot actually, as Korsgaard has put it, ask whether the desire is a reason to act, for, as I will argue in the next section, desires are not reasons to act: they track reasons for action, but the only reasons which they can ‘provide’ are those which they track. It is this second sort of stepping back with which I am concerned as practical reasoning is deliberation about reasons for

35 Recall from Chapter 1: a ‘realizer desire’ is what an agent wants to do at a particular moment; ‘intrinsic desires’ are desires that certain states of affairs, those that we value, be the case.
action, not reasons for desire. The question posed is not ought we to want to do, but what ought we to do.\textsuperscript{36}

3 Practical reasoning is deliberation about reasons for action, not desire: the role of desires in practical deliberation is not as reasons but as tracking reasons.

In the first sort of stepping-back that I have just mentioned, the agent is considering features of her situation as reasons that count in favour of, or against, justification of her wanting to do whatever it is that she wants to do; as reasons that justify a desire. In the second sort she is considering them as reasons that count in favour of, or against, her actually doing whatever it is that she wants to do; as reasons that justify an action. Although we can deliberate about whether or not we ought to want to do something, when engaging in practical deliberation, we are deliberating not about whether or not we ought to want to do whatever it is we want to do, but about whether or not we ought to act on that desire; whether or not we ought to do that which we desire to do. Thus when engaged in practical reasoning we are examining reasons for action, not reasons for desire. If an agent were to be examining reasons for desire, then she would not be answering the question of what she ought to do, for although, as I argued in Chapter 1, realizer desires track reasons for action, it is not necessarily the case that the question of ‘ought I to desire to Φ?’ has the same answer as the question of ‘ought I to Φ?’ A desire can be completely justified without thereby justifying the corresponding action. The easiest example of this is cases where an agent has two justified desires, but can fulfil only one. To deny that this is possible would be to argue that the only desires that are justified are those that are to do whatever it is we ought to do. But desires are not like judgements in this effect: one can hold

\textsuperscript{36} There is a third way in which an agent might step-back from a desire, which is to ask why she has the desire. This is a question which is answered by consideration of empirical evidence, and thus is asked from the theoretical stance. Of course an agent might ask all three of these questions: in asking herself whether her desire for an argument is justified, she might consider why she has the desire. The answer to this question, if it is that she’s in a bad mood as a result of her builders, might lead her to answering the normative question of whether the desire is justified, in the negative. This realization might lead her to answering the normative question of whether she ought to act upon the desire in the negative, also.
desires that produce ‘conflict without contradiction’. The conflict between desires is not a conflict of rationality. The agent need not necessarily revise either of the desires in order to avoid rational contradiction, as she need do in the case of beliefs. This is because, as I argued in Chapter 1, all beliefs have the same goal, that of truth, thus reasons for belief are justified from one source. Desires, however, are grounded by reasons that come from values, and an agent has multiple values. At times these values will produce conflicting realizer desires, but this does not impact upon the justification of her having those desires.

For example, Kath is on her second holiday in New York, and has time to visit just one collection in the Metropolitan Museum of Art. Last time she visited she encountered the Impressionists for the first time, and has a desire to visit it again, for she knows she values depth of experience: she will have the opportunity to appreciate the paintings in a different way, now that they are no longer novel. However, she wonders if she had rather visit a different gallery, and forms a desire to visit the Renaissance collection, for it has been recommended to her and she values new experiences, especially in terms of broadening her exposure to art. She has competing desires: on the one hand to go to the Impressionists collection, on the other to go to the Renaissance collection. There is nothing irrational about her having either of these, as they are both justified by perfectly good reasons. One might want to argue that once she has made her decision she ought not to continue to have the desire that goes against the choice, but this does not translate into the desire being irrational at the time she first had it.

I'm borrowing this phrase from Sabine Döring (2008), who employs it in relation to cases where an agent's emotion and judgement conflict without contradiction, in that the emotion need not, rationally, be revised in the light of her better judgement. Like emotions, desires can conflict with judgement without contradiction (one can judge that one would be better off without the desire), and can also conflict with other desires (including second-order desires, desiring to not have the desire (Schiffer, 1976)), without rational contradiction, though if one acted upon a desire one judged (correctly) one ought not to have, or desired (justifiably) not to have, then there would be an element of irrationality in the action. This irrationality, however, would come from the fact that one is not doing what one has most reason to do, that is, from the fact that the reasons that justified the conflicting judgement or second-order desire outweighed the reasons that justified the desire upon which one acted. The irrationality does not stem from the mere fact that one had a competing judgement or desire.

In this case, she would have an additional reason, one that counted against having the desire, and might tip the balance in favour of the desire not being justified. This reason could not count against the justification of the desire initially, because it did not exist at that time: the additional reason is that she has decided against acting upon that desire. This may count against having the desire because continuing to hold it may interfere with her enjoyment of the Impressionists,
Kath has most reason to spend her time in the Impressionist galleries and does so, that does not mean that her desire to visit the Renaissance galleries was something she ought not to have had in the first place. She can’t do both things, so she must choose between them and do whatever she has most reason to do, but she can desire both things, and thus is not required to ‘choose’ which one to desire on the basis of which she has most reason to. Thus it is not automatically the case that the answer to the question, ‘ought I to desire this’ answers the question, ‘ought I to do this’.

In practical deliberation an agent is considering her reasons for action. There are, as I have already said, cases in which an agent enquires as to whether she ought to desire to do something (either whether she ought to have this ‘realizer desire’ - she shouldn’t want to argue with her colleague, there is no reason to want to do so; what she ought to want is to confront her lazy builders – or whether she ought to have an intrinsic desire – she shouldn’t desire that she always get her own way, for it conflicts with values which are more important), but these are not the usual case. In the usual case, the role for realizer desires is to prompt deliberation; as Korsgaard put it,

… I desire, and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn’t dominate me and now I have a problem. Shall I act? (1996, p. 93)

The presence of the desire to act prompts deliberation about whether one should act, for,

The reflective mind cannot settle for perception and desire, not just as such. It needs a reason. Otherwise, as least as long as it reflects, it cannot commit itself or go forward (ibid)

The reflective mind – the agent engaged in practical reasoning about what she ought to do – needs a normative reason to justify the acting on the desire, for she is engaged in the question of whether she ought to so act; she is looking for

or might prompt reconsideration of the decision, which would be inefficient: as Scanlon (1998) puts it, she has a second-order reason to not reconsider, as a general policy of reconsidering one’s decisions without good reason (such as realizing one has overlooked something) would be “a serious impediment to doing anything at all” (pp. 240-1).
normative justification that makes it the case that this is what she *ought* to do. Thus, the first role for desires in practical reasoning is to *prompt* deliberation. In the language of Dual Processing theories, System 1 has generated a desire to Φ, but has also activated System 2 to answer the question of whether she *ought* to Φ. The desire to Φ somehow conflicts with the deliverances of other of her reason-tracking mechanisms, such that Φ-ing would conflict with reasons that derive their normative force from other of her values, thus an ‘error signal’ is generated and detected by System 2, and conscious deliberation is engaged to decide whether on the balance of *all* her reasons, not just those which justify her desire to Φ, she ought to Φ. In Chapter 6 I argue that emotions play a fundamental role in this signalling process, prompting deliberation. Desires (and emotions) do not just lead to activation of practical reasoning as realizer desires, but they also, as standing desires, track the reasons that conflict with the realizer desire and thus give rise to the need for practical reasoning.

As I argued in Chapter 1, standing desires lie in the back of our minds and occasionally generate thoughts, feelings and actions of the relevant sorts. These standing desires track our various normative reasons for action. Normative reasons for action derive their normative force from the values by virtue of which they count in favour of performing a certain action, thus if an agent values state of affairs A she has an intrinsic desire that A be the case, this gives her reason to desire that A be brought about or maintained, which gives her reason to desire as to act so as to bring it about that A obtains, which transmits its justification to the realizer desire to perform action Φ that will promote A. Thus as well as having the realizer desire to Φ, the agent will have a number of standing desires that derive their justification from some of her other values. And these standing desires will generate thoughts and feelings of the relevant sorts as they track the reasons for action that are relevant to them, and will thereby ‘pick out’ reasons other than those tracked by her occurrent desire to Φ.

The agent’s desires pick out the normative reasons that are relevant to the question of whether or not she ought to Φ, and they *also* affect how much normative force she considers her various reasons to have. She sees her various
reasons for action from the perspective of her various desires; they are not themselves usually the thing about which she is deliberating; rather than looking at her desires in practical reasoning, the agent is looking at her reasons for action through her desires, with her intrinsic and standing desires disposing her to see certain considerations as reasons, and to afford them certain normative weight. To borrow Simon Blackburn’s lively example;

If I am a miser, the cost takes my attention; if I am a gourmet, the quality of the food does; if I am prudent, the durability of the cloth; if I am not a knave, the fact of the promise. If I am extravagant, or a glutton, or concerned only with my appearance today, or if I am a knave, none of these features presents itself as important. There is not typically a second-order process of standing back, noticing that the cost is obsessing me, and deciding to endorse that fact about myself, or alternatively deciding to try to change it ... Deliberation is an active engagement with the world, not a process of introspecting our own consciousness of it (1998, p. 254).

To be a miser is to value spending as little money as possible, thus to have an intrinsic desire to spend as little money as possible, and it disposes one to see cost as a reason to take into consideration when deciding what to do, and towards giving such reason more normative ‘weight’ than would a non-miser. Being a miser means that the agent, in deciding whether or not she ought to buy the item that she is considering, let’s say it is a coat, takes its high price to be a reason against doing so, and a reason that has substantial normative weight. For the agent who has the opposite disposition towards extravagance, the expensive price is a reason for buying it, and might even be a reason with considerable weight (if, for example, she wishes to be seen as extravagant by others). Thus, as I argued in chapter 2, the same fact can be both a reason for and against the same action: the expense of the coat is a reason which the miser has against buying it, and a reason which the extravagant has for buying it. Both the miser and the extravagant have a desire to buy the coat, and they also both have the reason that it is expensive, but for the miser this reason that she has against acting upon her desire, whereas for the extravagant it is a reasons that she has in favour of action upon her desire.
The cost has taken the miser’s attention, and prompted her to engage in practical reasoning about whether or not she ought to act upon her initial desire to buy the coat. Her disposition towards saving money has tracked the cost of the coat as a reason against acting on the desire, thus there is a conflict between the deliverances of System 1 that presented the buying of the coat to be justified (perhaps she is disposed to want clothes in her favourite colour of imperial purple) and the reason-tracking disposition of miserliness. This conflict prompts practical reasoning to rule over what she ought to do, and she ‘steps back’ from the impulse to act on her realizer desire to buy the coat to deliberate. But when the miser steps back from her desire to buy the coat, she steps back not to a position where she is free from all desire, but to a position from where she sees the features of her situation as reasons from the wider perspective of standing desires. She does not step back to a position from which she can use cold, calculated reason devoid of all the influence of desire, but to the position from which she can employ her standing desires in order to work out what she ought to do. Thus, as I shall now argue, as well as prompting practical reasoning, desires play a fundamental role within it, and thus should not be seen as opposing reason.

4 The fundamental role of desires in practical reasoning

An agent has to employ her desires in practical reasoning because her desires track both her reasons and their normative force, and thus inform her as to what she has most reason to do. The fundamental role for her desires in her practical reasoning is in enabling her to reach a decision about what she ought to do. My argument is not that ideally an agent’s decision would be independent of whatever desires she happened to have; my argument is that, as a matter of contingent fact, the only way in which human beings can answer the question of what they ought to do is from a position which is, far from being independent of desire, the position from which they can ask the question from the perspective of desires. Not from the perspective of their realizer desire – from this they ‘step back’ - but from the perspective of their standing desires and, ideally, as far back as their intrinsic desires. This is because the judgements a practically deliberating
agent has to make – deciding whether something really is a reason that counts in favour of a particular action, how much normative force that reason has, what other reasons she has and how one consideration weighs up against these, what action is most justified by the combined forces of her reasons – all require the employment of her various desires, as I shall now argue.

I have already argued that what an agent ought to do is dependent upon her values, and that her desires track the reasons for action which she has by virtue of the fact that they promote these values. I am now arguing that because her desires track her reasons, she needs to employ her desires in practical deliberation in order to work out what she has most reason to do. This means that reason and desire are not at loggerheads in practical reasoning: the role of reason is not to ‘control’ unruly desires, nor is it to be their slave. The role of reason is to reflect upon the deliverances of desires, employing them to deliver putative reasons and putative normative forces, so that the agent can come to a decision about what she has most reason to do. My picture of practical reasoning goes against both the dominant positions, which portray desire and emotion as contrary to reason. The Kantian position has it that the role of reason is to ‘control’ emotion, which would suggest that the deliberative stance is one that is, ideally, devoid of emotions, since, as Kant sometimes goes so far as to say, “it must be the universal wish of every rational being to be wholly free from them” (Kant, 1964, p. 428). The Humean position has reason as “the slave of the passions” (Hume, [1739/40] 1978, p. 415). Both propose a dualism between reason and emotion. The position for which I argue does not, and leaves room for my argument in Chapter 6 that emotion is part of reasoning, and thus rationality. There I will argue that reasoning cannot occur, in humans at least, without emotion, but this does not mean that reason is emotion’s slave, being passively blown around by whatever motivational states the agent happens to have. I have already rejected this Humean position as failing to do justice to the distinct phenomenology of practical reasoning as being under our control. The phenomenology seems to be better represented by the Kantian picture of practical deliberation, which acknowledges the motivational role of the passions in trying to steer us towards a particular course of action, but postulates a different
source of motivation which, far from being the slave to the passions, is an independent source of a fundamentally different kind of pressure: reason. However, the temptation of such a view is to see the deliberative stance as being something where reason ‘rules’ over desire without being subject to it, which I shall now argue we should not accept.

In section 2 I used the Platonic metaphor of the person as a ship to illustrate why the Humean picture of reason being bullied into fulfilling ends set by desire is so unattractive. Simon Blackburn uses a similar metaphor to illustrate the Kantian picture of practical reasoning:

[S]tanding above [the crew – the agent’s various desires], on the quarter-deck, there is another voice – a voice with ultimate authority and ultimate power. This is the Captain, the will, yourself as an embodiment of pure practical reason, detached from all desire. The Captain himself is free. But he always stands ready to stop things going wrong with the crew’s handling of the boat (1998, pp. 246, italics added).

This ‘Kantian Captain’ is ‘free’ because she is self-legislating, master of herself and slave to no other; she is “a will which is itself a supreme law giver [and therefore] cannot possibly as such depend on any interest” (Kant, 1964, p. 99 [§ 72]). The Captain is not part of the crew, but rather impartially surveys the crew. This picture is intuitively very attractive as it portrays reason as taking into account all of the deliverances of her crew of desires, but as not being influenced by their motivational forces: the Kantian Captain she does not do whatever it is that the strongest of her crew members tell her to do, thus she is not directed by the strength of her desires. She can choose whether or not to act on any of her desires, and is not being controlled by them: she – reason – is in control. This satisfies our intuition of deliberating with mastery, and is certainly the sort of picture that we find suggestions of in the philosophical literature. For example, Korsgaard describes our desires as the “passively confronted material upon which the active will operates” (1996, p. 24; italics added), and Thomas Nagel argues that when practical reasoning one is in the position of judging what one ought to do “in a way that does not merely flow from [one’s] desires and beliefs, but operates upon them – by an assessment that should enable anyone else also to see what is the right thing for you to do against that background” (1997, p.
110). But, for all its attractiveness, this picture does not accurately represent human practical reasoning, because, as I shall now argue, when an agent deliberates, her desires are *not* passive, and reason does not conduct some sort of objective assessment ‘upon’ them. Practical reasoning is not conducted from the position of the Kantian Captain, but this does not entail that instead reason is passion’s ‘slave’.

I have already argued that in practical reasoning an agent is not assessing her *desires*, as Nagel suggests in the quotation above, but is assessing her reasons for action. She may, of course, ask questions about her desires, for example why she desires to \( \Phi \), which will yield up the reasons for which she has the desire to then be assessed as *reasons for action*. Although she can step back from a particular desire to ask what reasons it ‘provides’, she is doing so in order to decide whether or not she ought to *act* on the desire; the reasons it ‘provides’ are those reasons for *action* which the desire is *tracking*. Already then, we can see the problem with the picture of practical reason as the Kantian Captain as it is not the *crew of desires* which the agent is examining, but the *reasons for action* which those desires track. And although an agent can step back from a particular desire to reflect upon the reasons for which she has it and their normative force, she cannot stand apart from the *entirety* of her dispositional desires and view all of her reasons from the impartial perspective of the Kantian Captain. Not only is this something of which we are not capable, it is not even something which we attempt, for we *need* to view our reasons through our various desires and concerns in order to be able to come to a decision about what we ought to do. This is because, not only do our desires pick out the reasons about which we deliberate, they also determine what we decide we have *most* reason to do. This does *not* mean that they will determine that we decide to do whatever we have most desire to do; it does not mean that we are the Platonic ship. We still *choose* our course of action, and we do so on the basis of *reasons*, not on the basis of desires. But without desires, we would not be able to *make* a choice between the courses of action available to us, because we have to make *decisions* about our reasons for which we need to employ our desires.
Desires pick out our reasons for action, and they also pick out the normative force of our reasons. Recall the miser from the previous section: her intrinsic desire to spend as little as possible not only meant that the coat's high price was a reason which she had against buying it, but also it meant that she gave this reason more importance than would someone less miserly. When we deliberate we decide how much normative force our various reasons for action have, and we do so on the basis of how important the values from which they derive their normative force are to us. The normative force of our reasons is transmitted via the desires they justify, thus our desires track not only what reasons we have, but what normative force these reasons have for us. A reason for action does not have an intrinsic weight that is independent of the agent who has the reason. Just as an agent’s values determine the reasons for action which she has, they also determine the normative force that reason has for her. The greater she values the state of affairs by virtue of which $R$ is a reason she has to $\Phi$, the greater normative force that reason will have for her. Just as the ‘having’ relation in which an agent stands to normative reasons is motivational, the having relation in which she stands to their normative force is motivational. The more she values the relevant state of affairs, the more motivation she will have to promote that state of affairs, and thus the more normative force the reasons provided by that value will have for her. Two agents can have all of the same reasons for and against a particular action, but that does not entail that they ought to make the same decision about that action, as would be the case if the normative force of reasons were intrinsic properties of reasons. And not only can two agents have all the same reasons yet reach opposing decisions, the same agent can have the same set of reasons at time1 as she has at time2 and reach a different decision at time1 than she does at time2. To repeat the example from the first chapter of deciding whether or not to go to yoga, I might decide to go on Monday but on Tuesday decide to not go. This does not entail that on Monday I did what I ought to do and on Tuesday did not (or vice versa), nor does it indicate that I must have had different reasons at on Monday and Tuesday; although it could be true that by going on Monday I then, on Tuesday, had an additional reason which counted against going to yoga (my muscles were sore or something like that), this is not necessarily the explanation if it is true that on both days I made the right decision; that on Monday I ought to have gone to yoga and on Tuesday I ought not to. The explanation
could be that, because my dispositions fluctuated, the strength of my standing desires through which I saw my reasons fluctuated, thus although I had exactly the same reasons to take into consideration, I assigned one or more of them different weights on Monday and Tuesday. For example, my standing desire to get a good night’s sleep picked out as a reason that counted in favour of going to yoga the fact that yoga aids sleep on both Monday and Tuesday. On both days I took this reason into account when deciding whether or not to go, but on Monday getting a good night’s sleep was more important to me than it was on Tuesday, thus on Monday this reason tipped the balance of my reasons in favour of going, whereas on Tuesday it did not, or on Tuesday some other consideration that was less important on Monday was then more important to me and this tipped the balance in favour of not going. Thus the reasons which I had did not change between Monday and Tuesday, but the weights I assigned to these reasons did, and this accounted for the difference between my decision on Monday and my decision on Tuesday.

That I made a different decision on Monday than I did on Tuesday does not of course entail that I was right to do so: it could be that on both days what I in fact had most accessible reason to do was to go to yoga, and that I was wrong in undervaluing getting a good night’s sleep on Tuesday. Thus it may be true that there is an objective fact about how much normative force each reason for action has, and that given a particular set of reasons, there is an objective fact as to what action they most support which is neutral as regards to the agent who has them. But if this is true, it is not consistent with the phenomenology of deliberation. As Richard Nozick expresses this intuition,

reasons do not come with previously given precisely specified weights; the decision process is not one of discovering such precise weights but of assigning them. At least, so it sometimes feels (1981, p. 294).

Our experience is of deciding how important each of our reasons is to, and thereby deciding what we have most reason to do. The phenomenology of deliberation is not that of performing an impartial cost-benefit weighing-up of reasons, but of deciding how much weight we will give each of our reasons. We step back from the initial desire to act and decide whether the reasons for which
we have the desire are reasons that justify our acting on that desire. To do that we need to have access to these reasons, but we also need to consider opposing reasons – why are we not already convinced that acting on the desire is what we ought to do? What reasons to we have to not act upon it? Clearly the question of what we have most reason to do is not answered by the working out of which side has the greater number of different reasons. We need to decide how much each reason is worth: how much it counts in favour of or against the considered action. In doing so we might completely discard certain of our reasons as being obviously not important enough as to make a difference to our decision. We might discover additional reasons as we consider those that we already have, for example by considering the consequences of our action we might realize we have more reasons that count in favour of it than were represented by the initial desire. We do this by considering the features of our situation from the perspective of a wider variety of our standing desires than that of the initial desire for action, and in doing so we assign our reasons various weights. These weights are of course open to revision.

In order to decide what she has most reason to do, the agent has to decide how much weight to give each of her reasons, so that she can ask whether “this is more important / consistent with / to be ignored compared with / that” (Blackburn, 1998, p. 255) and thus compare them to decide which option has the most justification. Her standing desires give her access to how much weight she thinks each reason has. If she has a desire to \( \Phi \) from which she steps back, she will then consider what reasons she has against \( \Phi \)-ing. After consideration of these reasons if she still wants to \( \Phi \) then her resilient desire to \( \Phi \) is informing her that the reasons she has so far considered do not outweigh the reasons which it is representing. She may then conclude deliberation and decide to go ahead and \( \Phi \), or she might still remain unconvinced that she really ought to \( \Phi \) and so examine those reasons that her desire is tracking to decide whether her desire is misrepresenting their real normative force. After doing this, she might have decided that certain of the reasons her desire was based upon do not have the justification behind them that her desire was purporting they did, and thus decide that she in fact ought not to \( \Phi \). The desire to \( \Phi \) may then fade, but it may be
recalcitrant, and persist despite her decision that it ought not to be acted upon. Its recalcitrance could be an indication that there are reasons for which it is justified that she has not properly considered, and thus may prompt re-deliberation, or its recalcitrance may persist simply because its motivational power exceeds its normative power, as a desire to eat bread might persist even when one is coeliac and thus ought not to. Desires can give false representation of the normative force of the reasons that justify them and thus persist in the light of contrary judgements as to what one has most reason to do after those reasons have been examined. But that the agent can decide this, that she can decide to act contrary to a desire, to resist a very strong urge or to grit her teeth and perform an action for which she feels no realizer desire, means that she is not being ‘blown around’ by her desires. She is not bound to perform whatever she feels the strongest desire to do, and although the virtuous person might be fortunate in that the outcome of their practical reasoning about what they ought to do is not just a judgement that they ought to Φ but also an unwavering desire to Φ and the complete absence of any contradictory desire, the normal rational agent might not. This is why the issue of practical reasoning is not just a judgement as to what she ought to do, but an intention to do that thing.

I have argued that practical reasoning is not a process of trying to be free from our desires and concerns. We do not ‘step back’ to position from which our reasons for action are viewed from an impartial perspective for we need to employ our partiality in the deliberation process in order to decide what reasons we have, what weight to give them in this decision, and therefore what we think we have most reason to do. Thus our desires are employed by reason, and although they can present false represent false representations of our reasons for action and their normative force, and even persist despite our judgements of such, they are not contrary to reason, for their purpose is to deliver the representations about which practical reasoning deliberates. Thus we do not deliberate from a position devoid of desires, as a Kantian model might suggest, but not are we at the mercy of our desires, as a Humean model might. Both of these models see reason as the ‘self’, and desires as assailing the self, but reason and desire do not need to be seen as competing in this way. Just as in the language of Dual Processing
theories System 1 and System 2, although separate, are both as much as *constitutive* of the agent as each other, reason and desire are both as much as constitutive of the self. Borrowing another analogy from Blackburn, we are “no more *passive* when our concerns are contending for a controlling say in our [acting], than a parliament is passive when it debates a law” (1998, p. 251). Reason and desire, System 2 and System 1, work *together* to guide an agent’s behaviour in line with the norm of doing what she has most reason to do. System 1 *alerts* System 2 when it needs guidance and, as I shall argue in the remainder of this thesis, *emotions* play a vital role in our faculties working together like this to promote our being practically rational.
Chapter Four

Emotions

In the previous three chapters I presented my picture of practical rationality. I argued that practical rationality consists in doing what one has most normative reason to do, and that the normative force of reasons for action is derived from the fact that they promote something of value. I argued that what it is for an agent to ‘have’ a particular reason for action is for that consideration to be able to move her towards the action by virtue of its normative force; by virtue of the fact that it promotes something which she values. I argued that an agent is practically rational to the extent that she is guided by the regulative norm of doing what she subjectively ought to do; what she has most accessible reason to do. Her action will be so guided to the extent that it is responsive to her reasons for action, and the capacity for practical reasoning enables her to be so responsive and to guide her actions via the regulative norm. Although actions do not have to be the direct result of practical reasoning in order to be so guided, they do need to be under her active regulative guidance control, such that if she needed to engage in practical reasoning in order to do what she has most reason to do she would do. I argued for a picture of practical reasoning that, rather than having reason and desire as opposing forces, had them working together in order to guide an agent’s actions towards the goal of practical rationality. This picture of practical rationality makes room for the overwhelming evidence that our emotions, not just contribute to, but are a fundamental part of our practical rationality.

There has, in recent years, emerged what Karen Jones (2006) has dubbed a “pro-emotion consensus”, replacing the old assumptions that emotions disrupt practical rationality as being irrational, or at least arational, and arguing instead that emotions are evolutionarily-adaptive reason-tracking mechanisms that enable us to make fast decisions in response to problems we encounter in our environments. Emotions promote our “ecological rationality”, our ability to cope successfully with our environments (Mameli, 2004), and mammals need them in order to get along successfully in the world. This claim is not controversial: no
one doubts, for example, that fear plays an important part in mammals’ successfully avoiding predators, disgust in their avoiding foods and acts that are unhealthy, emotional facial expressions in their effectively communicating with each other (ibid., p. 159). The controversial claim is that emotions contribute to our \textit{practical} rationality; our ability to make the right \textit{choices}.

The dominant philosophical positions, as I have already argued, see desire and reason as being \textit{conflicting} forces when it comes to our making decisions about how to act; as desire either leading reason astray, or as reason needing to control desire. They likewise see emotion as having either a disruptive influence on reason, or else something that reason needs to control. That emotions contribute to our \textit{ecological} rationality is consistent with either of these positions, for their being \textit{evolutionarily} adaptive does not mean that they are beneficial to us in our current lives. Fear might be beneficial to non-human animals for without it they would not ‘know’ that they ought to run away from a predator, but humans can know that they ought to run away without feeling fear, for we have the capacity to \textit{reason}; to \textit{decide} that something is dangerous and to \textit{decide} that we ought to run away from it. More than that, we can \textit{distinguish} between when something really is or isn’t dangerous, and between situations where such danger means that we ought or ought not to run away, whereas non-human animals will react \textit{automatically} to feeling fear and thus run even when they ought not to. That they prime us for certain actions \textit{automatically} poses a threat to our practical rationality: emotions produce ‘false-positives’: they \textit{mislead} us because there is, as Peter Goldie argues, an “environmental mismatch” between the environments in which these emotional capacities developed, and the environments in which we now live (2008, p. 154ff). This means that, although emotions still have a function in enabling us to behave in ways that coincide with what we ought to do in those situations that are sufficiently similar to those in which they developed (for example, staying a safe distance from a lion whilst on Safari), they are a disabling influence in our being able to do what we have most reason to do in situations where what we have most reason to do is not what emotions prime us to do, and these situations make up the majority of those in which we find ourselves in the modern world.
In particular, they are held to be a disruptive influence on our practical *reasoning*, for they prime us for automatic behaviours, present our situation as calling for those automatic behaviours, and lead us astray with their motivational pulls. As Goldie puts it, the emotions “systematically mislead us, and they can do so in ways that are systematically hard to detect and correct” (2008, p. 149). They are thus seen as something we’d be better off, from the perspective of practical rationality, without. We may not be able to escape their effects, but we should try to do so to the best of our abilities so that we can mitigate their misleading us into acting in the way they have primed us for instead of acting in accordance with what we have most reason to do. Fear might alert us to the fact that something is a *potential* threat, but we should use rationality to decide whether it is or is not a real threat, and how we really ought to respond to it. The problem with emotions, so the story goes, is that they pull us towards acting in the way for which they primes us, and thus make reason their ‘slave’, with nothing to do but fulfil the goals they automatically set for us, or else are the wild horses that reason has to control and steer in the right direction. They prime us to behave in ways that conflict with what we in fact have most reason to do, and are thus disruptive influences on our practical rationality. We’d be better off without them, or at least we’d be better off without them in all of those situations where they might lead us astray. Their usefulness is limited to those situations in which they prime us for automatic behaviours that are ecologically rational, such as running away from bears and lions. The rest of the time, we should try to be as little influenced by them as possible, trying our best to remain in a cool, calm and deliberative state, resisting their influences on the ways in which they make us think, feel and act. As soon as we feel the pull of emotion we should stop and reflect, prevent it from manifesting into a state that has any say in how we behave. We should stay as unemotional as possible, using *reason* to decide what to do. Emotions, on the traditional view, might occasionally be useful, but for the main we’d be more practically rational without their influence.

My proposal is that we would *not* be more practically rational without the influence of emotions; we would be *less* so. Despite their ability to mislead us, they are
something without which we could not guide our behaviour in line with the norm of practical rationality. This is a contingent claim: given the sort of creatures we are we need the influences of emotion. I am not ruling out the possibility that there could be a creature which could be perfectly practically rational without any emotional capacities, though such a creature would need to be wildly different to us to be able to be practically rational in our world. I am also not denying that emotions can lead us away from doing what we ought to do, and in Chapters 7 and 8 I shall argue that they can make us weak-willed and make us lose self-control, both of which will prevent us from being practically rational. My argument is not that in every situation emotions assist our being practically rational, but that they are part of the capacities by which we are ever able to be practically rational. I have argued that an agent is practically rational to the extent that she guides her behaviour in line with the regulative norm of doing what she has most accessible reason to do, and that she does this by responding to her reasons. My argument is that emotions are integral to the system by which she guides her behaviour in this way. They do not just enable her to respond appropriately in those stereotypical evolutionarily-adaptive situations such as running away from a bear, but that they play a number of key roles in the whole system by which she guides all of her behaviour. They are a vital part of the capacities and mechanisms by which human agents guide their behaviour in line with the norm of practically rationality, and thus they are, far from being a force that runs contrary to reason, an integral part of the whole system by which an agent is practically rational. Emotions should not be seen as either ‘for’ rationality nor ‘against’ rationality: they are simply part of rationality.

In Chapter 5 I shall argue that emotions ‘provide’ normative reasons for action in two ways. The first is by being reasons per se, but this is not the usual case: the usual role for emotions is to give the agent the access to her normative reasons for action which she needs in order to be able to respond to them, and to thus act for them. I argue that emotions provide this access in at least three ways: they provide access in situations where such is needed more quickly than can be gained by any other route; they furnish us with concepts we need in order to respond to certain of our reasons; and they may be, in some cases, the only route
by which we can access certain of our reasons. Thus emotions contribute to our practical rationality by providing us with the access we need in order to respond to our reasons for action and thus guide our behaviour in line with the regulative norm of doing what we have most reason to do. In Chapter 6 I argue that emotions play a number of roles in practical reasoning, and thus contribute to our practical rationality by enabling us to guide our behaviour via this regulative capacity. Emotions play a vital role in some, or indeed all, of the stages of conscious deliberation such that without emotion we would, at least in some cases, fail to engage in practical reasoning when we ought to, fail to be able to assess the normative forces of our reasons, or fail to come to a decision about what we ought to do. Certainly, there is sufficient evidence to argue that emotions play an important and positive role in the reasoning process and so, far from being contrary to reason, they enable reason to do its job.

But before I can do this, we need to know what we are talking about when speaking of an ‘emotion’. I am not concerned to argue for any specific theory of emotions, and so the purpose of what follows is to give the reader an understanding of the background of philosophical literature of theories of emotion. In much of the twentieth century, the emotions were neglected. Recently, they have enjoyed something of a resurrection, not only in philosophy, but in other disciplines too. Interdisciplinary exchange – particularly with psychology, cognitive science and economics – has had an impact on philosophical theories of emotions, which seeks to develop a systematic theory of the phenomena which fall under the label of ‘emotions’. Such phenomena include what are known as ‘episodic’ emotions, such as fear, joy, pride, anger and compassion, and also, under some theories, moods, sentiments, emotional dispositions and character traits. I am primarily concerned with episodic emotions, and so, unless I indicate otherwise, I am referring to episodic emotions whenever I use the term ‘emotion’.
There is a lack of consensus on the question of what emotions really are, and indeed whether they are in fact any one sort of ‘thing’ at all. I do not seek to provide an argument for any particular theory of emotion, although my sympathies lie with a broadly perceptual-cognitive approach. Instead, I will introduce what are generally accepted as being the central features of emotion. These are that they involve bodily sensations, or feelings, and have particular phenomenologies; they are intentional, being directed at intentional objects, and also have ‘formal objects’, or emotion-proper properties; they are subject to epistemological standards, as well as accuracy conditions fulfilled by a mind-to-world direction of fit; they supply *prima facie* reasons for evaluative beliefs, analogous to the way in which perceptions provide such; and they focus our attention on aspects of our environment, making certain features salient. I shall provide a brief overview of the variety of philosophical approaches towards a theory of emotion, drawing from these what are generally taken to be common features that most, if not all, emotions have in common.

The most simple theory of emotions is that they are a class of feelings. A prominent version of this, the ‘James-Lange’ theory, proposes that emotions are feelings caused by physiological changes in our autonomic and motor functions. The emotion is constituted by our awareness of our bodily responses, such that “we feel sorry because we cry, angry because we strike, afraid because we tremble, and [it is] not that we cry, strike or tremble because we are sorry, angry or fearful, as the case may be” (James, 1884, p. 190). This brings me to the first central feature of emotions, that they tend to involve bodily sensations or feelings, and also they have a phenomenology. In this way, emotions contrast with desires, which are also usually regarded as affective states. Desires are not ‘felt’: though there is perhaps a phenomenology of having an urgent desire, there is not a phenomenology of having a specific desire – there isn’t something it specifically ‘feels like’ to desire to get seats in the front row of the Grand Tier at The Royal

---

39 See Ronnie De Sousa’s twenty-nine answers to the question of ‘What is an emotion’ (de Sousa, Emotional Truth, 2011, p. 27) for twenty-nine possible ‘things’ they could be, and Amélie Rorty’s claim that emotions do not form a natural class (Rorty, 1980).
40 After William James and Carl Lange
Opera House for Wozzeck.\textsuperscript{41} However, so-called ‘feelings theories’ that simply identify emotions with bodily sensations rule out the possibility of unfelt feelings, and such theories fall prey to a number of objections, two of the most prominent being that they fail to give an adequate account of the differences between emotions, and that they fail to account for the various relationships emotions have to rationality.\textsuperscript{42}

More recent, sophisticated feeling theories have to an extent overcome these objections by integrating the observation that emotions are ‘intentional’; that is, they are typically directed at intentional objects. It is these intentional objects that differentiate emotions, and also account for emotions’ links to rationality. For example, a tension in one’s stomach could be ‘fear’ if the intentional object one is approaching is a tiger, or ‘excitement’ if the intentional object one is approaching is an admired celebrity. Emotions’ link with rationality is grounded in their being ‘feelings towards’ things in the external world, such that they can reveal things about the world, for example that there is something frightening in one’s immediate environment (Goldie, 2004).

This, then, is the first central feature of emotions: that they are \textit{intentional}. This term is not used in the sense of their being ‘intended’ as in voluntary, but in the sense that they have intentional objects. It has been argued that this intentionality is what distinguishes emotions from moods and other affective states, though the distinction is not universally held.\textsuperscript{43} There are a wide range of potential intentional

\textsuperscript{41} There are some urgent desires that do seem to have a specific phenomenology, such as hunger, thirst, and others associated with physiological needs, but such can be seen to form a different ‘class’ of desires that are more like ‘tugs’. For example, see Stephen Schiffer’s distinction between “reason-providing” and “reason-following” desires (Schiffer, 1976).

\textsuperscript{42} The first of these objections was first voiced by Walter Cannon in 1915, on the basis that since the visceral reactions characteristic of distinct emotions such as fear and anger are identical, such cannot be what enable us to distinguish these emotions from each other (Cannon, 1929). A huge number of studies were then undertaken in order to try to establish the physiological differentiators of the emotions, the vast majority of which concluded that there were none that were consistent and clear-cut (e.g. (Schachter & Singer, 1962)), thus the differentiae of specific emotions must be something other than the physiological. An example of the second objection is that they fail to account for how emotions can provide reasons for action.

\textsuperscript{43} For example, Julien Deonna and Fabrice Teroni reserve the term ‘emotion’ for episodic states and specifically distinguish them from moods and other affective states (Deonna & Teroni, 2012, p. 8). However, Goldie (2000) argues that moods have the \textit{whole world} as their intentional object. Deonna and Teroni suggest that this would imply that moods were always unjustified, else they
objects for an emotion. These intentional objects provide the *reasons* for the emotion and include such things as people, the relevant facts about the situation, aspects of those facts and the causal role of those aspects.\textsuperscript{44} For example, if Clive is angry because his friend Vernon has said that he will go to the police to inform them that Clive failed to report as a witness to a crime the objects of Clive’s anger can be identified as being Vernon, Vernon’s words, Vernon’s betrayal of Clive, and many other aspects of the situation which Clive can take to justify his anger at Vernon.\textsuperscript{45} If an agent has emotion \(e\), then there is some object \(o\) or proposition \(p\) such that the agent is \(e\) at or with \(o\), about \(p\) or that \(p\), or in virtue of \(p\).

In addition to having these various sorts of objects, an emotion has a ‘formal object’, what I shall refer to, following Peter Goldie, as an “*emotion-proper property*” (Goldie, 2004, p.94; Goldie 2004b), which is a property implicitly ascribed by the emotion to its target or propositional object, in virtue of which the emotion is intelligible, and which is essential to the definition of that particular emotion. For example, the emotion-proper property of disgust is of being disgusting, of fear, being frightening, of shame, being shameful, and so forth. This is another feature neglected by simple feelings theories that is generally now accepted as a central feature of emotions: emotions involve *evaluations*.

Early cognitivist accounts of emotion held that emotions in fact were evaluative judgements or beliefs. These followed the Stoics and psychological appraisal theories characterizing emotions primarily in terms of their associated cognitions, usually in terms of propositional attitudes, for example by identifying emotions with judgements (Solomon, 1980), or judgements plus other elements such as complexes of beliefs, desires and feelings (Oakley, 1992).\textsuperscript{46} The phenomena of

\[\text{Page 110 of 224}\]
recalcitrant emotions poses a problem for such theories, as if emotions are simply judgements, in such cases one would be simultaneously holding two conflicting judgements and would thereby be guilty of blatant irrationality, when it seems that in such cases we have rather a ’conflict without contradiction’ (Döring, 2009).47

That emotions have intentional objects allows us to criticize emotions as being justified or unjustified. This is another central feature of emotions: they are subject to *espistemological standards*. The justification of emotions depends upon the reason the subject has for them. For example, if Clive’s indignation at Vernon was based upon a claim made in a down-market tabloid newspaper about what Vernon had said, Clive may not have any good reason, normatively speaking, to be angry, given that the report is likely to be grossly exaggerated, at best. Emotions share this feature with other cognitive states, such as beliefs, but not, come argue, with moods, for which we do not request reasons as justification (Deonna, 2012, p. 7).

That emotions have emotion-proper properties also allows us to criticize them in term of whether or not they are appropriate, that is, they are subject to *accuracy conditions*. Like many other cognitive states such as beliefs and perceptual experiences, emotions have a content such that it is possible to assess whether or not they are ‘fitting’ to the facts. For example, the emotion of disgust is appropriate only if the object of the emotion does indeed have the emotion-proper property of being disgusting. This seems to distinguish emotions from other psychological states, such as dispositions, moods and desires, which do not have this ‘mind-to-world’ direction of fit.48

It is worth here distinguishing the judgement of the appropriateness of an emotion from a standard of propriety or morality (D’Arms & Jacobson, 2000). An emotion may be appropriate, yet be morally wrong. For example, envy would be

---

47 A recalcitrant emotion is one that ‘refuses’ to fall into line with one’s conflicting evaluative judgement, persisting despite one’s belief that it is not justified or accurate.
48 Indeed, desires have precisely the opposite; a world-to-mind direction of fit (Searle, 1983), in that desires aim to bring the world into line with what the mind conceives as desirable.
appropriate if the object of one’s envy had the emotion-proper property of being enviable, but still one might want to say that it is always wrong, morally, to be envious.\textsuperscript{49} There might also be prudential standards to which emotions are subject (Deonna & Teroni, 2012, p. 7), for example anxiety in response to one’s imminent \textit{viva voce} may be both justified and appropriate, yet still, one would be better off without it as it may bring about the very thing about which one is anxious: poor performance in the exam.

The mind-to-world direction of fit is inspired by the model of perception, and so we come to perceptual accounts of emotion, wherein my sympathies lie.\textsuperscript{50} This view of emotions has them as playing an essential role in our gaining evaluative knowledge of the world; they enable us to see the world as it is. This variant of cognitivism construes emotions as being analogous to perceptions more than to beliefs and can deal with recalcitrant emotions, with which other cognitivist accounts struggle. Visual illusions that persist despite contradictory judgements that they are illusions, such as the Müller-Lyer illusion, are an analogous phenomenon in perception: we are able to hold the conflict without contradiction.

Another benefit of perceptual theories is that they can provide a plausible criterion of success that depends on correctedness with respect to some objective property, for example an emotion-proper property. The sense of ‘truth’ applicable to emotions is closer to that of ‘veridical’ applicable to perceptions, than is to that applicable to beliefs. Just as the perceptual system can, at times, produce false representations (as in the case of hallucinations), emotions can mislead us into incorrect judgements. The content of our perceptions are only \textit{prima facie} reasons for the relative empirical belief (that is the belief that the world really is as we perceive it to be); analogously, the content of our emotions are only \textit{prima facie}.

\textsuperscript{49} For example, see Gabriele Taylor’s discussion of envy and jealousy as subjects for vice (Taylor, 1988), but also Goldie’s defence of their being appropriately assessed “in the round”, and possible of being virtues (Goldie, 2000, Ch8).

facie reasons for the relative empirical belief (that the object of our emotion really does have the relevant emotion-proper property) (Goldie, 2004, p. 251).

The representations produced by our emotions can be defeated. For example, when taking an evening stroll, one may be frightened as a bird takes flight from the ground, momentarily perceiving the creature as having the emotion-proper property of being frightening (for in the fading light, that it was just a bird was initially unperceivable, it appearing to be much greater in size, and appearing as if from nowhere, it having blended in to the shadows). The characteristic bodily sensations of fear are experienced – quickened heart rate, tightened muscles – until the prima facie representation is defeated by the realization that it is just a bird, and as such does not actually have the emotion-proper property of being frightening. The bodily sensation then subsides. This is the typical case, though there are, of course, non-typical cases, where the object may still seem to have the emotion-proper property (recall Barbara’s persistent fear, despite judging the cliff to pose little danger). The justifying reasons for an emotion are also those reasons which justify the ascription to the object of the emotion-proper property: it is not one’s perceiving the object as having the emotion-proper property that justifies one’s emotion, nor is it one’s having the emotion that justifies one’s perceiving the object to have the emotion-proper property, but rather the fact that the object does have the emotion-proper property justifies both one’s having the emotion, and one’s perceiving the object as having the emotion-proper property (Goldie, 2004, pp. 254-5).

A variety of perceptual accounts has emotions as ‘ways of seeing’ rather than as articulated propositions. Emotions are construed as determinate patterns of salience among objects of attention (de Sousa, 1987) and inferential strategies (Roberts, 2003) that makes features of the situations in which we are more prominent by ‘filtering’ our perception of our surroundings (Wollheim 1999, Döring 2008). Ronnie de Sousa argues that emotions enable us to solve what is known among cognitive scientists as the ‘Frame Problem’, which is that the number of goals which one could pursue at any time is virtually infinite, as is the number of strategies by which at attain them, and so without some method by which the
majority of these are discounted *a priori*, a decision would never be reached. Emotions constitute one of the chief mechanisms whereby our attention is constrained and directed, so that only a tiny proportion of the conceivably relevant facts are rendered salient – a process labelled (and criticised) by Dylan Evans as the ‘Search Hypothesis’ of emotions (Evans, 2004). I am most sympathetic to de Sousa’s view, or at least to the observation that whether or not this is what emotions *are*, it is a central feature to emotions that this is one of the things that they *do*. For example, fear focuses our attention on the object perceived to have the emotion-proper property of being frightening.

Thus we have what I assume to be the central features of emotions: they typically involve bodily sensations or feelings, and have particular phenomenologies; they are intentional, being directed at intentional objects, and also have ‘formal objects’, or emotion-proper properties; they are subject to epistemological standards, as well as accuracy conditions fulfilled by a mind-to-world direction of fit; they supply *prima facie* reasons for evaluative beliefs, analogous to the way in which perceptions provide such; and they focus our attention on aspects of our environment, making certain features salient. In the next two chapters I argue that these features of emotion mean that they are able to contribute positively to our practical rationality.
Chapter Five

Emotions Provide the Access to Normative Reasons and thus Enable Practically Rational Action

I have argued that practical rationality consists in doing what one has most reason to do, and that what it is for an agent to ‘have’ a particular reason for action is for it to be able to move her towards that action by virtue of its normative force. Agents are practically rational to the extent that they guide their behaviour in line with the regulative norm of doing what they have most accessible reason to do, and that they will do this to the extent that they respond to their reasons for action. To be able to respond to a normative reason the agent needs to have access to that reason, and in this chapter I argue that emotions provide that access in at least three ways: in many cases we need to have very quick access to reasons in order to be able to respond to them under time constraints, and emotions provide a quicker route to these reasons than can conscious deliberation; emotions furnish us with concepts which we need in order to be able to respond to the related reasons; and the only route by which we can access certain reasons is emotion. Without access to a reason an agent cannot respond to that reason, and thus emotions contribute to practical rationality by enabling an agent to respond to, and thus act for, the normative reasons for action which she has. Thus, emotions contribute to our being practically rational, contrary to the traditional view of their necessarily interfering with rationality.

I shall begin this chapter by arguing that when we cite emotions as explanatory reasons we are doing so as a short-hand way of indicating the likely psychological states that are the real explanatory reasons of an agent’s action; that is, we are not usually making the claim that emotions per se are explanatory reasons for an action. In section 2 I argue that the same is true for when we cite emotions as normative reasons for action: although there are cases in which the emotion per se is a normative reason, in the usual case it is the emotion-proper properties of the situation which are the real normative reasons for action, and our citing the emotion is a short-hand way of indicating the likely emotion-proper properties which are the normative reasons that justified the agent’s action (or those
considerations she treated as normative reasons), with the emotions providing access to these reasons.\(^{51}\) I then argue that there are at least three ways in which emotions provide such access. In section 3 I argue that emotions provide ‘quick and dirty’ access to reasons by focussing our attention on our reasons and priming our bodies to respond appropriately to those reasons. In section 4 I argue that emotions furnish us with concepts that we need in order to be able to respond to reasons in the future. In section 5 I argue that emotion may be the only path of access to particular reasons, and propose the emotion of love as a likely candidate. In providing access to our reasons for action, emotions contribute to our practical rationality, but, more than that, they are part of the mechanisms by which human agents can be practically rational. We are rational to the extent that we guide our behaviour in line with the norm of doing what we have most reason to do, and an action is rational to the extent that it responds to our reasons for action. We can only respond to those reasons to which we have access and so, by furnishing us with that access, emotions are part of the capacities we need in order to effectively guide our behaviour in line with the norm and be practically rational.

1 Emotions as Explanatory Reasons

We commonly cite emotions as explaining actions; that is it is common practice to talk as if emotions were explanatory reasons. For example, the explanation given for Otello murdering Desdemona is that he is jealous, for Fred proposing to Mary that he is in love with her, for Barbara’s running away that she is frightened. We also cite emotions (perhaps anthropomorphically) as explanations for the behaviour of infants and animals: we say things such as that the dog bounded up to us wagging its tail because it was happy to see us, that the squirrel ran up the tree when the fox approached it because was afraid, that the cat swiped at the child who pulled its tail because it was angry, that the baby cried because it was

\(^{51}\) Recall fro the previous chapter, an “emotion-proper proper[y]” (Goldie, 2004, p.94; Goldie 2004b) is the ‘formal object’ of an emotion: the property implicitly ascribed by the emotion to its target or propositional object, in virtue of which the emotion is intelligible, and which is essential to the definition of that particular emotion. For example, the emotion of fear ascribes to the thing feared the property of being dangerous.
distressed. In this section I argue that when we are doing this, we are not actually saying that the emotion *per se* is the explanatory reason for action, for the explanatory reasons of human action are the agent's psychological states of beliefs and desires. When we cite emotions as explaining actions we are rather indicating the likely beliefs and desires that are the actual explanatory reasons for the action. First, let us recall the difference between normative reasons for action, and explanatory reasons of action.

In Chapter 1 I drew distinctions between normative reasons, operative reasons, and explanatory reasons. Normative reasons are propositions which are true and count in favour of the action for which they are normative reasons. Operative reasons are those propositions *for* which an agent acts. Operative reasons may or may not be normative reasons; an agent can act for operative reasons which do not reflect normative reasons, but her action would not be justified as only normative reasons can transmit justificatory force. Explanatory reasons are facts that explain why the agent did as she did, and include facts about her psychology, such as her beliefs and desires. For example, say that raw kidney beans are poisonous for humans, and that this counts in favour of ensuring kidney beans are cooked before eating them, so ‘they are poisonous when raw’ is a normative reason to ensure kidney beans are cooked before eating them. James checks his kidney beans are well cooked before eating them for this reason, so his operative reason for checking his kidney beans are cooked is that ‘they are poisonous when raw’. However, what explains James’ action of checking his kidney beans are well cooked are his desire to avoid being poisoned and his belief that checking his beans are well cooked is a good way to fulfil this desire: his explanatory reasons are his belief and desire, his psychological states. In this example because it is true that kidney beans are poisonous when raw, James’ operative reason is a normative reason, and so his action is justified. However, were it not true that raw kidney beans were poisonous, although the explanatory reasons for James’ action would be the same, and the operative reason for which James acted would be the same, the operative reason would not be reflecting a normative reason, and so James’ action would not be justified.
Citing the psychological states of an agent as explanatory reasons for her action makes possible a rationalizing explanation of the agent’s action. Rationalizing explanations of action usually attribute to an agent a pro-attitude, such as a desire, and a cognitive state, such as a belief. The belief-desire pair makes the agent’s action intelligible, and distinguishes it from non-rationalizing explanatory reasons. For example, Christina’s leaping over the fence into the next field to escape a bull can be made intelligible by citing the rationalizing explanation that she desired to escape the bull and believed that leaping over the fence would enable her to fulfil this desire. Such a rationalizing explanation cannot be given for the example of my driving through a red light because I wasn’t paying attention, because although my not paying attention explains why I went through the red light, I did not have a desire that I believed would be fulfilled by doing so.

How do emotions figure in explanations of action? Citing an emotion as an explanation for an agent’s action is not usually a claim that the emotion per se is the explanatory reason for the action, but rather it indicates the sorts of pro-attitudes and cognitive states that may have been the reasons that explain the agent’s action. For example, if we say that it was because Christina was afraid that she leaped across the fence once she realized there was a bull in the field, we are indicating that Christina had cognitive states typical fear, such as ascribing the emotional-proper property to the bull of being frightening, and the characteristic pro-attitudes of fear, such as a desire to be safe, and it is these beliefs and desires which provide the rationalizing explanation for her action. In this way, citing the emotion as an explanation is a short-hand indication of the likely explanatory reasons. If this were not the case, and citing the emotion was intended to be the full explanation for her action, it wouldn’t be a very good explanation. It would leave something unexplained, namely, why did Christina’s being afraid mean that she leapt over the fence, rather than that she performed some other action? ‘Because she was afraid’ would equally well explain a range of other actions that are characteristic of fear, such as freezing still. Thus why she leapt rather than froze would remain unexplained if the entire explanation was

---

52 Donald Davidson calls the pair of the pro-attitude and the related belief the ‘primary reason’ that explains why an agent performed an action (1963).
‘because she was afraid’. Citing the emotion is not, therefore, usually the whole explanatory story, but we can infer from the indicated emotion what the whole story is, given the action, as we can infer which of the psychological states that are characteristic of fear intelligibly explain that particular action.53

There is a distinction between the above sort of rationalizing explanations, and those in which the emotion is part of the content of the relevant psychological states. For example, Maeve wakes up from a bad dream feeling afraid and switches on the light. The explanation for her switching on the light is that she was afraid. However, there is nothing of which she is afraid, so she is not assigning the property of being dangerous to anything from which she desires to escape, a desire which we could intelligibly say she thought switching on the light would fulfil. She is not switching on the light in order to facilitate escape, for example, by illuminating the way to a place of safety from a source of danger, such as a monster in the closet.54 In cases like Maeve’s, the explanation 'she was afraid' indicates a rationalizing explanation in which the emotion is the content: Maeve believes that she is afraid, and believes that she ought not to be afraid, and so desires to cease being afraid and believes that turning on the light will fulfil this desire.

Emotions can also figure in non-rationalizing explanations of action. For example, if the explanation for my driving through the red light was ‘because I was angry’, this could explain why I wasn’t paying attention to the traffic lights, for my attention was directed towards the object of my anger (perhaps going over the situation

---

53 Similarly, Davidson argues that “many explanations of actions in terms of reasons that are not primary do not require mention of the primary reason to complete the story” (1963, p. 688), and gives two examples involving emotions: “Jealousy is the motive in a poisoning because, among other things, the poisoner believes his action will harm his rival, remove the cause of his agony, or redress an injustice, and these are the sorts of things a jealous man wants to do. When we lean a man cheated his son out of greed, we do not necessarily know what the primary reason was, but we know there was one, and its general nature” (ibid., p. 689).

54 If she was attempting to illuminate the way to a place of safety in order to escape then the case would be like the usual ones, as illustrated in the preceding paragraph, although the operative reason for which Maeve acts – that there is a monster in her closet – is not a normative reason – since there is no monster in her closet – thus she would not be doing what she has most normative reason to do. She might, however, be doing what she has most accessible normative reason to do if we told the story such that all the normative reasons she had to which she had access (thus discounting the operative reasons of there being a monster in her closet) supported her doing so.
about which I was angry in my head) rather than on the road ahead. There could also be a rationalizing explanation of this situation, for example in my anger I *desired* to exhibit aggression and *believed* that stubbornly refusing to obey the rules of the road would satisfy this desire, but this would be describing a different action. In the first example, I did not intentionally drive through the red light, and in the second I did.

In all of these cases, citing the emotion as an explanation for the agent’s action is a short-hand way of indicating the likely psychological states which constitute the explanatory reasons. However, there are also cases in which an emotion is given as an explanation for something which an agent does, where it is *not* a short-hand way of indicating the agent’s beliefs and desires. This is the case for what we might call *expressions of emotion*, for example, jumping for joy, throwing something in anger, or embracing someone out of love. 55 Such expressions of emotion seem to fall short of being what we might call ‘reasoned’ actions because they do not have rationalizing explanations in terms of beliefs and desires (Goldie, 2004). As Rosalind Hursthouse argues, we do not say of an expression of emotion, such as jumping for joy, that “the agent [did] it for a reason in the sense that there is a true description of action in the form ‘X did it (in order) to …’ or ‘X was trying to …’ which will reveal the favourable light in which the agent saw what he did, and hence involve, or imply, the ascription of a suitable belief …” (1991, p. 59). In such cases, “the mere fact that she was in [the emotion’s] grip explains the action as much as anything else does” (ibid). In such cases as these, the emotion *is* the *whole* explanation: there is nothing else more ‘basic’ that explains the action, nor nothing more to add, such as that she believed she was experiencing that emotion. 56 Because such actions are not done for reasons, they are not subject to rational criticism: there is no way in which the agent can be

---

55 Rosalind Hursthouse has many examples of such actions that are explained by various emotions, including shame, horror and pride (1991, p. 58)
56 There may be tokens of these action types that can be given rationalizing explanations in terms of a belief-desire pair. For example, one may throw something *in order to communicate one’s anger*, or *in order to feel better*, or *in order to try to give the impression of being more angry than one actually is*. Such actions would not be genuine expressions of emotion (see Goldie, 2000, p. 26, Hursthouse, 1991, p.59) because they were done *in order to achieve some purpose*, that is, as a means to some further end. They can therefore be given the normal belief-desire rationalizing explanation.
‘getting it wrong’. Such actions are, to borrow Hursthouse’s term, ‘arational’ (Hursthouse, 1991, p. 58), rather than irrational, (following the way in which actions can be amoral rather than immoral) in that they are not in the business of trying to respond to reasons.

Thus emotions are often given as explanations of actions, usually as a short-hand description of the explanatory reasons of beliefs and desires which explain the action, but sometimes as elementary explanations, as with expressions of emotion. However, my thesis is concerned with emotions as normative reasons for action; that is with emotions as justifying actions. In the case of expressions of emotion, I take it that the brute fact of the emotion justifies the action, in as much as the action requires justification. If we accept that expressions of emotion are arational, then, being excused from rational criticism, they do not require any further justification. However, reasoned actions, that is actions which are explained by belief-desire rationalizing explanations, do. In the next section I argue that just as citing an emotion as an explanation for an action is usually a short-hand way of indicating the likely beliefs and desires which are the actual explanatory reasons for the action, citing an emotion as a justification for an action is usually an short-hand way of indicating the likely emotion-proper properties that are the actual normative reasons that justify the action, or at least count in favour of the action.

2 Emotions as Normative Reasons

We have seen that when emotions are cited as explanations for reasoned actions, this is usually a short-hand way of indicating the rationalizing explanation of the agent’s action – the beliefs and desires which are the explanatory reasons. Beliefs and desires are, as I argued in Chapter 1, rarely the agent's operative reasons, that is, the reasons for which she acts. The explanatory reasons for Christina’s leaping over the fence are her desire to escape and her belief that leaping over the fence will fulfil that desire. Her operative reason is that the bull is dangerous. In Christina’s case, her operative reason reflects a normative
reason: ‘the bull is dangerous’ is true, and counts in favour of leaping over the fence to escape it. Thus Christina’s action is practically rational, for it is responding to her normative reasons, namely that the bull is dangerous. The normative reason justifies Christina’s desire to escape and this justification transmits to those actions which will fulfil that desire, and thus counts in favour of her leaping over the fence.\textsuperscript{57} The normative reason thus transitively justifies her action.

\begin{center}
\begin{tabular}{|c|c|c|}
\hline
\textbf{Normative reason} & \textbf{Rationalizing Explanation} & \textbf{Action} \\
\hline
Bull is dangerous & \textmd{justifies} & Desire to escape \textmd{justifies} \\
\hline
Belief that leaping over the fence is a good means of escape & \textmd{justifies} & Leaping over fence \\
\hline
\end{tabular}
\end{center}

Where do emotions fit in to this picture? Recall from the previous chapter that emotions involve ascribing emotion-proper properties to their objects. In the case of Christina’s bull, her emotion of fear ascribes to the bull the emotion-proper property of being dangerous.\textsuperscript{58} That the bull is dangerous is in this case is a normative reason for leaping over the fence, hence in having the emotion of fear Christina is \textit{latching on to} the normative reason of the bull being dangerous and that means that she can leap over the fence \textit{for} the reason that the bull is dangerous: Christina’s emotion of fear enables the normative reason that the bull

\textsuperscript{57} In Chapter 1 I argued that normative reasons for action derive their normative force from the values by virtue of which they count in favour of the particular action. The normative force of is transitional, such that the valuing of \( A \) justifies the intrinsic desire that \( A \) be the case, which provides the justification for the \textit{instrumental} desire to do whatever it is that will bring it about that \( A \) is the case, which then provides the justification for the realizer desire to perform the action that will promote \( A \).

\textsuperscript{58} Note that ascribing to the bull the emotion-proper property of being dangerous need not involve a \textit{conscious belief} that the bull is dangerous.
is dangerous to be her operative reason for leaping over the fence. Her emotion of fear thus enables her to be rational, since it enables her to do what she has reason to do.

Emotions thus provide us with normative reasons for action, in that because we have the emotion we have access to these reasons. Having access to a normative reason means that we are able to act for it, that is, it can be one of our operative reasons for action. ‘Access’ here does not imply that the agent need have a conscious representation of the reason as a reason, but there does need to be a certain sort of relation between the fact and the action, such that the fact counts in favour of the action, and the agent needs to have some sort of awareness of this relation, though not a judgement that it exists. It is enough, referring back to Chapter 2, that the agent’s reason-tracking mechanisms are tracking the reason.59

When we cite an emotion as a normative reason for action, for example when we say that Christina’s fear justified her leaping over the fence, this is a short-hand way of indicating the actual normative reason that justifies the action, just as citing the emotion as an explanatory reason is a short-hand way of indicating the actual explanatory reasons that explain her action. The normative reason that actually justifies Christina leaping over the fence isn’t that she ‘happens’ to be scared, but is rather the fact that justifies her being scared, and transitively thereby justifies her action of leaping over the fence to escape it; the normative reason being that the bull is dangerous. By saying that Christina was frightened of the bull, we are implicitly attributing to her the (unconscious) judgement that the bull was dangerous. If we say that her being frightened justified her leaping over the fence, we are saying that her judgement of the bull as dangerous was correct, and the content of this judgement - the fact that the bull is dangerous – was indeed a normative reason for her action, and justified her leaping over the fence.

59 The propositional content of emotion is more like that of perception than of belief. Just as one does not have to hold a conscious judgement that the bull has horns in order to perceive that the bull has horns, one does not have to hold a conscious belief that the bull is dangerous to perceive that the bull is dangerous.
Why is it not simply the brute fact that she is afraid that justifies Christina leaping over the fence? Compare Christina’s case to one in which the emotion-proper property has been mis-ascibed. Imagine Pauline, who is staying in a holiday cottage and spots a dead mouse on the floor in the kitchen. Terrified, Pauline locks herself in the bedroom and refuses to come out. Pauline’s fear ascribes to the dead mouse the emotion-proper property of being dangerous. This property provides a reason to desire to escape, and coupled with her belief that locking herself in the bedroom would be a good means of escape, provides a rationalizing explanation for her action of locking herself in her bedroom. However, clearly a dead mouse is not dangerous, hence although there is an operative reason for Pauline’s action – that the mouse is dangerous – there is no normative reason which can justify it. If it were simply the fact that she was afraid that provided a normative reason, given that it is true that Pauline has the emotion of fear, there would be a normative reason that justified her action of locking herself in her bedroom, and this would transitorily justify her action. This would entail that many things that we think of as irrational would be deemed rational just by virtue of the agent happening to have an emotion, and so we should reject the idea that the brute fact of the emotion is a normative reason in this case. Most cases are like the one outlined above, where it is not the emotion per se that is the normative reason for action. There are some cases where an emotion is a normative reason, and it would be helpful to get these cases out of the way now, so that we can then see why most cases are not of this kind.

2.1 Emotion as a normative reason per se

The case of Maeve turning on the light because she felt afraid when she awoke from a bad dream is an example of a case where the emotion per se is the normative reason for action. There is nothing ‘behind’ the emotion which could lend justification to her action, and Maeve doesn’t turn on the light in order to achieve anything more than to get rid of her feeling of fear. There is nothing further to be achieved than emotion-regulation. Cases like Christina’s are not like
this: Christina doesn’t just desire to stop feeling afraid, but to actually escape.
Were it the case that all she desired was to stop feeling afraid, then if she could take a pill to get rid of this desire, there would be nothing wrong with doing this. But there would be something wrong about Christina taking a pill to get rid of her fear, instead of leaping over the fence to escape the bull. If Christina took a pill to get rid of her fear, there would still be a normative reason for her to be afraid and to escape (the reason being that the bull is dangerous). If Maeve took a pill to get rid of her fear, there would not remain a reason for her to be afraid and switch on the light. If there remains a reason to have the emotion and perform the action, if the agent took a pill to get rid of the emotion, then it is not the emotion per se that is the normative reason for action, but rather citing the emotion is a short-hand for citing the emotion-proper property as the normative reason for action.

2.2 Emotion-proper properties as normative reasons

When Pauline locks herself in the bedroom because she is afraid of the mouse, her action is not justified because the emotion-proper property which would justify her action – i.e. being dangerous - is not a property which can correctly be ascribed to the dead mouse. This case is different to Maeve’s turning on the light to alleviate her fear because Maeve isn’t ascribing the emotion-proper property of being dangerous to anything, whereas Pauline is. Pauline locks herself in the bedroom in order to escape from the mouse. Pauline’s fear is irrational rather than arational, because it is in response to Pauline taking something to be a normative reason – the mouse being dangerous – which isn’t actually a normative reason (since the mouse isn’t dangerous). Her fear is presented as being something which is responding to a normative reason, but the reason which would justify it doesn’t actually exist. Her desire to escape is therefore not justified, since there is nothing from which there is actually a reason to escape, and thus her action of locking herself in the bedroom is not justified as it is not responding to any normative reason which could justify it. The reason for which Pauline locks herself in the bedroom cannot justify her action because it is not a normative reason, and only normative reasons can transmit justification to action.
We can, of course, say that she was *instrumentally* rational in choosing an appropriate means to satisfy her desire to escape, but given that reasons to act on a desire come not from the existence of the desire itself but from the facts which justify the having of the desire, she was *not* acting for reasons, and thus was not *practically* rational.\(^6\)

Christina *is* responding to reasons. The reason that justifies Christina’s fence-leaping is the *same* reason that justifies her having the emotion of fear. Christina is right to be afraid because the bull *is* dangerous, and she is right to desire to escape because the bull *is* dangerous. If we agree that the normative reason is that the bull is dangerous, then the relationship would be as follows:

\[\begin{align*}
\text{Normative reason} &\quad \text{Emotion} &\quad \text{Action} \\
\text{Bull is dangerous} &\quad \text{Fear} &\quad \text{Desire to escape} &\quad \text{Leaping over fence} \\
\text{Belief that} &\quad \text{justifies} &\quad \text{justifies} &\quad \text{justifies} \\
\text{leaping over} &\quad \text{the fence is a} &\quad \text{the} &\quad \text{the} \\
\text{the fence is a} &\quad \text{good means} &\quad \text{bull is} &\quad \text{bull is} \\
\text{good means of} &\quad \text{of escape} &\quad \text{dangerous} &\quad \text{dangerous}
\end{align*}\]

It is important to see here that it is not the emotion *per se* that is justifying the desire and therefore the action, but that it is the normative reason of the bull being dangerous that provides the justification. The normative reason exists independently of the existence of the emotion. There would be a reason for Christina to leap over the fence to escape the bull, even if she didn’t actually feel

\(^6\) We can imagine a situation in which Pauline would also be instrumentally irrational. For example, she might lock herself in the same room as the dead mouse, which would clearly not be a suitable way of satisfying her desire to escape the ‘danger’ posed by it.
the emotion of fear (if, for example, she had taken a pill which inhibited any feelings of fear). However, Christina might not be able to access the normative reason to escape if she does not have the emotion. Specifically in this case, without feeling fear she might not be able to respond quickly enough to the fact that the bull is dangerous in time to escape it. That is because emotions are activated by a ‘quick and dirty’ route which prompts us in to action more quickly than can conscious deliberation. In some situations, as I shall now argue, emotions assist us in being practically rational by providing us with access to our reasons at a speed which is necessary in order for us to be able to respond: we could not respond to the reason quickly enough by a non-emotional route.

3 Emotions provide ‘quick and dirty’ access to normative reasons

In Chapter 2 I introduced the picture of Dual Processing theories which hold that humans have two distinct mechanisms for thought processing that operate largely independently: one fast, automatic and non-conscious (System 1), the other slow, controlled and conscious (System 2). The table below shows the features attributes by various theorists to the two systems of cognition, as collated by Frankish and Evans (2009, p. 15):

---

61 The terms ‘System 1’ and ‘System 2’ are the most neutral terms (compare, for example, Goldie’s (2008) use of ‘intuitive thinking’ and ‘deliberative thinking’) but Dual Processing accounts of human behaviour are appealed to in so many different ways by different cognitive psychologists that it would be a mistake to assume that there is a clear consensus on how they are thought of, and whether there are in fact (just) two ‘systems’. Some propose that there are two underlying cognitive systems, others that there are only two processes, and the degree to which there are two distinct categories is at issue. However, almost all agree on a distinction between processes that are rapid, automatic and unconscious, and those that are slow, deliberative and conscious. For a review of the appeals made to dual processing in cognitive and social psychology, see Evans (2008).
<table>
<thead>
<tr>
<th>SYSTEM 1</th>
<th>SYSTEM 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evolutionarily old</td>
<td>Evolutionarily recent</td>
</tr>
<tr>
<td>Unconscious, preconscious</td>
<td>Conscious</td>
</tr>
<tr>
<td>Shared with animals</td>
<td>Uniquely (distinctively) human</td>
</tr>
<tr>
<td>Implicit knowledge</td>
<td>Explicit knowledge</td>
</tr>
<tr>
<td>Automatic</td>
<td>Controlled</td>
</tr>
<tr>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>Parallel</td>
<td>Sequential</td>
</tr>
<tr>
<td>High capacity</td>
<td>Low capacity</td>
</tr>
<tr>
<td>Intuitive</td>
<td>Reflective</td>
</tr>
<tr>
<td>Contextualized</td>
<td>Abstract</td>
</tr>
<tr>
<td>Pragmatic</td>
<td>Logical</td>
</tr>
<tr>
<td>Associative</td>
<td>Rule-based</td>
</tr>
<tr>
<td>Independent of general</td>
<td>Linked to general intelligence</td>
</tr>
</tbody>
</table>
Dual processing theories hypothesize that System 1 is designed to very closely track what we might call ‘evolutionary rationality’, that is, survival at the level of the gene (Stanovich & West, 2003, p. 178), which would be achieved by reproductive success, whereas System 2 is attuned to normative rationality, that is the individual’s goals. In most cases, evolutionary rationality and normative rationality will coincide, for example being able to accurately navigate around objects in the natural world is evolutionarily adaptive, but it also serves normative rationality in that it enables us to achieve our goals in the modern world (Stanovich, 1999, p. 150).

There is much evidence to suggest that System 1 processing has an important survival value. Because System 1’s processing is automatic and quick, whereas System 2’s processing is deliberative and relatively slow, System 1 can generate solutions to problems before System 2. In many situations, System 2 wouldn’t be able to generate a solution in time because it is too slow and action needs to be taken quickly if it is going to be taken at all. Emotions are activated by System 1 – quickly, automatically and unconsciously – when specific stimuli which convey particular information comes to the organism’s attention (LeDoux, 1996, p. 175). For example, if System 1 perceives something as dangerous, the emotion of fear automatically occurs (Hanoch & Vitouch, 2004, p. 440). The emotion has two effects which would assist the agent in being practically rational by priming her to respond for the perceived reason: it focusses the agent’s attention on the potential reasons for action, and it triggers the associated bodily changes which get her ready to act. Because emotions are triggered so quickly, they enable the agent to act for the perceived reason more quickly than she would be able to do via conscious deliberation.

---

62 The thesis that emotions are evolutionary adaptive, while appealing and popular, is not, of course, uncontroversial. Some argue that the evolutionary lineage of emotions is compatible with explanations other than that of survival value, for example that they are by-products of genes who are adaptive for other purposes (Elgin, 2008, p. 36), but it doesn’t matter for my argument whether emotions were or were not evolutionarily adaptive for since my concern is with their ability to aid normative, not evolutionary, rationality.
3.1 Emotions focus an agent’s attention on her reasons for action

Emotional arousal reduces the range of information to which we are capable of paying attention (Easterbrook, 1959), and focusses our attention upon the features of our environment that are relevant to the emotion. For example, evolutionary-relevant threatening stimuli (e.g. snakes) capture visual attention more readily than non-threatening stimuli, and the effect is even further facilitated if the stimulus is emotionally provocative (e.g. the person is particularly fearful of snakes) (Öhman, Flykt, & Esteves, 2001). Other features of the situation receive significantly less, if any attention, as a result. For example, the attention of individuals who are robbed with a weapon is so focussed on the weapon, that afterwards they can vividly recall it in detail, at the expense of being able to recall other aspects of the event (Hanoch & Vitouch, 2004, p. 437). As I have already argued, emotions provide access to reasons for action by latching on to the emotion-proper properties of the situation. The features of our environment made salient by emotions are these emotion-proper properties. In the example of people who are robbed with a weapon, the weapon will have the emotion-proper property of being dangerous, and the emotion of fear will focus the person’s attention on to this property. Thus emotions focus our attention upon emotion-proper properties by reducing the range of other features to which we can pay attention.

By reducing the range of information to which we can pay attention, emotions ‘by-pass’ conscious deliberation. For example, the weapon is perceived as being dangerous, and information which might count against its being dangerous is not considered. We do not have to deliberate about whether or not the weapon is dangerous; we automatically judge it (non-consciously) to be so. This is advantageous, for in some situations – possibly those similar to those in which the emotion evolved, if indeed they did so – if we don’t perceive the reason as a reason for action immediately, we won’t have time to act. The speed of emotional arousal enables organisms to react quickly enough to reasons, as evolutionary psychologist Robert Zajonc explains:
A rabbit confronted by a snake has no time to consider all the perceivable attributes of the snake in the hope that he might be able to infer from them the likelihood of the snake’s attack, the timing of the attack, or its direction. The rabbit cannot stop to contemplate the length of the snake’s fangs, or the geometry of its markings. If the rabbit is to escape, the action must be undertaken long before the contemplation of even a simple cognitive process – before, in fact, the rabbit has fully established and verified that a nearby movement might reveal a snake in all its coiled glory. The decision to run must be made on the basis of minimal cognitive engagement (1980, p. 156).

This is a key feature of the success of System 1 processes: making decisions on the basis of minimal cognitive engagement, in the interests of time. Emotions enable us to respond appropriately to our situation precisely because they make ‘decisions’ for us before conscious deliberation can be engaged. Perception of the emotion-proper property will give rise to the associated desire, for example to escape in the face of danger. We will then search (not necessarily consciously) our environment for solutions to satisfy those desires, which results in action. The second way in which the emotion primes the agent to respond appropriately to the emotion-proper property as a normative reason for action is by preparing her physically to do so.

3.2 Preparing the body for action

There is a substantial body of research which supports the claim that some emotions are associated with emotion-specific autonomic nervous system activity which prepares the body for certain actions (Ekman, 1992). Such evidence provides support for the view that emotion-typical behaviours were evolutionary adaptive. For example, blood flows to the hands when one is angry, which would be consistent with the hypothesis that fighting was an adaptive action in anger, and flows to the large skeletal muscles in fear, which would be consistent with the hypothesis that fleeing would be the adaptive action in fear (ibid). Thus as well as providing the agent with access to a reason to escape, the emotion triggers the body to be ready to escape: it furnishes us with some of the means by which we can fulfil the desire that arises. The action occurs automatically, without conscious deliberation, in response to specific stimuli (LeDoux, 1996, p.
175), unless the information does not conform to the pre-programmed criterion, in which case a more elaborated deliberation can occur (Hanoch & Vitouch, 2004, p. 440). Thus System 2 will step in if and only if the stimulus doesn’t trigger the automatic behaviour. For example, say I am in a museum and come across a stuffed tiger in a pouncing stance. I might feel fear, which would entail ascribing to the stuffed tiger the emotion-proper property of being dangerous. However, I don’t run away screaming; presumably because the awareness of being in a museum, and there being a plate of glass between me and the tiger, for example, entail that what I perceive does not conform perfectly to the pre-programmed criterion. System 2 can step in and I can decide to stay still despite feeling afraid, as I can reason that the tiger doesn’t really have the emotion-proper property of being dangerous. My emotion provided a prime facie reason to run away, but this was rejected as being sufficient to justify acting on the desire by System 2.

In the example above, I was able to decide not to run away from the tiger. If I came across a live tiger in the wild, I would not be able to make this decision. This is because the perception of the live tiger would trigger the automatic fleeing reaction before System 2 could step in; I would not have the moment of hesitation in which I could decide whether or not to run away. As I argued in Chapter 2, this lack of conscious deliberation does not undermine the claim that I would be acting for reasons. An agent does not need to be consciously aware of her reasons in order to be reason-responding; she is implicitly guiding her behaviour in line with her reasons, as if she were to deliberate she would endorse the course of action. If System 2 took the prima facie reason to not be a reason to run away, as with the case of the stuffed tiger, it would overrule the System 1 deliverance and I would not run away. That I would run away from the live tiger, and not from the stuffed tiger, indicates that my reason-tracking mechanisms are functioning as they should do, in that they are prompting me to run away from perceived danger when and only when there really is a reason to run away. An animal would not be able to do this. My cat continues to react to the appearance of the vacuum cleaner as if it were dangerous – hissing, fluffed up tail and so forth. He might learn in time, through repeated conditioning, to react differently, but he cannot recognize
that he feels fear and reason that the machine does not in fact pose a threat to his life, and decide to not behave as if it did.

In this way emotions enable us to respond to reasons and thus be practically rational: they draw our attention to features of the environment which we might otherwise have missed, or failed to realize were indeed reasons for action in time for us to be able to act for them. Without fear, we would not, like Zajonc’s rabbit, have the opportunity to respond to the reason ‘it is dangerous’, and thus one role that emotions play in practical rationality is enabling us to respond to reasons that we otherwise would not have a chance to, in ways that are effective at fulfilling the desires to which they give rise. Another way in which emotions furnish us with reasons is by providing us with access to concepts that we need in order to be able to respond to associated reasons in the future.

4 Emotions provide us with access to concepts

One could, it seems, respond to the reason ‘it is dangerous’ without an occurrent experience of fear in certain situations. For example, when choosing a route for a walk I might decide against one which involves going through a field which contains a bull. I can make this decision for the reason that the bull is dangerous without having to have an occurrent experience of fear. It does seem, though, that in order to be able to act for the reason ‘it is dangerous’ without an occurrent feeling of fear, that I would have to know what danger is; that I would need to possess the concept of danger. Although I do not have space to argue that this is the case, it is plausible that we may need to have had, at some point, an occurrent feeling of fear in order to know what danger is, analogous to the thesis that in order to possess the concept of red one must have had a visual experience of redness (Peacocke, 1984, p. 365). If this were so it would mean that in order to have access to the reason ‘X is dangerous’ then at some point, either now or earlier, I must have experienced the emotion of fear, whether towards a bull, or towards something else. Possessing the concept of danger, I would then act for
the reason that something is dangerous, even if I had never come across that thing before. For example, I have never seen a live lion, yet having been told that lions are dangerous, and possessing the concept of danger from past experiences of fear of other dangerous things, I would decide against walking up to a lion to stroke it, for the reason that the lion is dangerous, even if I was not having an occurrent feeling of fear towards it. Thus as well as emotions providing necessary quick access to reasons, they may also furnish us with concepts that we need in order to be able to respond to certain reasons without the relevant emotion. The third way in which they promote our practical rationality by providing access to our reasons is by providing the sole path of access to certain of our reasons, to which I shall now turn.

5 Emotions as the only path of access to particular reasons

The emotion of love is a possible candidate for an emotion that is, at least sometimes, the sole path of access to a normative reason for action. By the emotion ‘love’ I refer to the occurrent emotion, or ‘surge’, of love, rather than to the sentiment of love, the latter being often characterized as a dispositional state to feel a characteristic and diverse range of occurrent emotions, including but not restricted to, surges of love (Goldie, 2010, p. 63), such as pride at the loved one’s achievements, and jealousy towards potential rivals (people or other things) for their attention. However, the two are obviously related, in that one is likely to develop the sentiment of love towards one to whom one feels the emotion of love, and those towards whom one has the sentiment of love will be, for the most part, 63

---

63This is an important distinction, as it seems to me that much of the controversy about the nature of love is about the nature of love as a sentiment, rather than as an occurrent emotion. ‘Love’ may usually be used to refer to the sentiment, but this does not mean that it does not also refer to an occurrent emotion. As an occurrent emotion it has been largely ignored in both the philosophical and psychological literature, perhaps, as I suspect, because of the very features which, I shall go on to argue, make it the most likely candidate for an emotion that provides unique access to reasons: it has been presumed to be an irrational, or arational emotion, as not occurring for reasons. For argument that there is in fact such a thing as a distinct occurrent emotion ‘love’, despite its omission from most contemporary literature in both philosophy and psychology, see Shaver et al. (Shaver, Morgan, & Wu, 1996), which argues that not only is love an emotion and not just a sentiment, but that it should in fact be counted amongst the ‘basic’ emotions as it meets the relevant criteria for such (ibid, p.84). Richard Lazarus makes a similar argument (Lazarus, 1991).
those who are the objects of one’s emotional episodes of love.\textsuperscript{64} This does not necessarily make love substantially different to other emotions, as it may appear to do at first glance: it is perfectly possible to have analogous sentiments of other emotions, such as to be disposed to hate a particular person. Having the sentiment of hate towards another would, as in the case of the sentiment of love, be a dispositional state to feel a characteristic and diverse range of emotions, such as glee at the other’s misfortunes.

What is the reason to which love gives us access? The emotion-proper property is something like ‘being loveable’, but crucially, this must mean something like ‘being loveable \textit{by me}', for given that every person is loved by some other person (and every – or nearly every domesticated pet, and all of the other animate and inanimate objects, and concepts and other things which form the class of objects of human love), without this qualification every person (and pet, and so forth) would have the emotion-proper property, and so one should feel the emotion of love towards every person (and pet), and we do not want to say that a person is irrational if she does not feel the emotion of love towards every potential candidate for the object of a human’s love. I do not think it is crucial to my account to spell out precisely in what the property consists, but a plausible approach is David Velleman’s (1999) notion that it consists in the property of being \textit{able} to be loved by me, rather than in being \textit{worthy} of being loved by me. Were the latter true, the properties which made it the case that something or someone was worthy of my love would also make it the case that another who substantiated those properties would be equally worthy of my love, and yet although I may “know that people whom we do not happen to love may be just as eligible for love” I do not love them (\textit{ibid}, p. 373). Love is \textit{de re} for, as Harry Frankfurt puts it, a lover “cannot coherently consider some other individual to be an adequate substitute for his beloved, regardless of how similar that individual may be to the

\textsuperscript{64} Talking of ‘the’ sentiment of love might be misleading, for there is good reason to believe that there are different ‘types’ of love; different dispositional states giving rise to particular clusters of feelings, thoughts and actions. For example, Ronnie de Sousa distinguishes lust, limerance (what we might call romantic love) and long-term attachment, which differ in their duration, their targets (with limerance’s being exclusive), their characteristic hormonal influences and the dispositional thoughts, feelings and behaviours to which they give rise (2011, especially Chapter 16; 1987).
one he loves” (2004, pp. 79-80): as Peter Goldie says, when one is in love, “a doppelganger will not do” (2010, p. 66).

But we look to be in danger of an accusation of circularity here: if the reason for my emotion of love is the property of being able to be loved by me, arguably this property would be acquired by the object of my love simply by virtue of the fact that I happen to have the emotion towards that object. We don’t want to say, ‘love me for a reason, let the reason be love!’ Such would suggest that the emotion of love was more like an urge, like that to urinate: reason-providing rather than reasoning following, or in some similar way arational, and not for reasons independent of the occurrence of the emotion. But, clearly, love is for reasons, since it can be rational or irrational (Solomon, 2003, p. 56ff). ‘Being loveable by me’ cannot be the reason for love, for this would be circular. There must be independent reasons. The intuition that love is not love if it is for independent reasons is perhaps born out of the realization that not just any sort of reasons will do. The sorts of reasons that Mr Collins gives to Elizabeth in Jane Austen’s Pride and Prejudice, for example, such as having appropriate family connections, are the wrong sort of reasons (ibid), for they are not reasons to love the person.

What sort of reasons are the right sort of reasons; those that are reasons to love the person, de re? There is clearly a difficult in pinning them down. Goldie brings out the problem with trying to do so by comparing the reasons for love with the reasons for fear:

Fear is an emotion that one ought to have towards an object in virtue of its having certain determinate fearsome properties; whereas love is not an emotion that one ought to have towards an object in virtue of its having certain determinate loveable properties (2010, p. 65).

Goldie’s intuition is that fear and love are therefore different in their direction, in that the object of our love acquires the property of being loveable by virtue of our

---

65 As did the ‘boy band’ Boyzone in the 90s, covering The Osmonds of the 70s. As Bob Solomon says, “It is a bit embarrassing to admit that virtually everything I will have to say here has been anticipated by one or another Country and Western song and in love poetry both great and awful, but such is the nature of love in our society” (2003, p. 51).
having the emotion of love towards it, but I do not think we need accept this. Just because the reasons for love are not accessible to the intellect, does not necessarily mean that they do not exist independently of our emotion. Rather, I suggest that the fact that these properties are not intellectually accessible is precisely what makes love a plausible candidate for an emotion that provides unique access to certain reasons; the reasons for love.

For what sort of reasons do we love, then? Solomon argues that they are not what he calls properties-of-the-beloved-type reasons, but Aristophanic reasons which are those to do with the relationship one has to the beloved: the concept that one has of one’s self as being intertwined with the other (2003, pp. 58-60). Velleman argues that, “what we respond to, in loving people, is their capacity to love” (1999, p. 365), “a capacity for valuation like ours, which can be constrained by respect for ours, and which therefore makes our emotional defences against them feel unnecessary” (ibid, p. 366). The reason we love some people rather than others, according to Velleman, is that we can only ‘see’ some people’s value as a person, or personhood, for “the human body and human behaviour are imperfect expressions of personhood, and we are imperfect interpreters” (ibid, 372). How do we ‘see’, or perceive such? I suggest that it is exclusively via the emotion of love. Although we may ‘know’ that every person may be ‘worthy’ of love, we can only come to access the reason that makes them so via ‘perceiving’ it with the emotion of love. The only path of access to the emotion-proper property of love is the emotion of love.

A potential objection to my claim is that we can act for the reason of love without having the occurrent emotion of love. Yes, perhaps we might at some point have needed to experience the emotion of love in order to act for the reason of love, but we also need to have at some point experienced the emotion of fear in order to act for the reason of fear (i.e. ‘is dangerous’), so in what way is love different? Well, firstly we must note that in order to act for the reason ‘x is loveable’ one

66 Solomon calls reasons for love Aristophanic after Aristophanes’ speech in Plato’s Symposium (1953, pp. 189c-193d) on the source of love as stemming from a desire to be once again fused with she whom one was once ‘one’ with as a double-faced prehistoric creature split apart by Zeus.
needs to have experienced the emotion of love towards x, whereas to act for the reason ‘x is dangerous’, one needs not to have experienced x as the object of one’s fear: one can abstract from the experience of having felt fear towards one bull that all bulls are dangerous, and beyond the concept of bulls that other things are also dangerous. Not so for love. As I have already argued, intellectually one ‘knows’ that all persons are worthy of love, but not that they are able to be loved by one’s self: that knowledge can only be acquired through the experience of having that emotion towards that particular object. Even so, the objection might go, one needs to have had the experience of the emotion of love towards that particular object once in order to know that that object is one whom one is able to love, but we often act for the reason that we love someone without experiencing the surge of love: we do so all the time for those to whom we have the sentiment of love.

I have two responses to this objection. The first is to insist that at the very least we must have a regular reminder of the reason, by having the experience of the emotional episode, in order to continue to appraise the object of our love as ‘loveable’, or else we will cease to appraise her as such and no longer have access to the reason. Of course we may continue to perform the same sort of actions as we used to when we did have surges of love towards her, but we will be performing those actions for different reasons, such as duty. We would behaving ‘as if’ we had loving feelings towards her, although we do not. We might visit a relative in hospital because we love them, or we might visit a relative in hospital about whom, were we to be asked if we loved them, we would answer in the affirmative, despite not having felt any surges of love towards them for decades. In the latter case, although we behave as if we had the emotion of love towards them, our real motive is duty.

My second response is to suggest that if we are not performing an action because of an episodic experience of love, then we might be performing it for the emotion-proper property of a different emotion; one associated with the sentiment of love, such as compassion, but not that of the emotion of love. The relative we are
visiting in hospital might be one to whom we truly do have the sentiment of love rather than duty; to who we do still sometimes feel surges of love towards, but on hearing that they are in hospital, it is a surge of compassion which motivates us to visit them, rather than a surge of love. Under one description it would be perfectly accurate to say that we are visiting them ‘because we love them, but under this description ‘love’ refers to the sentiment, which includes the disposition to feel compassion more readily towards them. We are not visiting them because of our recognition of their being someone whom we are able to love: we are visiting them because they have the emotion-proper property of being in some sort of trouble, though the sentiment of love may make us more likely to have the emotion of compassion towards them and so be disposed to respond to them.

Thus, one of the roles that emotions play in practical rationality is in providing access to practical reasons. They do this in at least three ways that provide a positive contribution to practical rationality that is not provided by deliberation. Firstly, they provide access more quickly than can conscious deliberation, and thereby enable us to act in time when deliberation would not. Secondly, they are (plausibly) necessary in order to provide us with the concepts that are the emotion-proper properties of our emotions. Thirdly, they are sometimes the only access we have to particular reasons. In these ways emotions contribute positively to practical rationality. This runs contrary to the view of emotions as interfering with practical rationality, as does my argument in the next chapter; emotional capacities are utilized in practical reasoning, and practical reasoning would be worse off, not better off, without them.
Chapter Six

The Roles of Emotion in Practical Reasoning

I have argued that an agent is practically rational to the extent that she guides her behaviour in line with the norm of doing what she has most accessible normative reason to do by responding to her normative reasons for action. I argued that the capacity for practical reasoning is vital to her ability to properly respond to her reasons and thus to her ability to guide her behaviour in line with this norm, such that if she needed to engage in practical reasoning in order to do what she has most reason to, then she would do. Thus practical reasoning is at the core of practical rationality as a regulative capacity that is active even when not occurrent. In Chapter 3 I argued for a picture of practical reasoning which has reason and desire working together in order to guide an agent’s actions towards the goal of practical rationality. In this chapter I argue that the above picture of practical reasoning makes room for the significant body of evidence that our emotions contribute positively to our ability to be practically rational and, further, that it can be argued that emotions are a fundamental part of the reasoning process; emotions play a number of roles within practical reasoning that we could not do without. Thus, as well as providing us with access to our normative reasons for action, thereby playing a fundamental role in our ability to act for reasons, and thus be practically rational (a view for which I argued in the previous chapter), emotions are also part of the system by which agents maintain regulative guidance control over their behaviour: practical reasoning. The main claim I make is that practical reasoning is not ‘better off’ without emotion. Emotions are employed in the reasoning process, and possibly have to be so employed in order for us to reach decisions. Thus the emotions, far from being disruptive influences that either enslave or must submit to reason, are part and parcel of human reasoning and human rationality.

In this chapter I argue that there is substantial evidence to support the argument that emotions play a fundamental role at three stages of the process of practical
reasoning. In the first section I argue that, at least in some cases, and in a wider range of cases than might first appear obvious, emotions instigate practical reasoning; that is, without emotions agents would, at least in some cases, fail to engage in practical reasoning, when doing so is needed in order for the agent to guide her behaviour in line with the norm of practical rationality. In the second section I argue that once practical reasoning has commenced, emotional capacities are employed in the process of reasoning and enable the agent to reach the right decision: they enable her to correctly judge what she has most accessible normative reason to do. In the third section I propose that emotions are also responsible for bringing practical reasoning to a close, and thus assist practical rationality by enabling an agent to cease reasoning and commence action.

1 Emotions instigate practical reasoning

It is presumably uncontroversial to claim that we sometimes enter into reasoning about what we ought to do because of some emotion or other. A feeling of guilt, for example, can prompt reasoning about whether we really ought to go ahead with something. We may have emotions which give rise to competing desires, thereby prompting reasoning about which desire we ought to fulfil. We also engage in reasoning about which course of action would best satisfy a desire that arises from an emotion. I claim that the range of cases in which emotions instigate reasoning is wider than that which contains only the obvious cases above.

The ‘obvious’ cases of the preceding paragraph are those in which the agent is, to some degree at least, aware that they are experiencing an emotion, or taking into account an emotion, at the time of reasoning. It seems relatively uncontroversial to claim that there are cases in which we are in an emotional state but are not aware of it at the time. William Lyons gives the example of an academic, O’Reilly, who,

is so taken up by the discussion at the curriculum meeting that he does not realise that he is becoming very angry with Macdonald who is
suggesting that the central texts in the first year should consist only of the writings of the Existentialists. It is only later on, when Macdonald curtly remarks to O’Reilly that there was no need to get so heated, and he overhears MacFee wonder why he got so angry, that O’Reilly realises that he must have become very angry during the meeting (Lyons, 1980, p. 6).

Recall from Chapter 4 that a common feature of emotions is that they involve feelings or bodily sensations, and that these feelings are caused by physiological changes in our autonomic and motor functions. Cases such as the above suggest that these physiological changes need not be ‘felt’ or attended to in order to have the effects on our automatic and motor functions. This may be because the subject is not paying sufficient attention to her bodily sensations, and thus fails to notice them, as is plausibly the interpretation of the case of O’Reilly (Leighton, 1986), or because some of the physiological changes associated with emotions are such that, at least in certain circumstances, we cannot attend to them.

There is a plethora of evidence from cognitive psychology to support a hypothesis that the physiological changes can be, and often are, non-conscious, such that they occur below the level of conscious awareness, and so can affect our nervous system, thinking and behaviour. Evidence from neuroscience suggests that such basic affective reactions are mediated largely by brain structures deep below the cortex, such as the amygdala (LeDoux, 1996). These subcortical structures are thought to have evolved early, and to carry out limited operations that are essentially preconscious, and associated with System 1 processing. The cortex is thought to have evolved much later, and is involved in consciousness.

67 For example, subjects in a gambling task involving choosing playing cards from different decks generated anticipatory skin conductive responses (which are a measure of emotional arousal) whenever they pondered a choice that turned out to be risky, even though they indicated that they had no awareness that the choice was risky – the participants reported, in the researchers’ words, that they “did not have a clue about what was going on” (Bechara, et al., 1997, p. 1293). This ‘pre-hunch’ phase progressed into expressing a ‘hunch’ for which decks were riskier, and for most participants into a ‘conceptual’ stage where they expressed knowledge about why the riskier decks of cards were disadvantageous. The participants who did not reach this conceptual stage still chose advantageously, even though they did not consciously ‘know’ that the decks they were choosing were advantageous. The researchers concluded that the participants were experiencing emotional feelings of which they were not consciously aware in response to the outcomes of their initial choices. In another experiment, participants were unable to report a conscious feeling at the same time as their behaviour indicated the presence of an emotional reaction which the experimenters had elicited through the subliminal use of happy, sad and angry faces (Winkielman & Berridge, 2004).
and System 2 processing. Because basic affective reactions originate in the subcortical reasons, below the cortex, it is possible that they occur below the level of consciousness, and so are unavailable for conscious awareness (Winkielman & Berridge, 2004, p. 122).

Current research is limited to very basic affective reactions, so we cannot draw the conclusion that full-blown emotions can also be non-conscious on the basis of such research. The majority of the studies are tracking simply positive-negative valence of feelings, though evidence from neuroscience suggests that there may be some qualitative differentiation between different emotions. For example, neuroimaging studies reveal differential activation of the amygdala in response to consciously presented facial expressions of fear versus anger (Whalen, 1998). We may insist that to constitute full-blown emotions such feelings need to become conscious or at least available to consciousness. Still, these basic constituents of emotions are very plausibly responsible for instigating practical reasoning in cases where we are not consciously aware of them, thus the range of cases in which reasoning is instigated by at least constituents of emotion, would be wider than that which is obvious, supporting the claim that emotional capacities are part of human practical reasoning, and thus human practical rationality.

As I have already explained, on Dual Processing accounts, conscious deliberation is instigated when System 2 ‘detects’ that there may be some sort of error in the deliverances of System 1, for example “when an event is detected that violates the model of the world that System 1 maintains” (Kahneman, 2011, p. 24), or when a task is disfluent (Alter, et al., 2007). The ‘cues’ that are detected by System 2 are generated by System 1, and System 2 processes, including practical reasoning, are activated only when needed. The sorts of situations in which System 2 would be needed are the sorts of situations likely to generate, if not full-blown negatively-valenced emotions, at least negatively-valenced affective feelings, such as lack of satisfaction with the response of System 1, or detection of conflict between its deliverances. These sorts of situations will produce what is referred to in psychology as ‘cognitive dissonance’; when an agent is aware that two or more things are not consistent with each other, she will
seek, in a variety of ways, to make them more consistent (Festinger, 1962). If an event violates the model of the world that System 1 maintains, there is an obvious inconsistency between the perception of the event and the previously-endorsed model; if an agent has no obvious course of action when a course of action is called for, there is an inconsistency between the way the world appears to be and the way she desires it to be; if the deliverances of one sub-system of System 1 conflict with the deliverances of another sub-system, there is an inconsistency between these deliverances. Cognitive dissonance is held to be psychologically uncomfortable (Elliot & Devine, 1994), thus is likely to produce negatively-valenced affective states, whether conscious or not. It is also held to be a motivating state (Festinger, 1962, p. 93), producing arousal (Zanna & Cooper, 1974) to motivate the individual “to seek and implement a strategy to alleviate this aversive state” (Elliot & Devine, 1994, p. 382).

Cognitive dissonance looks to be the sort of thing that would produce negatively-valenced feelings, and also the sort of thing that would prompt reasoning, since reasoning would involve seeking a strategy to solve the problem, for example deciding which of the two deliverances to accept, or which course of action is actually better. Psychologist Gier Overskeid argues that all problem solving is motivated by feelings, since an agent is only in a problem state if that state feels aversive, thus problem solving always begins with an aversive feeling, conscious or not (Overskeid, 2000) (though we might not want to call such non-conscious states ‘feelings’, but perhaps ‘affective reactions’). Overskeid defines a problem as when “a situation is in a given state, the problem solver wants the situation to be a goal state, and there is no obvious way of transforming the given state into the goal state” (p. 285).68 “When a task can be performed without any prior testing of possible solutions, the task is nonproblem” (ibid.), thus when System 1 turns up a solution that is an obvious way for the agent to transform the state of affairs into her goal state, no aversive feeling (or affective reaction) is experienced and no problem solving instigated. It could be that we have problem-solving strategies other than engaging in practical reasoning, but certainly practical reasoning is

68 Cf. Richard Wollheim’s theory of the characteristic history of an emotion as arising in response to either the satisfaction or frustration of a desire (Wollheim, 1999, pp. 15-16).
one way, and the obvious way, in which we attempt to solve the problem of how to achieve our goals. If, as Overskeid argues, an aversive feeling (or affective reaction) towards one's present state (which includes situations in which a mentally represented state becomes a goal by eliciting more pleasant feelings than those elicited by attending to the present state) with no obvious way of changing that present state is always the motivation for problem solving, it is plausible that all practical reasoning is instigated by at least the basic components of emotions; valenced affective states.

I am not claiming that the range of cases in which emotions, if just the basic affective feelings, instigate practical reasoning, is necessarily so wide as to encompass all cases of practical reasoning, but the current evidence from cognitive psychology and neuroscience certainly does not rule such a claim out. There is evidence that the demand for conscious intervention may be evaluated in part by the monitoring for conflicts in information processing (which would produce cognitive dissonance) by the anterior cingulate cortex, an area of the brain which also responds to the occurrence of emotional conflict (Botvinick, et al., 2001). This supports a hypothesis that at least in some cases the ‘cues’ that are detected for the need for practical reasoning are emotional. There is also evidence that the ‘shift’ to employing logic in order to come to a decision depends the ventromedial prefrontal cortex and anterior cingulate (Houdé, et al., 2001), areas of the brain thought to play important roles in both decision making and emotion (Bechara, et al., 2000) (Rogers, et al., 2004) (Jackson, et al., 2006). There is therefore reason to claim that one of the positive contributions emotional capacities play in practical decision making is, at least in part, in some cases, in prompting practical reasoning. And since a fundamental part of practical agency is engaging in practical reasoning when it is required, emotional capacities thus enable an agent to be practically rational.

2 Emotions enable the agent to reach the right decision
I have just argued that the range of cases in which emotional capacities are responsible for the instigation of practical reasoning is wider than that which encompasses that of obvious cases, and that the evidence from cognitive science does not rule out the possibility that the range is actually so wide as to encompass all cases. I am now going to make the same claims about the role of emotion in the ‘rounds’ of practical reasoning: the range of cases in which emotion plays a determining role in the rounds of reasoning is much wider than that which is at first obvious, and that there is reason to hypothesise that emotional capacities are involved in every case.

As I argued in Chapter 3, when we reason we step back not to the position of the impartial ‘Kantian Captain’, but to a position from which we assess our reasons from the multitude of our desires and concerns. Our reasons for action are ‘picked out’ by our desires and concerns in that, to repeat an example from that chapter, being able to finish my paper is a normative reason I have that counts against going to yoga only if I value finishing my paper. I also argued that our desires and concerns are also necessary in order for us to assign weights to our reasons, such that from an entirely impartial position, devoid of any desires or concerns, no effective reasoning could take place.

Reasoning often occurs in “rounds”. The agent ascertains what reasons she has that count in favour of a particular action, and what reasons she has against. After doing so, if it is not clear whether she ought or ought not to perform that action, then she will need to examine those reasons in order to ‘weigh’ them up. As I argued in Chapter 3, to do so she will need to assess her reasons against each other, in the light of her various desires and concerns. She will need to ask, “Is this more important than/consistent with/to be ignored compared with/that?” (Blackburn, 1998, p. 255). After she has examined her reasons like this, some of them may be discarded as being now obviously not strong enough in the face of certain considerations. This may happen to reasons on both sides. The agent now has a modified list of reasons for and against the proposed action. She may have added to, as well as subtracted from, the list, as in pondering her reasons she may have discovered previously-overlooked ones, for example further
consequences which were not immediately obvious. It now might be clear whether or not she ought to perform the proposed action. If it is not, then she will need to engage in another ‘round’ of reasoning, assessing this new combination of reasons. At some point a decision will need to be made: the agent will come to an all-things-considered judgement about what she has most reason to do.

There are some cases in which it seems uncontroversial to claim that emotions play a part in the round of decision making. For one thing, as we saw in Chapter 5, in some cases emotions can themselves be reasons for action, as in the case of Maeve turning on the light when she is frightened from a dream. Being frightened in this case is a reason to turn on the light, and it might be a reason that needs to be weighed up against reasons to not turn on the light, for example waking up someone else. The other role for emotions, of providing access to reasons (discussed in Chapter 5), is also relevant to the rounds of reasoning: emotions may be generated when we are thinking about our various reasons, and these may provide access to reasons which were before overlooked. For example, in deciding whether or not to become romantically involved with a particular person, I might imagine the consequences of doing so, one of which is enjoying being with that person in a friendship group. I might then feel an emotion that gives me access to further reasons for or against, for example a feeling of guilt when thinking about being in that friendship group might alert me to something previously overlooked, such as that someone else in the friendship group once had romantic feelings towards this person. In thinking about the potential outcomes of our actions or inaction, emotions can be generated which provide access to further reasons. Another consequence of such might be that certain *prima facie* reasons are then discarded before the next ‘round’ of reasoning. For example, the reason that I originally thought counted in favour of my becoming romantically involved with the particular person, that of enjoying spending time as a couple with her in a certain friendship group, might then be eliminated as something that counts in favour of becoming involved, as I realize that, given the previously-overlooked fact that it might upset someone, I would not actually want to spend time as a couple with her in that group. Thus there is
certainly at least a range of cases where emotions would play a part in the rounds of reasoning.

However, just as I argued that the range of cases in which emotion is involved in the instigation of reasoning is wider than that which contains just those obvious cases in which the agent has conscious awareness of her emotion, so too I will now argue that the same is true for the rounds of reasoning: that it is not just in those obvious cases where we have conscious access to our emotions that emotions are at work in the rounds of reasoning. And, as in the argument regarding instigation of deliberation, the evidence that gives reason to hold that emotion can play a role in a wider range of cases does not rule out the possibility that it plays a role in all cases. There are at least three ways in which emotions play a role in non-obvious cases: they rule out some alternative courses of action from the beginning, they determine whether something is a reason that an agent has for or against a particular action, and they determine the normative force or ‘weight’ of an agent’s various reasons for action.

2.1 Emotions rule out some options from the beginning

In section 1 I argued that evidence from the cognitive sciences suggests that basic affective reactions can be non-conscious. Such evidence supports argument that these basic constituents of emotions play a role in all of the rounds of practical reasoning, even if they do not develop in to consciously-accessible emotions. In the first place, there is reason to believe that emotions play an important role in narrowing the range of possible actions that an agent even considers in the first place: emotions rule out some possibilities before ‘thoughts’ of such possibilities can even become conscious.

It is advantageous for some possible courses of action to be ruled out \textit{a priori}, for there is a potentially infinite number of options available as potential solutions to a problem, and, as Ronnie de Sousa argues, “unless some drastic preselection can be effected among the alternatives their evaluation could never be
completed” (1994, p. 276). Even if we were capable of deducing all possible implications of an alternative course of action, “we need to know whether a consequence will turn out to be relevant before drawing it”, for if it is relevant and we have not considered it we may fail to do that what we have most reason to do, but “if it is irrelevant and we have already drawn it, we have already wasted time” (de Sousa, 1987, p. 194). When we are reasoning about whether or not we ought to Φ, the decision has to be made within a given time frame: there is no point in coming to the all-things-considered judgement that one ought to Φ once the opportunity to Φ, or the benefit of Φ-ing, has passed. Human practical rationality is as a matter of contingent fact ‘bounded’: human agents have to make decisions under the constraints of having limited, possibly unreliable, information regarding the possible alternative courses of action and their consequences (since we are not omniscient), with a mind that has a limited capacity to evaluate and process the information that is available, and within a limited amount of time (even if the decision remained open for an entire lifetime, that agent’s lifetime is of course finite).69

Relative to theoretical decision, time restrictions on practical decisions increase the effects of the first two constraints: we can ‘withhold belief’ on a number of theoretical issues, awaiting further evidence, and it is perhaps less important in most cases that we ‘get things right’ on theoretical issues the first time we reason about them. We might feel foolish if we realise in later life that a belief we formed in our youth turns out to be irrational, but leaving aside that we may have acted on the assumption that the belief was true, it is not too late to change our minds once we realize this mistake. We do not have this comfort when making practical decisions. And given limited time in which to make a decision, we will have a further-reduced amount of information which we can take into consideration, as finding out further information would take more time, and a further-reduced capacity to evaluate and process the information that we do have, since such

69 These three constraints comprise the theory of ‘Bounded Rationality’ (Simon, 1982), which argues that human rationality operates within limitations and is an alternative model of decision making to mathematical models employed in economic theory. It is a prominent theory, influencing economics, political science, psychology and artificial intelligence. The theory of Bounded Rationality is a normative one. I am not arguing that these constraints are normative, but rather descriptive of the way in which human agents, as a matter of contingent fact, make decisions.
processing and evaluation also takes time. Pre-selection amongst alternatives aids our being able to come to the decision in time, thus pre-selection aids practical rationality even more than theoretical.

Emotions play a role in this pre-selection of alternatives, and thus aid practical rationality, by, as I argued in Chapter 5, reducing the amount of information to which an organism pays attention, and focussing that attention on the emotionally salient factors, which, when all is going well, will be the relevant information. That an assailant has a knife is relevant to the question of what one ought to do, that the assailant has blue eyes, is not, thus a victim’s focus will be on the knife and not the eye colour (and thus what she will remember is the knife rather than the features of the assailant, since at the moment of attack her decision is one of how to escape, not of how to identify the villain). Emotions thus solve what is known as the ‘frame problem’ (de Sousa, 1994, p. 276), or the ‘search problem’ (Evans, 2004, p. 183): how to focus attention on relevant information and keep this amount of information small enough so that the mind can actually perform the necessary computations upon it.

By restricting the information to which an agent can pay attention, most logically-possible courses of action will not even occur to the agent. For example, if when walking in the forest Jing-Jing stumbles upon some bear-cubs and thus angers their mother, there are an infinite range of potential courses of action which are available to her. She could attempt to foxtrot with it, sing to it, play with its cubs and so forth – none of which would be advisable, and most of which would not even occur to her. Jing-Jing has read that running away from a bear is not advisable as bears are likely to chase in such circumstances and will almost certainly catch up with a human, but she can’t remember what action is advisable. Options such as attempting a foxtrot do not even occur to her, despite being possible, and a good thing too, since she does not have time to ponder the potential consequences of such actions in order to draw the conclusion that they are unlikely to be successful. The reason these things do not occur to her, I propose, is that emotionally-salient information ‘floods’ her consciousness before non-emotionally-salient thoughts can arise (as I said in Chapter 5, System 1 acts
more quickly than can System 2, thus the deliverances of System 1, which include information about emotional-proper properties, occur before the deliverances of System 2, which include non-emotionally salient information), and because human minds have a limited capacity to process information, there is no ‘room’ for the non-emotional information, nor can practical reasoning be engaged in attempting to discover such, as the mind’s reasoning capacities are engaged in dealing with the emotionally-salient information which has just been presented. Emotion thus aids practical rationality by making salient what is likely to be (from an evolutionarily adaptive perspective at least) the most relevant features of the agent’s situation, and delivering this into consciousness quickly, which results in non-emotionally salient possibilities, which are likely to be less-relevant, not even being considered, at least in the first ‘round’ of reasoning. If this first round does not produce a satisfactory result, more information can be considered. This information is not that which is immediately emotionally-salient but, as I will argue in the next sub-section, emotion is employed in assessing the relevance of this information in the subsequent rounds of reasoning.

In this section I have argued that emotions rule out some potential courses of action before they can consciously occur to the agent. The examples I have used have been what would be classed as ‘obvious’ cases, those in which the agent is obviously in an emotional state, of which she is likely to be consciously aware. I propose that if emotions solve the frame problem in the way I have argued above, and if it is true that emotions, or at least basic affective reactions, can be non-conscious, then there is good reason to suppose that emotions play this role in a wider range of cases than those that are obvious, i.e. that non-conscious affective reactions can also solve the frame problem. In every case in which an agent reasons practically the frame problem will arise, not just in those cases which obviously involve emotion. The pre-selection of relevant information occurs non-consciously, and if emotion, or at least the basic constituents of emotions, can occur below the level of consciousness, they could pick out the relevant features of an agent’s situation to be ‘delivered’ into consciousness. The problem here is that to do this, these non-conscious emotions would have to have intentionality and formal objects in the same way as fully-blown emotions do, and there is, at
present, a lack of evidence from the cognitive sciences to support such. As I said earlier in the chapter, the focus of research in neuroscience and cognitive psychology has been on basic valenced reactions, and so hypotheses about the role of non-conscious emotion in human reasoning that are based on such, such as those of Damásio (1994) and Mameli (2004) which are discussed below, may be assuming an overly-simplistic account of what emotions are, thus limiting the extent of their claims. However, the scientific evidence does not exclude the possibility that in the future there will be evidence that we can have non-conscious fully-blown emotions, or at least non-conscious states that are closer to such than are basic affective valenced reactions. The discovery of differential activation of the relevant areas of the brain for the different basic emotions is relatively recent, and supports a hypothesis that the distinct phenomenologies of different emotions may not be entirely available to conscious awareness.

It would be a very speculative claim indeed to suggest that emotions are always involved in the stage of practical reasoning discussed in this section; that emotion is always involved in the pre-selection of potential courses of action. Though the claim would be consistent with the Somatic Marker Hypothesis, which I will discuss in the next section, I think it too great a leap to be tenable, as it would effectively transpose the very problem that I’ve argued emotions help solve – that of the set of possible courses of action being in principle infinite and thus unmanageable by the human mind – to the level of non-consciousness. I shall return to this argument at the end of the next section after discussing the relevant neurobiological theories.

2.2 Emotions provide the necessary access to all of the reasons for action which an agent has

The Somatic-Marker Hypothesis, proposed by neuroscientist António Damásio (1994), holds that decision-making is influenced by ‘marker signals’ generated by

70 See also Dylan Evans, who argues that theories that argue that emotions solve the search problem (or frame problem) are “vacuous unless we have some independent account of emotions to flesh [them] out” (Evans, 2004, p. 190).
emotional mechanisms that rapidly provide implicit and explicit knowledge about the prospective consequences of an action. These ‘somatic marker states’ (bodily responses that hallmark an emotion, from the Greek *soma*) influence working memory and help endorse or reject response options that are brought into mind during practical reasoning, at both the conscious and non-conscious level (Bechara & Damásio, 2005, p. 343). In this way, argues the hypothesis, emotion is beneficial to decision making, not just because it enables the endorsement or rejection of reasons that is needed to come to a decision *per se*, but because it helps the agent to come to the *right* decision.\(^71\) Evidence for this claim comes from studies of patients with damage to the ventromedial prefrontal cortex (vmpfc) area of the brain, an area, as mentioned earlier, thought to be involved in both emotion and decision-making. These patients have found to show compromised ability to experience and express emotions (Damásio, 1994, p. 138), and also impairments in decision-making in both real life (*ibid.*, pp. 193-4) and in laboratory studies (Bechara, et al., 1997), in that it takes them a long time to reach decisions and also they often choose disadvantageous options. For example, in the gambling task described in footnote 1 earlier in the chapter, a group comprised of patients with this ventromedial prefrontal cortex damage consistently chose the more risky, disadvantageous options, whereas the non-vmpfc-impaired participants quite quickly learnt to choose the advantageous options. What is more, the non-impaired participants, as I said in the footnote, generated anticipatory skin conductance responses (SCRs), and then ‘hunches’, which appeared to guide their choices to being advantageous before they had conscious access to their learned knowledge of which decks were advantageous, whereas the impaired patients not only had no anticipatory SCRs or hunches, but even when they had conscious knowledge of which decks were advantageous, they *continued* to choose disadvantageously (*ibid.*). The Somatic-Marker Hypothesis proposes that the damage to the emotional mechanisms in vmpfc-

\(^71\) The hypothesis is restricted to decisions where the emotion is relevant to the situation (Bechara & Damásio, 2005, p. 337). The parameters of relevance are determined by whether the emotions are a response to the situation at hand. This hypothesis finds support in studies of the role between emotional arousal and decision-making: if one is *generally* aroused decision-making may not benefitted, and if an agent is in a state of emotional arousal that is unrelated to the decision at hand, decision-making is impaired by that emotional arousal. This is thought to be because the emotion will focus the agent’s attention on the emotionally-salient information, which in these cases is *not* the relevant information to the current decision (Hanoch & Vitouch, 2004).

Page 153 of 224
impaired patients impairs the patients’ decision-making because they do not generate the emotional feelings in response to their situation which would, in non-impaired people, indicate, non-consciously, the advantageous option.

How do emotional feelings indicate the advantageous strategy? They let the agent know ‘what it feels like’ to be in a given situation, though this feeling need not be conscious (Bechara & Damásio, 2005, p. 341), thus signal whether the situation would be advantageous for the agent (indicated by positively-valenced feelings) or disadvantageous (indicated by negatively-valenced feelings). Having a positively-valenced feeling in response to the considered option lets the agent know that the option would be advantageous. The feelings can be “processed unconsciously” (ibid., p. 343) because induced changes in the neurotransmitters that are released can be detected in areas of the brain that are below the level of conscious access. They can play this role in reasoning because they are either innate responses to evolutionarily-relevant stimuli (such as a snake), or else they are learned through the course of experience and they can then be triggered in response to the thought of the potential situation:

When the bad outcome connected with a given response option comes into mind, however fleetingly, you experience an unpleasant gut feeling [which is the somatic marker] which forces attention on the negative outcome to which a given action may lead, and functions as an automatic signal which says: Beware of danger ahead if you choose the option which leads to this outcome. The signal may lead you to reject, immediately, the negative course of action and thus make you choose among other alternatives … When a positive somatic marker is juxtaposed instead, it becomes a beacon of incentive (Damásio, 1994, pp. 173-4).

Somatic markers are learned affective responses in the same way as other responses are learned through the process of classical conditioning; via punishment and reward which build up distinct neural pathways that are then automatically activated when the stimulus, either concretely or as a thought, is re-encountered. They can be reconstituted by either the ‘body-loop’, which is Damásio’s term for the cases in which the physiological changes associated with the original emotional experience are activated and then perceived by the somatosensory cortices (the system in the brain which perceives physical
sensations), or by the ‘as-if-body-loop’, where the thought of the potential situation activates the somatosensory cortices without there being any associated physiological changes. The agent may become conscious of the fact that the sensory cortices have been activated, in which case she will know that she has the feeling, or she might not become conscious of the activation, in which case she will not know that she has the feeling. Whether or not she knows of its existence, according to Damásio, it will move her either towards or away from consideration of the considered outcome as a potential course of action.

If this is the case, then we have the second role for emotions, or at least for emotional feelings (or even just basic affective reactions if we don’t want to follow Damásio in calling somatic markers ‘feelings’), in practical rationality: they determine whether something is a reason that an agent has for a particular action (those potential consequences which have positive somatic markers) and those that she has against the action (those which have negative somatic markers). The neuroscientific literature which argues for the Somatic-Marker Hypothesis appears to be claiming that this role is enacted before a ‘weighing up’ of reasons is performed; that somatic markers play the role that was discussed in the previous section, that of determining viable possibilities, before practical reasoning takes place. If this is indeed the hypothesis, then I think it cannot be right for this would, as I indicated at the end of the last section, be effectively claiming that the frame problem is solved in the non-conscious; that every conceivable action is evaluated non-consciously, via somatic markers. Although this is logically possible, given that the problem is that the number of potential possible courses of action is so vast as to seem impossible to be computated by the conscious mind, it would be an empirically-odd move to claim that the non-conscious mind can actually perform this problematic task, and that it can do so in a fraction of a second. However, I do think it plausible that somatic markers can play this role in the rounds of reasoning, marking consciously considered options and outcomes with a conscious or non-conscious emotional feeling that renders them as being able to move the agent towards or against finding in favour of that particular action; i.e. in determining whether they are a reason that an agent has for or against a particular action. The absence of any feeling (positive
or negative, conscious or non-conscious) connected with consideration of a particular fact means that it is not a reason which the agent ‘has’ at all (it may be a reason, but the having relation is one of ability to move the agent towards or away from the potential action): it is irrelevant to the decision and likely to be overlooked as the agent’s emotional capacities will focus her attention on the information which is relevant (though it may be ‘picked up’ in a later round of reasoning, as I will argue below).

What I am proposing is that emotions, or at least the constituents of emotions feelings, play an even more fundamental role in practical decision-making, and therefore in practical rationality, than the Somatic-Marker Hypothesis claims. Emotional feelings are the means by which an agent has access to reasons, thus without emotional feelings an agent does not have access to any reasons for action about which she can deliberate, since without emotional feelings she would not be moved towards or against favouring a particular action. There is no non-emotional cost-benefit analysis because what makes something a cost or a benefit is determined by how it makes the agent feel, or rather, how she knows that something is a benefit is via the positively-valenced feeling she experiences (consciously or non-consciously) upon contemplation of it, and how she knows that something is a cost, is via the negatively-valenced feeling she experiences upon its contemplation. I am not proposing that what makes something a reason that counts in favour of an action the fact that it makes the agent feel good (which might strike one as morally bankrupt), but that the way in which an agent comes to have access to reasons (which exist as reasons independently of her, though she may not have them) is via valenced affective feelings. Mine is a very different claim, akin to Aristotelian concepts of virtue as involving having the right feelings (Aristotle, 1925, p. 1105b25–6), as opposed to egoism. 72 Neither does my

---

72 Niko Kolodny argues that some virtues are dispositions to respond correctly to certain kinds of reasons, “[k]indness is a disposition to respond to the needs of others, for example, and justice is a disposition to respond to considerations of fairness” (2005, p. 553). An agent needs to have access to reasons in order to be able to respond to them, thus, if an agent must have the requisite affective reaction in order to have access to a reason, then she must have the right emotional capacities to access any reason. Thus if to have the virtue of kindness is to be disposed to respond to the needs of others, one has to have the right affective reactions to the contemplation of others needs in order to respond to that need as a reason. One has to have the right emotional dispositions, therefore, to be able to have that virtue.
proposition entail that that agents are simply ‘blown around’ by their various emotions, for it is not the case that an agent is bound to do whatever she most feels like doing at the moment of choice: practical reasoning enables her to assess her reasons, and in doing so her feelings, their intensities and her consciousness of them will change as reasons are discarded and discovered and weighed up against each other.

If emotional feelings are necessary for human decision-making, how do we account for the fact that vmpfc patients ever reach decisions, if their emotional capacities are impaired? Matteo Mameli, who similarly argues that in humans the choice between different actions is always determined by the emotional feelings generated by the contemplation of possible outcomes, a position he calls a New Somatic-Marker Hypothesis (Mameli, 2004, pp. 170-1), suggests that vmpfc patients lack, not any emotional feelings at the thought of possible outcomes, but specifically emotional outcomes at the thought of non-immediate predicted outcomes (Mameli, 2004, p. 173), thus the predicted non-immediate future does not enter into the cost-benefit analysis. This explanation seems to me to be consistent with the finding in the gambling task, which I described earlier) that vmpfc patients could state that they knew the advantageous strategy yet continued to choose the disadvantageous one: the knowledge could not move them towards the advantageous strategy because successful gambling is, of course, dependent upon being able to preference long-term advantages over the immediate thrill. Mameli suggests that his new hypothesis would account for the behaviour of psychopaths, who know what is morally right, but don’t have the right somatic markers attached to that knowledge, so they experience a positively-valenced effect at the thought of the consequences of murder, which moves them towards it, rather than the negatively valenced feelings which move the rest of us away from it (Mameli, 2004, pp. 171-4). This could explain addictive behaviour: the reinforcing behaviour provides immediate outcomes which are motivationally overwhelming; so much so that the positively-valenced feelings associated with the addictive behaviour are so strong that they render the negatively-valenced feelings triggered in response to the longer-term consequences too weak in comparison. Thus, although the addict is aware of the multitude of reasons
against their drug-seeking behaviour, these reasons fail to move her away from the behaviour, because the feelings generated upon contemplation of them are not strong enough to come down in favour of doing so.\textsuperscript{73} This brings me to the next role that emotions play in practical reasoning: they determine the normative force that the agent attaches to her various reasons for action.

### 2.3 Emotions determine the normative weight an agent assigns to a reason for action

As well as determining whether an agent has access to a reason, I propose that emotions determine the normative force an agent assigns to a reason for action; they determine how much ‘weight’ she decides to give each of her reasons. As I argued in Chapter 3, reasons do not have intrinsic normative force. As the agent moves through the rounds of reasoning, her feelings (conscious and non-conscious) will change. Newly-contemplated reasons will generate new feelings and discarded reasons will alter the balance of valence. The ‘weight’ that she assigns to each reason during the course of her reasoning indicates to her how much she values this in comparison to that. As she contemplates a reason, the intensity of the affective reaction generated will determine how much weight she gives that reason. For example, contemplation of the reason ‘it will promote a good night’s sleep’ I have in favour of going to yoga might, on Monday, generate only a mildly-intense pleasant state, whereas contemplation of it on Tuesday generates a strongly-pleasant state. Thus on Tuesday I assign more weight to the reason ‘it will promote a good night’s sleep’ than I do on Monday, and this might tip the balance in favour of going to yoga on Tuesday, but not on Monday.

As Mameli points out, the claim that emotional feelings, caused by the contemplation of potential outcomes, are necessary for correct deliberation, does not entail that they are sufficient for such: the ability to infer the likely consequences of different actions is also needed (Mameli, 2004, p. 174). Yet

\textsuperscript{73} Addiction is defined as “a primary, chronic disease of brain reward, motivation, memory and related circuitry .... Addiction is characterized by impairment in behavioural control, craving, inability to consistently abstain, and diminished relationships” by the American Society of Addiction Medicine (Duhigg, 2012, p. 69).
there is even reason to suggest that emotion plays a role in the ability to predict possibility (Shackle, 1961), by responding immediately to patterns that either repeat or alternate (Huettel, et al., 2002). Thus as well as assigning normative weight to reasons, emotions may also assign likelihood to their leading to the valued states of affairs by which they are reasons. Thus we have a third role for emotions in practical rationality: as determining the normative force of an agent’s reasons for action. The forth role, for which I shall now argue, is in bringing deliberation to a close, leading to the formation of intention and, ideally, action.

### 3 Emotions Close Reasoning

In section 1 I argued that emotional feelings instigate practical reasoning. Negatively-valenced feelings indicate that there is a problem that needs solving, and positively-valenced feelings on contemplation of an alternative to one’s current state provides incentive to reach that alternative state. If a negatively-valenced feeling is necessary in order to motivate reasoning, then it is plausible that its absence would mean that there was nothing to motivate the agent to reason, and so, unless some other source of motivation had replaced it, the agent would cease reasoning. This would happen if the agent was no longer in the state that generated the negatively-valenced feeling, which could occur for a number of reasons, such as that she had temporarily or permanently put a thought that generated the feeling ‘out of her head’.

The other reason why the agent is no longer motivated to reason is, of course, that she has come to a decision, and thus solved the problem. Emotional feelings play a role here, too. We automatically experience a “pleasure response” when we find a solution to a problem (Bechara & Damásio, 2005, p. 340): finding a satisfactory conclusion to reasoning is marked by a positively-valenced affective feeling. I propose that this feeling signals to the agent that a satisfactory solution has been found: she comes to know that the solution is a good (or good enough) one to her problem via the positively-valenced feeling that the thought of it generates. It is not that what makes the solution good is that it happens to make
her feel good, but that her feeling gives her access to the fact that the solution is a good one. The feeling is a somatic marker that signals that the proposed action would be advantageous. This feeling must be in some way different from those generated at contemplation of alternatives during the reasoning process: were it just that something felt pleasant then the agent would be able to cease reasoning as soon as she came across any option that felt pleasant. If this were the case then she would rarely be practically rational, not considering potentially better alternatives, and not taking into account the negative feelings generated by the reasons she has against this option which first spring to mind. The positive feeling that signals a satisfactory solution must therefore be either qualitatively or quantitatively different to the positive feelings generated during the rounds of reasoning.

Either the balance is tipped far enough in the direction of positively-valenced feelings over negatively-valenced feelings in contemplation of a proposed solution (with the rounds of reasoning adjusting this balance until it reaches the tipping point) and the ‘pleasure response’ is simply an awareness of enough positively-valenced feeling, or else the response is something separate that can ‘triggered’ by the tipping point. The latter is a preferable explanation, since it leaves open the possibility that the “pleasure response” can be triggered by states other than a certain threshold of positively-valenced feelings having been met. Thus even if it is not the case that emotions are necessary for weighting reasons, it remains a possibility that they can signal to the agent that reasoning can cease.

4 Conclusion

In this chapter I argued that there is substantial evidence to support argument that emotions, or at least emotional capacities, play a number of fundamental roles in the stages of practical reasoning. In the first section I argued that, at least in some cases, and in a wider range of cases than might first appear obvious, emotions instigate practical reasoning; that is, without emotions, or at least the
basic constituents of emotions, agents would, at least in some cases, fail to engage in practical reasoning when doing so is needed in order for the agent to guide her behaviour in line with the norm of practical rationality. In the second section I argued that once practical reasoning has commenced, emotional capacities are employed in the process of reasoning and enable the agent to reach the right decision: they enable her to correctly judge what she has most accessible normative reason to do by providing the necessary access to all of her reasons for action, and by determining the normative force that she assigns to each. In the third section I proposed that emotions are also responsible for bringing practical reasoning to a close, and thus assist practical rationality by enabling an agent to cease reasoning and commence action. Thus, on my picture of practical rationality, emotions, far from being the enemy of reason, work together with our capacity for conscious reflection to enable us to be practically rational.
Chapter Seven
Weak-Willed Action is Necessarily Irrational; Akratic Action is Not

In this thesis I have argued that practical rationality consists in doing what one has most normative reason to do, and that an agent is practically rational to the extent that she guides her behaviour in line with the regulative norm of doing what she has most accessible reason to do by responding to her normative reasons for action. I have argued that the capacity for practical reasoning is fundamental to an agent’s ability to so guide her actions, and that, although actions do not have to be the direct result of practical reasoning in order to be so guided, they do need to be under the agent’s active regulative guidance control, such that if she needed to engage in practical reasoning in order to do what she has most reason to do she would do. I further argued for a picture of practical reasoning that, rather than having reason and desire as opposing forces, presents them as working together in order to guide an agent’s actions towards the goal of practical rationality which made room for the overwhelming evidence that our emotions, not just contribute to, but are a fundamental part of our practical rationality. In this chapter I am going to apply my argument thus far to the question of whether or not incontinent action – action which goes against an agent’s action – can be rational. My central claim is that by distinguishing between weakness of will and akrasia we can find a satisfactory answer to the question of whether or not inverse incontinent action can be rational: yes, if it is inverse akrasia; no if it is inverse weakness of will. Inversely akratic actions can be under the agent’s regulative guidance control, and can thus be rational, whereas inversely weak-willed actions are, by definition, not under an agent’s regulative guidance control, and cannot therefore, by definition, be rational.

In the first section I argue that we should draw a distinction between akrasia and weakness of will, and thus depart from the majority of the literature that addresses the question of the rationality of incontinent action which identifies akrasia and
weakness of will as the same phenomenon. In section 2 I take up the example of Mark Twain’s Huckleberry Finn which the relevant philosophical literature maintains is a prime example of potentially rational inversely incontinent action. However, I argue that Huckleberry Finn is inversely weak-willed, rather than inversely akratic, and thus cannot be rational for, as I argue in section 3, inverse akrasia can be rational, but inverse weakness of will cannot. This is because an akratic agent’s action can be responding to her reasons and thus be guiding her action in line with the regulative norm of doing what she has most accessible reason to do, whereas a weak-willed agent’s behaviour is necessarily not responding to her reasons, and thus is not being guided so. The akratic agent’s action can be under her regulative guidance control even though it conflicts with her consciously-held judgement as to what she ought to do, but the weak-willed agent’s behaviour is, by definition of being the result of weakness of will, not under her regulative guidance control. Thus the inversely akratic agent can be practically rational, whereas the inversely weak-willed agent cannot.

1 Two types of ‘incontinence’: akrasia and weakness of will

Traditionally, the philosophical literature identifies akrasia and weakness of will as the same phenomenon: incontinence. The term ‘incontinence’ is used by Aristotle to describe an agent who, as a result of being either propeteia (impetuous) or astheneia (weakened by the influence of emotion), acts contrary to reason. The opposite to incontinence is continence (enkratia) which is when emotion runs counter to reason but reason wins out. Thus,

the incontinent man, knowing that what he does is bad, does it as a result of passion, while the continent man, knowing that his appetites are bad, refuses on account of his rational principle to follow them (Aristotle, 1925, p. 160 §1145b).

The incontinent agent who is ‘weak’ goes through a process of practical reasoning and decides what she ought to do, but then rather than act in

---

accordance with this she acts against it as the result of some *pathos* (emotion).\textsuperscript{75} The person who is impetuous does not go through the process of practical reasoning, but simply acts under the influence of an emotion and then comes to regret her action as being against the choice she would have made had she deliberated. For our purposes we need be concerned only with the first type of incontinence, that of acting against a reasoned choice as a result of the influence of an emotion, as this is the type of incontinence whose rationality is the subject of debate.\textsuperscript{76}

‘Incontinence’ is the English translation of *akrasia*. Traditionally ‘weakness of will’ and ‘akrasia’ are used interchangeably to describe Aristotle’s ‘weak’ agent. I am not concerned to argue for any particular interpretation of Aristotle’s own theory, for my concern us with the modern tradition of conflating an agent acting against her reasoned judgement with being weak willed. For example, Donald Davidson defines “an action that reveals weakness of the will or incontinence” thus:

\[\text{In doing } [\Phi] \text{ an agent acts incontinently if and only if: (a) the agent does } [\Phi] \text{ intentionally; (b) the agent believes there is an alternative action open to him; and (c) the agent judges that, all things considered, it would be better to do } [\Phi] \text{ than to do not-} [\Phi] \] (1980, p. 22).

However, that a weak-willed person intentionally acts against her atcj does not necessarily mean that in all cases of an agent acting against her atcj the agent is doing so because she is weak-willed: there are other reasons for which she might act against it. A distinction should therefore be drawn between the two phenomena, such that we can have cases of akrasia (intentionally acting against one’s atcj) that are not cases of weakness of will. I will argue that this distinction answers in the negative the question that is topical in the inverse akrasia debate,

\textsuperscript{75} There may be causes of incontinence other than emotion, also, but I shall not address such here.

\textsuperscript{76} The impetuous agent might actually be an easier case. Since there is nothing inherently irrational about acting under the influence of an emotion on my account, the absence per se of a reasoned choice does not make the agent irrational. If, as a result, she fails to do what she has most reason to do, then her action will be irrational; if she does what she does have most reason to do, then so long as the emotion that motivates her action is tracking those reasons she will be acting for them, and will thus be practically rational. If she later comes to regret her action because it wasn’t in fact what she had most reason to do, that will not add further irrationality to her already-irrational action. If she mistakenly regrets her action, this too will not add or detract from the original rational status of her action.
that of whether Huckleberry Finn can be rationally and morally praiseworthy for his ‘inverse’ incontinence. Inverse akrasia can be rational, but Huck is inversely weak-willed, and weak-willed action cannot be rational.

In the literature we find akrasia defined as “intentionally choosing the worse course of action” (Wiggins, 1987, p. 239), as action “against one’s best judgement” (Arpaly, 2000, p. 488), as acting against one’s “all-things-considered judgement” (Jones, 2003, p. 185) and as a “failure to act on one’s intentions” (Holton, 1999, p. 241). I shall adopt the definition of intentionally acting against one’s all things considered judgement, where an ‘all things considered judgement’ (atcj) is the judgement reached by an agent about what she ought to do based upon her reasons. As I argued in Chapter 3, the ideal outcome of practical reasoning is the formation of an intention to act in accordance with the agent’s judgement about what she has most reason to do, thus the outcome of her practical reasoning will be an intention to $\Phi$, on the basis of her judgement that she has most reason to $\Phi$. We should not conflate akrasia, weakness of will and acting against an atcj as the same thing, for although an agent might judge that she has most reason to $\Phi$ she could very well fail to form an intention to $\Phi$, and if she fails to form the intention this might have nothing at all to do with any weakness of will. It is also useful here to remember that an agent does not need to engage in practical reasoning in order to act intentionally.\footnote{I am assuming what might be called the ‘simple view’ of intention and intentional action (Bratman, 1987), whereby an intention is a mental state (contra Davidson, 1963; see Bratman 1987) in the form of a ‘practical commitment’ (Brandom, 1994, p. 256) to do something. The practical commitment is based upon certain reasons, such that it is revisable in the case of the agent revising her assessment of her reasons, and neither the assessment of reasons nor the commitment need be conscious. Intentions can be both future-directed and present-directed, or, as John Searle calls them, ‘prior intentions’ and ‘intentions in acting’ (1979), and are distinct from purposes and also deliberative action (Austin, 1979). Thus if an agent intentionally performs an action an intention is present, though it need not be a prior, future-directed intention, nor a conscious intention.}

I use the term ‘incontinence’ to describe all cases of an agent intentionally acting against her atcj about what she has most reason to do, ‘akrasia’ to describe all cases in which an agent forms an intention contrary to this atcj, and where this formation of a contrary intention explains her failure to act in accordance with her
atcj, and ‘weakness of will’ to describe all cases in which the agent forms an intention in line with her atcj, but fails to act upon this intention.\(^7\) We thus have two types of incontinence: akrasia where the agent’s practical reasoning produces an intention that conflicts with her atcj and upon which she acts; and weakness of will where an agent forms an intention to act, which may or may not be consistent with her atcj, upon which she fails to act.

For example, Bernado is deciding whether or not to get a taxi home. He reasons that the considerations that he has that count in favour of him doing so – that he is tired, and that a taxi will be the quickest form of transport home – do not justify him doing so in light of the considerations that count against him doing so – that it is by far the most expensive way to get home, it won’t be much quicker that getting a bus because of the rush hour traffic, it’s more economically friendly to use public transport or walk – and so, all things considered, he judges that he has most reason to not get a taxi home. He has formed the atcj that he ought not to get a taxi. ‘But’, he thinks to himself, ‘I want to get a taxi! I know I shouldn’t get a taxi, but I want to, and so I shall’. Bernado has acted akratically, because he has formed the intention to get a taxi against his atcj that what he has most reason to do is to not get a taxi (and ought to catch the bus or walk home instead).

Chino has the same decision to make, and the same reasons as Bernado. He also comes to the same atcj as Bernado – that what he has most reason to do is not get a taxi. But unlike Bernado, Chino forms the intention to not get a taxi, and to do otherwise. He forms the intention to get a bus home, and goes to the bus

\(^7\) In drawing a distinction between akrasia and weakness of will in this way I am departing from the traditional Davidsonian (1980) understanding of the topic. However, Christopher Peacocke (1985) also defines akrasia, as I do, as forming an intention, rather than just performing an action, that is contrary to one’s ATCJ; Alfred Mele (1987) also distinguishes from the normal cases of incontinence those cases in which one judges a particular action as right, forms an intention to do that, and then fails to act on it through weakness of will, and additionally points out that an agent can exhibit weakness of will in cases when he acts in accordance with his better judgement (1987, p. 7); Richard Holton (1999) argues that the majority of cases of weakness of will are not best characterized as cases of acting against one’s better judgement, and so also proposes a division of the terms ‘akrasia’ and ‘incontinence’; Amelie Rorty argues that the ‘akratic break’ can take place, in an Aristotelian schema, at four points, one of which is that between the formation of the judgement and the intention to perform the action judged as best, another of which is between the intention to act and the performing of that action.
stop intending to catch the next bus, which he know will be along in just a couple of minutes. However, a taxi approaches and, despite his intention to not get a taxi, Chino sticks out his hand to flag the taxi down, and jumps in. Chino has been weak-willed; he formed the atcj that he ought not to get a taxi, intended to not get a taxi, and yet here he is, in a taxi on his way home. He intentionally caught the taxi, but this intentional action contradicted the intention he formed as a result of his practical reasoning about what he ought to do. The difference between Chino and Bernado is that Bernado did not form an intention to act on his atcj that he ought not to get a taxi, and thus Bernado’s action of getting a taxi does not contradict his intentions; Chino did form an intention to act on his atcj that he ought not to get a taxi, and thus his action of getting a taxi does conflict with his intentions. Both Chino’s and Bernado’s action of getting a taxi conflicts with their atcj that they ought not to, but whereas Bernado’s action is akratic, Chino’s is weak-willed.

An obvious potential objection to my distinction is the argument that an agent cannot form an intention that is contrary to her atcj, thus if she forms the intention to \( \Phi \) she must actually judge that she ought to \( \Phi \), all things considered. Akrasia as I have defined it, then, would be impossible. It would follow that since Bernado formed the intention to get a taxi, then, despite evidence to the contrary, he must have actually judged that, all things considered, he really had most reason to get a taxi. I shall not dwell too long on this objection, as I think it is dealt with substantially elsewhere, and amounts simply to the Socratic position that intentionally acting against one’s judgement about what one ought to do is impossible.\(^79\) But if it is true that an atcj about what one ought to do, and intentionally doing that thing can come apart at all – that is, if it is true that *Chino* can form an atcj that he ought to not get a taxi, form the intention to do otherwise, and then intentionally act against this by getting a taxi - then an atcj and an intention must be able to come apart, thus an agent must be able to hold an

\(^79\) In the *Protagoras* Socrates declares that “no one who either knows or believes that there is another possible course of action, better than the one he is following, will ever continue on his present course when he might choose the better” (Plato, 1961, pp. 348-9; §358b-c). The literature from Davidson (1980) onwards denies the assumption that if an agent intentionally does something than it must be that she actually judged that she ought to do that thing.
intention that contradicts her atcj about what action she ought to perform. Thus if one was to deny that Bernado's case was not in fact one of incontinence - of intentionally acting against one's atcj - it would follow that Chino's was likewise not one of incontinence, and this would amount simply to the assertion that incontinent action per se is impossible.

In the ideal case an agent’s practical reasoning would conclude with an intention to do that which she has most reason to do, with her judgement about what she has most reason to do leading to the intention to do that thing, whether the commitment is to perform the proposed action immediately, or in the future. In the ideal case, her intending to do that thing would lead her doing that thing, when the time came to do it. Thus there are at least two places where the ideal result may not occur; between judgement and intention, as with Bernado, and between intention and action, as with Chino. As Amelie Rorty puts it, the “akratic break” can occur in more than one place (1980, p. 333). In akrasia, the akratic break takes place between the forming of the atcj and the forming of the intention; in weakness of will the failure takes place between the formation of the intention, and that intention translating into action. It is only if the agent fails to act on her intention when she ought to that the action is one of weakness of will, as I shall now argue. If the agent ought to have revised her intention, the action was not incontinent at all.

2 Weakness of will as failure to act upon one’s intention when one ought to

I have defined akrasia as forming an intention that contradicts one’s atcj about what one has most reason to do, and weakness of will as forming an intention that coincides with one’s atcj about what one has most reason to do, but then failing to go through with it. Clearly, not all cases of intentionally contradicting one’s intention to do something are cases of weakness of will. As Richard Holton

80 Or, at least to her attempting to do that thing: that the agent does not succeed does not negate the existence of her intention. As Elizabeth Anscombe remarks, one need not really say, when reporting an intention, “I am going to . . . unless I am prevented” (1957, p. 91).
says, “[s]ometimes we realize that our intentions were ill-judged, or that circumstances have changed to make them inappropriate” (1999, p. 241). Following Holton, I propose that weakness of will occurs “when agents are too ready to reconsider their intentions” (ibid.): their will in the direction of transforming their intention into action is too weak.

By ‘too weak’ I mean weaker than it ought to be, not just that it fails in this particular case to be strong enough, since in some cases we ought to revise our intention and thus a will that was strong enough to override such would be too strong, being the vice of stubbornness rather than a virtue. Sometimes we act against an intention because we have thought better of it. This might happen because new information comes to light which was either not available when we formed the intention to act, or was not given proper consideration when we formed the intention to act. For example, I might reconsider my intention to purchase standard-fare rail tickets over first-class rail tickets once I discover that the standard-class tickets are only a few pounds cheaper than the first class tickets, because when I initially formed my intention to buy second-class ones I’d assumed I’d thereby save a lot of money (new information has come to light), or I might reconsider my intention because, although the prices are as I expected, I now decide that when forming my initial intention I had given too much weight to the reason that I’d save money, and not enough consideration to the reason that first-class travel is more pleasant and better facilitates the work that I want to do on my journey (information was not originally given proper consideration). In both cases I have acted against my initial intention, but have formed a new intention upon which I have acted.

I propose that whether or not a case is one of weakness of will depends on whether the agent failed to act upon an intention that she ought to have acted

---

81 Holton also draws a distinction between akrasia and weakness of will, though on slightly different grounds from mine: his definition of akrasia I think would exclude many of the cases which my distinction would include, as he more readily attributes seeming cases of akrasia to cases in which judgement has in fact changed (n.d.). Holton’s definition of weakness of will, however, has heavily influenced mine.

82 See also Holton (1999) at page 252.
upon. If she ought to have acted on her intention, then it is a case of weakness of will; if she ought to have revised her intention, then it is not. But by 'ought' here I do not mean that the action one did end up performing was what one ought to do in the first place; I do not mean that if an agent ends up doing what was in fact the 'right' thing this entails that she was not weak willed, but that if she ends up doing what was in fact the 'wrong' thing that she was; weakness of will does not depend on either the objective or subjective oughts about what one ought to do of which I have been talking throughout this thesis (see Chapter 2, in particular). The question of whether or not one ought to act on one's initial intention to \( \Phi \) does not necessarily have the same answer as the question of whether or not \( \Phi \)-ing was what one has most reason to do.

When an agent's action is the result of weakness of will, her action is *intentional*, even though it goes against her prior intention: she intended to \( \Phi \), and then intentionally did not \( \Phi \). Whether or not her action was weak-willed or not depends on the answer to the question of *whether she ought to have acted upon her intention, or ought to have revised her intention*. If she ought to have revised her intention then the case is not one of weakness of will; if she ought to have acted upon her intention then it is one of weakness of will. Thus I am characterizing weakness of will as cases in which *an agent's intention failed to do what it ought to have done*. An agent's intention ought to be immune to reconsideration unless for good grounds. Thus if there are not got good grounds to revise the intention that she formed as a result of practical reasoning, she ought not to revise it; she ought to act upon it and if she fails to do so then her action is one of weakness of will.

### 2.1 An agent's intention ought to be immune to reconsideration unless there are good grounds

Agents should revise their intentions if and only if there are good grounds to do so. If there are no good grounds on which to revise one's intention, one should act upon that intention. In the rail ticket example, had I reached an atcj that I ought to purchase the standard class tickets and formed the intention to purchase
standard-class tickets, I ought to act upon that intention unless I have grounds to reconsider the intention; but had I good grounds to reconsider the intention, then I ought to do so. Let’s say that my initial intention was based upon the reason that the standard class fare was cheaper. When I get to the station I find out that in fact first class tickets are cheaper, and now my intention is that I have most reason to purchase first class tickets. This entails that I *ought* to revise my intention: it would be irrational of me to persist with my intention to purchase standard class tickets now that my intention is that I ought to purchase first class tickets.

If I did not have good grounds on which to revise my intention, then I ought to have acted upon it and purchase standard class tickets. Perhaps I am travelling with a friend, and my friend has already purchased a standard-class ticket, and I initially decided to purchase a standard-class ticket was so that I could sit with my friend. In this case whether or not I ought to change my mind would depend upon whether I still have more reason to travel standard-class than first class: if the cheapness of the first-class tickets is a reason that outweighs the reason of being able to travel with my friend, then I ought to change my mind and form the intention to purchase first class tickets; if it is not, then I ought to stick to my initial intention and resist the ‘temptation’ to revise it.

My intention *ought* to be immune to such ‘temptation’: *that is its purpose.*\(^3\) It ought not to be immune to revision *per se*, but to revision for insufficient reason. That is, my intention ought to *assist me in being practically rational*. The *point* of practical reasoning is to work out what one ought to do, that is what one has most reason to do, and its conclusion is an intention to act: its conclusion, therefore, ought to be an intention to do what one has most reason to do. The point of forming an intention about what one ought to do is for it to lead to an *intention* to do that thing, thus the intention ought, rationally, to be to do what one has most

---

\(^3\) For argument that such is the purpose of prior intentions, and indeed for a defence that talk of intentions is even legitimate, see Michael Bratman’s *Intention, Plans and Practical Reason* (1987). Gilbert Harman (1986) also defends talk of intentions as legitimate. Both Christopher Peacocke and Richard Holton, as already cited, argue that acting against a previously formed intention is crucial to the concept of incontinence.
reason to do. If the purpose of an intention is to move one to do what one has most reason to do, this intention ought to be acted upon only so long as it continues to reflect what one has most reason to do. Rationally, then, if one realizes that the initial intention does not reflect what one has most reason to do, then one ought not to act upon it. To continue to do so in the light of contrary reason would not be continence but stubbornness.\(^{84}\)

Just as what an agent ought to do is the 'subjective' ought of that which she has most accessible reason to do, whether an agent ought to act upon or revise her intention is dependent upon those reasons she has to which she has access. If I formed the intention to buy a standard-class ticket on the basis of it being cheaper, not knowing that in fact the first-class tickets were actually cheaper for this particular train, that I have the reason that first class tickets are cheaper does not mean I ought to revise my intention since I do not have access to this reason. If I revised my intention to buy standard-class tickets for some bad reason - for example I am swayed by the idea of having beverages brought to my seat which, while is a reason that counts in favour of travelling first-class, is not worth what I think to be the doubled price – and thereby end up doing what I in fact objectively ought to do, this does nothing to mitigate the extent to which I was weak-willed for giving in to the temptation of having beverages brought to my seat. Agents are practically rational to the extent that they do what they have most reason to do for those reasons, and since I didn’t revise my intention for the reasons that made it such that buying first-class tickets was what I had most reason to do, I didn’t buy first-class tickets for the reasons that made buying first-class tickets what I ought to do. Thus I ought to have stuck to my initial intention to buy second-class tickets and was weak-willed for not doing so. That I happened to do what I in fact had more reason to do is a ‘lucky coincidence’ and does not detract from my irrationality, for I am practically rational only to the extent that I do what I have most reason to do for those reasons. An intention should be adhered to only so long as it continues to reflect the agent’s judgement about what she has most

---

\(^{84}\) Nietzsche appears to have held that one should never revise an intention, advising, “Once the decision has been made, close you ear even to the best counterargument: sign of strong character. Thus an occasional will to stupidity” (1989, p. 84 ; §107), however this seems obviously to be ill advice.
reason to do, thus if she changes her judgement then her intention should fall in to line. If she has good grounds to reconsider her intention, then she ought to do so.

An agent is given good grounds to reconsider her intention if she suspects that it no longer reflects what she has most reason to do. This needs some qualification, for it could be that, for example, just as I am about to purchase my train tickets in person it occurs to me that I might be able to get them cheaper if I bought them online, thus I *prima facie* have grounds to reconsider my intention; yet let’s say that finding out whether it is true that I will be able to get them cheaper online entails returning home to access the internet, and I quickly calculate that the amount I could potentially save is unlikely to be worth the hassle of having to make a second trip to the railway station to buy them in person if it turns out that I was wrong and they are actually more expensive online. Thus I might suspect that my initial attempt might have not actually been spot-on, but the reason I now realize I failed to consider is not a good-enough reason to revise my intention. It may be that initially my intention did not reflect what I in fact had most reason to do, but in these revised circumstances it actually does reflect what I have most reason to do, for I now have the additional reason of avoiding a repeated journey to the railway station. Thus I might briefly reconsider my intention, but decide that it is still the best thing to do.

If an agent is not given good grounds to reconsider her intention then she ought not to reconsider it, and ought to act upon her intention. If intentions are to fulfil their function then they need to be resistant to reconsideration in the face of anything other than good-enough grounds, and the bar for what counts as ‘good-enough’ is relatively high. Intentions ought to have what Michael Bratman calls *stability*, that is, once formed they ought to persist for a future-directed intention to perform a certain action ought to lead the agent to perform the action, that is, in Bratman’s term, they ought to be *controlling* (1987, p. 16). The purpose of future-directed intentions is to ‘store’ our decision about what we ought to do, when the time comes, and so it needs to be stable so as to be immune to revision: were we to re-open deliberation when the time comes that would defeat the
purpose of the decision we had already made, thus unless the intention has a certain level of stability it would be practically pointless. There may be a variety of purposes for forming future-directed intentions, for example it may be that there won’t be time to deliberate when the times comes, or it may be that other actions will depend upon what we decide to do and we will need to perform these actions in the meanwhile. However, the purpose of intentions with which I am concerned is that I am aware that unless I resolve to Φ now, when the time comes to Φ I won’t Φ, or I’ll be unlikely to Φ.

Say I know that when it comes to buying tickets tomorrow I’ll be tempted by the lure of having beverages served to me at my seat: if I deliberate now, taking into account the reason of having beverages delivered to my seat, and form the intention to purchase standard-class tickets, then when tomorrow comes the lure of having beverages served to my seat is, I hope, less likely to tempt me, because I have already deliberated and formed the all (including this reason) things-considered judgement that I ought to purchase standard-class tickets. Because this reason is included in my atcj it ought not to ‘tempt’ me away from performing my intended action, as it might to without my strengthening my resolve against it doing so by deliberating and forming my intention to act away from the salience of this reason. Intentions in this way stand as a sort-of ‘placeholder’ for one’s reasons: they reflect the balance of normative force of all of the reasons which were considered in coming to form that intention. Of course they reflect the normative force of our reasons from the deliberative stance from which we formed the intention at the time we formed it, and may not reflect the normative force of our reasons as assessed from the deliberative stance in which we are at the time at which we are supposed to perform the action. In the cases in which I am particularly interested, this is the point: aware that our deliberative stance when the time comes may not reflect the balance of our intrinsic desires that we

---

85 Holton gives the example of wanting to paint his front door tomorrow, and having to decide today what colour to paint it for he needs to buy the paint today (1999, p. 244).
86 See also Holton, who suggests that, “it would be useful for me to form an intention now, an intention which will lead me directly to act when the time comes, and which will be somewhat immune to reconsideration in the light of the desires I shall have then” (1999, pp. 245, italics added). Holton further suggests that telling others of our intentions employs as an incentive their potential disapproval should we fail to act upon them (ibid, p. 246).
endorse, we form our intention to act at an earlier time at which we judge that our deliberative stance does reflect a balance of intrinsic desires which we endorse. That is we think that at this earlier time our atcj is more likely to accurately reflect what we in fact have most reason to do, and so judge that we ought to deliberate now, so that we form an atcj that is more likely to be correct, and thus form an intention to do what is more likely to be what we actually ought to do, and thus, if our intention does what it is supposed to do, leads us to actually do what we have most reason to do. That is, our future-directed intentions are a mechanism by which we guide our actions in line with practical rationality.\textsuperscript{87} They act, as I said before, as a sort of placeholder: if we do what we formed an intention to do, then via the intention we are acting for all of the reasons which were taken into account when we formed the intention.

An agent should reconsider her intention, therefore, \emph{only} if she has good grounds for suspecting that the outcome of her prior deliberation might have been wrong. This could be because circumstances have changed and thus there are previously unconsidered reasons, or because she did not take in to account a reason that she ought to have done, or because she incorrectly assessed the normative force of her reasons. I propose that weakness of will occurs when she revises her intention when she ought not to: either the previously-unconsidered reason is not good enough to change what she ought to do, or she misjudges the normative force that the ‘swaying’ reason actually has. This is why it is important for deliberation to be thorough: if the agent knows that she will face temptation, then she needs to ensure that she has taken into account all of her reasons, and their real normative force.

One way in which intentions can too easily fail is if agents underestimate the normative force of their reasons. For example, the dieter who forms the intention to not eat cake may severely underestimate the force of his reason of being hungry that counts in favour of his eating cake. As often happens with dieters,

\textsuperscript{87} See also Bratman, who argues that “intention formation is but one of several strategies for the resolution of practical conflicts” (1987, pp. 137-8).
those trying to quit smoking, and other common fruitful areas of weak-willed action, resolutions are made in a state where they are too easy to make. It is easy to resolve to never again eat cake after having eaten three slices in a row, since there is no current desire for cake, thus no affective state representing the normative force of the reasons for which one wants to eat cake. The reasons for which the dieter wants to eat cake are likely to be either ignored or too-easily dismissed as weak in the state of feeling sick as a result of a cake-binge.

Consider too the very common experience of swearing that one will 'never drink again' when sporting a hangover. One can't imagine ever desiring alcohol in that state: the reasons for which one does appear completely insignificant in the light of the salience of the reasons for which one doesn't, when one is experiencing the salience of those reasons against to their absolute extreme. It is perhaps foolish to make resolutions in such states, for they will not reflect all of one’s reasons, or the real normative force of one’s reasons, and thus will not be immune to revision when it transpires that the intention did not in fact incorporate one’s reasons properly. If the intention does not reflect the true force of one’s reasons, it is more likely to be erroneously revised: it may still be the case that the dieter ought not to eat the cake, but his resolution is more easily abandoned if the salience of his reasons for eating it so obviously exceed the balance of his reasons as assessed when he formed the intention. In the light of temptation, his reason of hunger, for example, might appear to have a much higher normative force than it ought to.

I have argued that we should distinguish between two sorts of incontinence: akrasia and weakness of will. Akratic agents deliberate about what they ought to do, form an atcj to that effect, but then form the intention to do otherwise, and thus act against their atcj. Weak willed agent deliberate and form an atcj, and form the intention to act upon that atcj, but they fail to follow through with the intention when they ought to follow through with that intention. What makes it the case that they ought to follow through with their intention is that they have not been given good grounds on which to revise their intention. If they have got good ground on which to revise their intention, then there is nothing incontinent about
their doing so and thereby acting against their initial atcj. Such agents are not incontinent: they have merely changed their mind about what they have most reason to do. They have formed a new atcj and then intentionally acted in accordance with that atcj, and thus they are not incontinent at all. The rationality of their actions will be judged as standard: they will be practically rational to the extent that their behaviour is guided by the regulative norm of doing what they have most reason to do.

I have so far been arguing about ‘regular’ incontinence: when an agent fail to do as she ought because she has failed to act in accordance with her atcj. I now turn to ‘inverse’ incontinence: when an agent’s atcj was wrong in the first place, and by acting against it she ends up doing what she in fact ought to have done. Inverse incontinence, like regular incontinence, can be either akratic or weak-willed. The question of the inverse incontinence debate is whether such action can be rational. Some, such as Arpaly (1999, 2000, 2003) argue that because the agent’s action is in line with what she has most reason to do, it is practically rational. Others, such as Döring (2013), argue that because the action contradicts her atcj, it cannot be. I am going to argue against both of these positions: an agent’s action is not irrational just by virtue of it contradicting her atcj since, as I have argued throughout this thesis, actions do not have to be the result of practical reasoning in order to be rational; but an agent’s action is not rational just by virtue of it being in line with what she has most reason to do, since only those actions under the agent’s guidance control, by virtue of being responsive to reasons, count as rational. The intuition that drives the debate, that of the seemingly praiseworthiness of some inversely incontinent actions (such as that of Huckleberry Finn’s protecting his slave friend Jim), can be satisfied by drawing the distinction between akrasia and weakness of will for which I have argued. Actions that contradict an agent’s atcj by virtue of weakness of will cannot be rational, but actions that contradict an agent’s atcj by virtue of being akratic, can. This is because actions that are the result of weakness of will are, by definition, not responding to reasons, whether or not they happen to be what the agent has

---

88 The concept of ‘inverse akrasia’ was introduced by Nomy Arpaly to describe cases in which an agent does the right thing, but does so against her best judgement (2003, p. 75)
most reason to do, whereas actions that are the result of akrasia can be responding to reasons, and thus can be rational. I shall begin with the paradigm example used in the debate – the inversely incontinent action of Huckleberry Finn – and argue that it is one of weakness of will, not akrasia. I shall then argue that, because weak willed action is, by definition, not responsive to reasons (for if the agent’s failure to act on her intention is the result of responding to reasons her action is, by definition, no longer one of incontinence), Huckleberry Finn’s action is not rational. Thus, the example around which the debate is centred does not prove that inverse akrasia cannot be rational; it demonstrates that weak willed action cannot be rational. I will then modify the example to be one of true inverse akrasia to demonstrate that such can be rational.

3 Huckleberry Finn is a weak agent, not an akratic agent

I have argued that there are two sorts of incontinence; akrasia and weakness of will. Akrasia is forming an intention to act that contradicts one’s judgement as to what one has most reason to do. Weakness of will is forming an intention to act that coincides with one’s judgement as to what one has most failure to do, but then failing to follow through with that intention because one has too easily revised that intention. In this section I will argue that under this division, Huckleberry Finn is a weak agent, not an akratic agent.

Mark Twain’s character of Huckleberry Finn was raised in a society in which slaves were seen as property, and this was not questioned. After having helped his friend Jim run away from slavery, Huck decides to turn him in. It is clear in the story that Huck truly believes that he ought to turn Jim in (his atcj is that he ought to do so), and he forms the intention to do just that:

I got to feeling so mean and miserable I most wished I was dead … My conscience got to stirring me up hotter than ever, until at last I says to it: ‘Let up on me – it ain’t too late yet – I’ll paddle ashore at first light and tell.’ I felt easy, and happy, and light as a feather, right off. All my troubles was gone (Twain, 1996, pp. 123-4).
However when the opportunity to hand Jim over to the slave hunters arrives, Huck fails to act on his intention, because his emotion of sympathy prevents him from doing what he believes he ought to do. The slave hunters to whom Huck intended to turn Jim in ask whether Huck’s companion (Jim) is white or black, and he lies to them instead of acting upon his intention:

I didn’t answer up prompt. I tried to, but the words wouldn’t come. I tried, for a second or two, to brace up and out with it, but I warn’t man enough – hadn’t the spunk of a rabbit. I see I was weakening; so I just gave up trying, and up and says – ‘He’s white.’ (ibid., p. 125)

Huck is incontinent because his action – protecting Jim the slave from the slave hunters by lying and saying “He’s white” – contradicted his atcj that he ought to turn Jim in. Huck is described in the philosophical literature as being ‘inversely akratic’ (Arpaly, 1999; 2003; Jones, 2003; Döring 2010; 2013; Bennett, 1974; de Sousa, 2010), but these authors do not draw a distinction between akrasia and weakness of will. If they did, they should agree that Huck is weak-willed rather than akratic, because he does form the relevant intention to act on his atcj -“I’ll paddle ashore at the first light and tell” - and even starts to act upon it. However, as he does so Jim thanks Huck for having been such a good friend, which “seemed to kind of take the tuck all out of [him]” and his paddling starts to slow down. He tries still to go through with it (“Well, I just felt sick. But I says, I got to do it – I can’t get out of it”), but as he says, “the words wouldn’t come … I see I was weakening; so I just give up trying and says – ‘He’s white’” (Twain, 1996, pp. 124-5). Huck intentionally lies as a result of weakness of will: he is acting against his intention to turn Jim in, the intention which he formed as a result of practical reasoning, in line with his belief that turning Jim in was the right thing to do.

I argued that weakness of will occurs when an agent fails to act on an intention that she did not have good ground to reconsider. Thus for Huck’s action to count as being weak willed, it must be that Huck did not have good grounds on which to reconsider his intention. If he ought to have reconsidered his intention, then he would not count as being incontinent at all. Had Huck rationally revised his intention to turn Jim in, then the rationality of his protecting Jim would not be at issue: had he changed his mind about what he ought to have done, and acted in
accordance with this act, his new act to protect Jim, his newly-formed intention to protect Jim, and his action of protecting Jim all would have reflected what he, in fact, subjectively had most reason to do, thus his action would have been perfectly rational. Thus for his action to count as incontinent at all, we must assume that he did not have good grounds on which to reconsider his intention.

The question of Huck’s rationality, and the rationality of any incontinent action, thus comes down to this: did he revise his intention to turn Jim in on good grounds. If the answer is ‘yes’ then Huck’s protecting Jim was rational, but it was not incontinent. If the action is ‘no’ then his action was incontinent, but it was not rational. Thus the real question of the debate is not whether Huck’s incontinent action was rational, but whether Huck’s action was incontinent or rational. The example of Huckleberry Finn cannot prove that inversely incontinent action can be rational, because if his action is correctly described as incontinent then it is, de facto, because it is weak willed, irrational. This is because Huck’s action is one of weakness of will, not akrasia. The example of Huckleberry Finn cannot prove that inverse akrasia can be rational, since Huck’s action is not akratic. It cannot prove that inverse weakness of will can be rational, since, by definition weakness of will is irrational: the thing that could make it rational – being responsive to reasons – also disqualifies it from being weakness of will. Thus, as I shall now argue, inverse weakness of will cannot be rational, but inverse akrasia can.

4 Inverse weakness of will is, by definition, not responsive to reasons and thus, by definition, not rational

In this thesis I have argued that an agent is practically rational to the extent that she does what she has most reason to do for those reasons. Inverse weakness of will and inverse akrasia both involve the agent doing what she has most reason to do, but only in the case of akrasia can an agent be acting for those reasons. A weak-willed agent’s action might be in line her reasons, but cannot be for those reasons since it is not responding to her reasons; it is merely tracking them. Thus
an inversely akratic agent can be practically rational but a weak-willed agent cannot. Huck is a weak-willed agent, thus his action of protecting Jim cannot be rational.

In Chapter 2 I argued that an agent is practically rational to the extent that she is guiding her action by the regulative norm of doing what she has most reason to do. I argued that the capacity to engage in practical reasoning was fundamental to her ability to do this and that, although actions do not have to be the direct result of practical reasoning in order to be so guided, the agent’s capacity to engage in such needs to be active, such that she would, if needed, engage in such. This means that her System 1 reason-tracking mechanisms are under her rational guidance control because if they deviated from the norm of doing what she had most accessible reason to do, this would be detected by System 2 and she would engage in practical reasoning in order to regulate her behaviour. As Karen Jones argues, the practically rational agent would not rely on a reason-tracking mechanism were it reasonable for her to distrust its deliverances: she “must be sensitive to when putative reasons are defeated and when they are outweighed”, which “requires the capacity to reflect on the status of the deliverances of those mechanisms that purport to latch on to reasons such as … emotion …, but also the capacity to reflect on reasoning itself – for it too can deliver false representations of the reasons that obtain” (Jones, 2003, p. 190).

In order to count as guiding her actions via her reasons, an agent must satisfy the condition of being a reason-responder. That means that she must satisfy the condition of self-monitoring her sub-systems and mistrusting their deliverances when she ought to. The inversely weak-willed agent does not do this. She ought to mistrust the deliverance of her reasoning because the inversely weak willed agent, by definition, has been given good grounds to (the grounds which she has been given are those that make her weakness inverse; those that make what she ended up doing what she objectively had most reason to do in the first place). If she hasn’t been given good grounds to mistrust the outcome of her reasoning then she is not incontinent because her action is what she has most accessible reason to do (even though it is not what she objectively ought to do). If she has
been given good grounds to mistrust the outcome of her reasoning and she does reconsider her intention, then she is not incontinent either: she has revised her atcj in the light of the grounds for which she has been given to mistrust the outcome of her reasoning, and is intentionally acting in accordance with this new atcj. It is only if the agent has good grounds to mistrust her original atcj and fails to respond to this by revising her intention that she is incontinent, for this is the only case in which she will be acting against her most recent atcj. And because she is acting against the intention that coincides with her atcj, she is weak-willed (rather than akratic). Thus, by definition, the weak-willed agent is failing to guide her behaviour by the regulative norm of doing what she has most reason to do, for she is not responding to her reasons by mistrusting the deliverances of her reasoning when she ought to do so. If she did respond to her reasons, she would mistrust the outcome of her reasoning, and would revise her intention.

Huckleberry Finn is weak-willed. Those that maintain that Huck is practically rational, despite his incontinence, need to establish that he was responding to his reasons. For example, Arpaly argues that Huck acts for the reason that Jim is a person, just like him, and that, because this is the reason for which Huck ought to protect Jim, Huck does what he has most reason to do, for those reasons, and thus is practically rational (2003). However, why is it that Huck is acting for the reason of Jim’s personhood, rather than just in line with it? Huck, crucially, does not consciously endorse this reason as a reason. His atcj is that he ought to turn Jim over to the slave hunters. And he does not mistrust his atcj in the light of his emotion; if he did, then yes, his action would be rational, but it would not be incontinent. But Huck does not question his atcj: he fails to act in accordance with by failing to act upon his intention. He ought to have questioned it and thus he is weak willed. In what way, then, can Huck’s weakness of will be attributed to his responding to reasons?

Arpaly’s argument is that Huck’s emotion of sympathy is tracking this reason, and that because his protecting of Jim is motivated by his emotion of sympathy, this means that Huck is acting for the reason of Jim’s personhood. However, an action is practically rational to the extent that it is under the agent’s rational guidance
control. That one of her reason-tracking mechanisms happened to track what she, in fact, had most reason to do does not make that action rational: it must have been under her regulative control. And Huck's protecting Jim was not. This is because had it been under his regulative guidance control he would have mistrusted his atcj. Recall that an action being under an agent's regulative guidance control entails that an agent would engage in practical reasoning if she ought to: reason-tracking mechanisms can motivate rational behaviour in the absence of atcjs formed as a result of practical reasoning iff the deliverances of those reason-tracking mechanisms would be subject to conscious reflection when they ought to be. Huck's intention to turn in Jim ought to have been subjected to conscious reflection in the light of the deliverances of his reason-tracking emotion of sympathy, thus his behaviour was not under his regulative guidance control. Had it been, he would have mistrusted the deliverance of his reasoning and subjected it to conscious reflection. Thus, because he did not consciously reflect on the accuracy of his intention to turn Jim in in the light of his conflicting emotion of sympathy, his behaviour was not being guided by his regulative guidance control. This means that, even if his emotion was tracking the reason of Jim's personhood, his action was not under his regulative guidance control and thus could not count as the practically rational action of a practical agent. Thus Huck's incontinence – and all weak willed action – is, by definition not rational. However, as I shall now argue, inverse akrasia can be rational, for it can be under the agent's regulative guidance control.

89 By the same lights, weak-willed actions such as Huck's do not have moral worth, contra Arpaly's (2003; also Arpaly and Tim Schroder, 1999). Arpaly and Schroder argue that Huck's protecting Jim has moral worth on the basis that he is "averse to turning Jim in for morally significant reasons" (1999, p. 164). The morally significant reason is that "while he 'knows' that Jim, being black, should be treated differently from other people, he can't help but experience Jim as human, as not so different from white people" (ibid.). This makes Huck morally praiseworthy, argues Arpaly, on the basis of her principle of \textit{Praiseworthiness as Responsiveness to Moral Reasons (PRMR): "For an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons – that is, the reasons for which she acts are identical to the reasons for which the action is right" (2003, p. 72). I am happy to concede this principle of moral worth, but maintain that under this principle Huck is not morally praiseworthy because he does not act for the relevant moral reason. However, since inversely akratic agents can be acting for reasons, an inversely akratic agent could do the right thing for the relevant moral reason, and so could, under Arpaly's \textit{PRMR} principle, be deserving of moral praise."
5 Inverse akrasia can be responsive to reasons and thus can be rational

Karen Jones maintains that inverse incontinent action will be rational just in case:

(1) the action is produced by a sub-system that reason-tracks because the agent reason-responded, and (2) the agent would have distrusted her [atcj] were her self-monitoring dispositions operating as they should (Jones, 2003, pp. 196-7)

Huck mistakenly trusts his atcj rather than his sympathy but, on Jones’ account, his incontinent action would be rational if it were the case that he would have distrusted his atcj had his self-monitoring dispositions operating as they should. Huck’s action is produced by his reason-tracking mechanism of sympathy and because Huck generally endorses this reason-tracking mechanism it is, on Jones’ account, reason-responding. She argues that, “our reason-tracking mechanisms can reason track because we, as agents, reason-respond” (2003, p. 196), which she thus explains:

sometimes we, as agents, initiate a method, or recalibrate a mechanism in order to latch-on to our reasons; other times, a mechanism will be maintained in place under the ‘virtual’ guidance of our reflective self-monitoring capacities (2003, p. 196 f.)

This would mean that Huck’s incontinent action was rational. However, although were it the case that Huck did endorse his sympathy, I would agree, he does not. On Jones’ picture, because Huck generally endorses his emotional capacities he is thereby endorsing this episode of emotion, however, this cannot be right. Jones herself argues that an agent is acting as a reason responder if and only if she would, if appropriate, question the deliverances of her reason-tracking mechanisms, but it is appropriate for Huck to question the deliverances of one of his reason-tracking mechanisms – his atcj – and he does not. Thus Huck is not acting as a reason-responder. If an agent’s actions were deemed to be under her regulative guidance control just by virtue of the fact that she generally endorses her emotional capacities, then we would be hard pushed to find any actions which failed to be so guided. Humans do, generally, trust their emotional capacities, so this cannot be enough to distinguish actions born out of emotionally-motivated weakness of will that are rational from those that are not. Huck has failed to
reason-respond, thus his action is not practically rational. However, this does not mean that incontinent action cannot be rational. Had Huck’s incontinence been akratic, rather than weak-willed, then it could have been.

If Huck had endorsed his emotion of sympathy as a reason-tracking mechanism then his action would be akratic rather than weak-willed and then it could be rational. Let us imagine an akratic version of Huck: ‘Schmuck’. Schmuck finds himself in the same situation as Huck, but instead of forming the intention to turn his slave friend in, he forms the intention to protect him. Schmuck has the same moral beliefs as Huck; he judges that morality dictates that he ought to turn Jim in. But Schmuck cannot shake off his feeling of sympathy. He feels sick at the idea of turning in his friend, and despite repeated examination of the situation he cannot locate the reasons for this nausea, which he attempts to do so that he can factor them in to his judgement and thus bring this recalcitrant feeling under the control of reason. Schmuck decides to trust his feelings and protect his friend, despite the fact that doing so contradicts his atcj about what he has most reason to do. He thinks to himself, ‘I know I ought to turn him in, I know that’s what I should do, but goddammit I don’t want to: he’s my friend. So I shan’t. I’ll lie to those slave hunters and protect him.’ Schmuck would be just like Huck except his incontinence would be the result of akrasia – forming an intention that contradicts his atcj – rather than as the result of weakness of will – performing an action that contradicts his intention. And whereas Huckleberry’s action cannot be rational because he fails to control his behaviour by the norm of doing that for which he had most reason, Schmuckleberry’s can, for he succeeds.

Inverse akrasia can be rational if the agent chooses to trust the reason-tracking mechanism that produces the action. It is not acting against her atcj that makes an incontinent action irrational, but why she acts against it; the weak-willed agent does not choose to act against her atcj and this is why her action is irrational: the action is not under her regulative guidance control. An agent can guide her behaviour by routes other than by her atcj, but her behaviour must be under her regulative guidance control – which entails that if she ought to trust the deliverances of her reason-tracking mechanisms then she would – for it to count
as rational action. Weakness of will undermines this guidance; akrasia does not necessarily undermine this guidance. Thus weak-willed action is necessarily irrational, but akratic action is not.

I have argued that we should distinguish between two types of incontinent action: akrasia and weakness of will. An action is practically rational to the extent that it is guided by the agent in line with the regulative norm of doing what she has most accessible reason to do. Action that is the result of weakness of will is necessarily not guided by this norm and so is necessarily irrational. Akratic action can be guided by this norm if the agent chooses to trust the reason tracking mechanism which produces the incontinent action over her atcj. Thus weak-willed action is necessarily irrational, but akratic action is not.
Chapter Eight
Loss of Self-Control

In this chapter I am going to apply my theory of the role of emotion in practical rationality to a question in the practical domain; that of what it means to ‘lose self-control’ in the context of a particular area of law. The relevant area is that of killing in response to a provocation, which is a defence to murder. The defence has been contentious, with the general consensus being that the requirement that the defendant ‘lost self-control’ ought to be removed because, as the Law Commission concluded, “[t]here is no satisfactory definition of loss of self-control” (2004, p. 35; §3.26). However, the defence has been recently reformed and, far from excluding the problematic concept of ‘loss of self-control’, it has been placed at the core of the new defence, with it constituting the title as ‘The Defence of Loss of Control’. In this chapter I will argue that my thesis of the role of emotion in practical rationality sheds light on this problematic concept and can provide a satisfactory definition of it. I propose that what it is to lose self-control in the context of the defence is for the question of whether or not one ought to Φ to have been settled in the affirmative prior to one’s having the chance to consciously consider whether or not one ought to Φ, and for this decision to then be immune to reassessment. The ‘control’ that is ‘lost’ is the regulative guiding control characteristic of the reason-responder. Understanding practical agency as reason-responsiveness, and understanding the role that emotions play within it as per my thesis, enables this coherent understanding.

1 Loss of Control as a defence to murder

Until very recently, there existed in the U.K. the defence of Provocation. This defence was a special, partial defence: ‘special’ because it applied only to murder; ‘partial’ because it was not a full defence to murder, but reduced the charge of murder to voluntary manslaughter. In the U.S. the defence still holds.

---

90 That it is only a partial defence means that the defendant is still found guilty of a crime, that of committing an unlawful killing. This differs from full defences, such as the defence self-defence, which would, if successfully pleaded, result in the finding of no unlawful killing having occurred.
In the U.K. it has been replaced by the defence of Loss of Control\textsuperscript{91}. Loss of Control is, in the same way, a special and partial defence that reduces the charge of murder to that of voluntary manslaughter.\textsuperscript{92}

The defence of Provocation was a common-law defence. It consisted of two tests, one objective and the other subjective. The subjective test was composed of two parts: whether the defendant was provoked in to losing her self-control; and whether she killed because she had lost control. Both subjective requirements had to be satisfied as well as the requirements of the objective test of whether the provocation was enough to make a reasonable person (of the defendant’s age and gender) do as the defendant did. The subjective test, in particular the requirement of loss of control, was the subject of much controversy and debate, much of which revolved around the meaning of the concept of 'loss of control'.\textsuperscript{93}

The Law Commissioner’s Report on Partial Defences to Murder (2004) concluded that,

> In summary, the requirement of loss of self-control was a judicially invented concept, lacking sharpness or a clear foundation in psychology. It was a valiant but flawed attempt to encapsulate a key limitation to the defence - that it should not be available to those who kill in considered revenge (§3.30).

The Report recommended that the provocation defence ought to be reformed, but the recommendations for reform notably \textit{excluded}, not just revised, the requirement that the defendant lost control (§3.115), and the consensus in academic discussion was that the concept had “serious logical and moral flaws” (The Law Commission Consultation Paper no.173 §6.14), such that it should be removed. Further, it was argued, the requirement was unnecessary, as “decisions on the provocation defence [could] be made correctly by relying solely on [the]...

\textsuperscript{91} Coroners and Justice Act 2009 ((CJA 2009) §54

\textsuperscript{92} The new defence is broadly similar to that of provocation in its requirements, however is thought to be more restrictive: The Ministry of Justice Impact Analysis of 2009 estimates that the changes will result in a further 10-20 murder rather than manslaughter convictions per year.

\textsuperscript{93} There was also much debate about the application of a reasonable person test in Provocation cases (Law Com No. 290 at § 3.31 and 3.113 (Commission, 2004). The reasonable person test is not part of the new defence, though there is a replacement objective test that a person of the defendant’s sex and age, with a normal degree of tolerance and self-restraint and in the circumstances of the defendant, might have reacted in the same or in a similar way to the defendant.
objective test” (Morgan, 2013, p. 120); that is, the question of whether or not the
defence of provocation should be available should be based solely on “whether
the defendant [had] acted in accordance with modern-day standards”, focussing
“strongly, if not solely, on an objective comparison of the act of the defendant with
the social views of the objective citizen” (Morgan, 2013, p. 120).

However, these recommendations were certainly not adopted in the reformation
of the defence. The requirement of loss of control was given such prominence
that the reformed defence was even entitled ‘Loss of Control’. Far from omitting
this purportedly problematic concept, the new defence has the concept at its core.
I will argue that the new defence has actually provided come clarification of the
concept of loss of control, and that it should be interpreted as the loss of the ability
to ‘restrain’ oneself form performing a specific action, because the question of
whether or not one should perform that action is beyond the reach of reason. This
endows the new defence with a very different ‘spirit’ than that with which it
originally developed, but, rather than entailing, as Chris Morgan has argued, that
“a loss of self-control, defined scientifically, can never really occur” (2013, p. 122),
it in fact reflects the evidence from cognitive science which, as I have argued
throughout my thesis, supports a conception of practical rationality as reason-
responsiveness.94

1.1 The new defence

The new defence of Loss of Control states that:

Where a person (“D”) kills or is a party to the killing of another (“V”), D is
not to be convicted of murder if—

D's acts and omissions in doing or being a party to the killing resulted from
D's loss of self-control;

the loss of self-control had a qualifying trigger; and

94 Jeremy Horder (1992) dates the development of the defence from the seventeenth century,
when a violent response to significant provocation was not just excused, but regarded as required
of a man of honour. Such responses were not regarded as a loss of control as the result of
emotion, but rather as a response informed and controlled by a rational understanding of the
offensive nature of the provocation. In the eighteenth century, anger came to be seen as
incompatible with rationality, and the test of whether or not the defendant ‘lost control’ emerged.
(c) a person of D’s sex and age, with a normal degree of tolerance and self-restraint and in the circumstances of D, might have reacted in the same or in a similar way to D (Coroners and Justice Act, 2009, §54-55).

The subjective test of the provocation defence is preserved in subsection (a); that D lost control, and killed V because she lost control. Subsections (b) and (c) substitute for the objective test of the provocation defence. Subsection (b) requires that the loss of self-control had a ‘qualifying trigger’, which is defined as being either D’s fear of serious violence from V against D or another identified person (§55.3), or else,

a thing or things done or said (or both) which—

constituted circumstances of an extremely grave character; and

(b) caused D to have a justifiable sense of being seriously wronged (§55.4).

The potential application of the defence to killing as a result of fear (§55.3) is a new concept, and substantially expands the scope of the defence.\(^95\) The second sort of qualifying trigger (§55.4) reflects the original provocation defence and is that which I shall address in this chapter: killing due to a loss of self-control as a result of anger, which I take to be uncontroversially the emotion that would arise as a result of having been “seriously wronged”. There is nothing further in the statute that defines what constitutes a loss of self-control, and as yet there is little case law involving the new defence. However, we can look to provocation cases to shed light on the nature of the requirement of loss of control. The concept is controversial because, as the Law Commission noted, there is no clear consensus in the case law as to what loss of self-control is, yet there is evidence of what it is not, and, as I shall argue, this evidence combined with the evidence from cognitive science, leads to a coherent definition of loss of self-control as loss

\(^95\) This expansion is designed to address the potential gender bias of the provocation defence, which was thought to, in “elevat[ing] the emotion of sudden anger above the emotions of fear, despair, compassion and empathy” (The Law Commission Consultation Paper no. 173), be biased towards excusing men, who were more likely to kill out of anger, and exclude the phenomenon of women killing out of fear as the result of cumulative domestic abuse (Clough, 2010, pp. 123-4)
of the regulative guiding control that is characteristic of practical agency as reason-responsiveness.

2 The concept of loss of control in case law

The Law Commission declared that the concept of loss of self-control was a ‘judicially invented concept’ that ‘lacked a foundation in psychology’. In section 3 I will argue that in fact the concept is consistent with current cognitive psychology, but first we need to explore the concept that the Commission decided had been ‘invented’ by the judiciary, i.e. the concept of loss of self-control that has emerged from provocation case law. The point of the loss of self-control requirement, as the Commission noted, was that it limit the defence as to not be available to those who, in the words of the new defence, “acted in a considered desire for revenge” (CJA 2009, §54.4). Yet the qualifying trigger requires the defendant to have a justified sense of being wronged which caused her action which means that that the defendant’s action was motivated by a desire to avenge this wrong, i.e. a desire for revenge. The operative word in the limitation, then, is that of a considered desire for revenge. If a defendant had acted from a considered desire for revenge then, no matter her emotional state, she would not be able to appeal to the defence.96

The rationale of this, I propose, is that, as Marcia Baron suggests, ‘cold-blooded’ killings are considered more heinous than hot-blooded killings because the former reflect more fully the real character of the agent and indicate a “wicked heart” (2004, p. 366). Such characters need both reformation and removal from society for the sake of protection, whereas those who have ‘reasonable’ characters but display usual ‘human frailty’ do not; we do not need to reform those

---

96 Although an agent acting from a considered desire for revenge would not be able to appeal to the defence of Loss of Control, they may have ‘lost self-control’ in the relevant way in pursuing their intention, I think. For example, in Moby Dick Ahab seeks to avenge his being crippled by the whale to the point of severe irrationality, though his initial intention to do so was not formed in the heat of passion. “Oh! Ahab,’ cried Starbuck, ‘not too late it is, even now, the third day, to desist. See! Moby Dick seeks thee not. It is thou, thou, that madly seekest him!” (Melville, 1950, p. 560).
who are already reasonable. If a defendant’s action is merely a result of the impulsive, or as Jeremy Horder puts it, the “shoot from the hip”, side of human nature (2005, p. 124), there is nothing to ‘reform’. But, we might think, isn’t impulsivity a trait of character; are we not responsible for being the sort of people that do not automatically act upon our impulses? Yes, but this is consistent with the defence as its being only a partial defence means that it only goes so far as to mitigate the crime to voluntary manslaughter: the defendant is still held to be responsible for an unlawful killing, and is still punished, nearly always with a custodial sentence. In addition, the defence is not available to a defendant who was particularly impulsive, as the objective test of whether the (in the provocation case law) the ‘reasonable person’ or (in the new Loss of Control defence) a person of the defendant’s age and sex with a normal degree of tolerance and self-restraint might have done the same needs to be satisfied, and the overly-impulsive defendant will fail this test. The person who was particularly unimpulsive would not perform the killing in such circumstances and so if they did kill it would reflect a different aspect of their character – a more stable trait of seeking revenge, perhaps, and since they would fail the subjective test of their killing being the result of a loss of control, they too would not be able to successfully plead the defence. We do not need to reform those whom we already deem reasonable – i.e. those who successfully plead provocation or loss of control – but society may still see fit to incarcerate them in order to fulfil other aims of legally-sanctioned punishment. But those we deem reasonable, because they are more than usually impulsive, or vengeful, or whatever else, we do seek to reform in order to bring their characters more into line with that which we deem reasonable. As Hume wrote,

Men are less blam’d for such evil actions, as they perform hastily and unpremeditatedly, than for such as proceed from thought and deliberation. For what reason? but because a hasty temper, tho’ a constant cause in the mind, operates only by intervals, and infects not the whole character (1978, p. 412).

With this being the rationale of the requirement that the defendant ‘lost self-control’ in mind, I will now argue that certain things are clear from the relevant case law, namely what the concept of loss of control does not entail. The relevant loss of control is not of the body, but of ‘mastery of mind’, yet not to the extent
that the defendant did not know what she was doing, nor to the extent that she is no longer an intentional agent, and the killing must have resulted from a loss rather than a lack of self-control.

2.1 The type of control lost

One of the supposed problems with the concept of loss of control in provocation defences is that the agent isn’t required to have entirely lost all types of control, such that she has no physical control over her bodily actions, or no conscious awareness of her physical actions. That the defendant had any sort of control at all has been argued to make it impossible to define what it is for him to have lost control in the relevant sense. For example, Chris Morgan claims that

The mere fact that a defendant has performed an act suggests that in some way he was in control, facilitating the argument that a loss of self-control, defined scientifically, can never really occur while the defendant remains able to act (2013, p. 122).

However, this argument is mistaken: that an agent has control in some sense does not entail that the relevant loss of control cannot occur. We can certainly see from the relevant case law that the extent of lost control is quite specific, and not as woolly as to merit a complete abandonment of the concept altogether.

The type of control that is relevant to the defence is, quite clearly, not control of one’s body, but of one’s mind. The so-called ‘classic’ definition of provocation is found in R v Duffy [1949] as:

some act, or series of acts, done [or words spoken] ... which would cause in any reasonable person, and actually causes in the accused, a sudden and temporary loss of self-control, rendering the accused so subject to passion as to make him or her for the moment not master of his mind (at

97 An agent who had lost physical and/or conscious control would instead appeal to the full defence of Automatism. Automatism excludes responsibility by negating the actus reus; that is it is found that the agent did not actually commit the act, for she was unaware of her actions that constituted it. Thus a defendant who is charged with an offence committed as a result of their body doing something over which they have no control, as a result of something like an epileptic seizure, night terror or other involuntarily-induced state that causes a “total destruction of voluntary control” (A-G Ref (No. 2 of 1992), 1993), would plead Automatism and, as it is a full defence, would be found not guilty of an unlawful killing.
Self-control in this context is, then, *mastery over one’s mind*, but its loss does not involve *complete* loss over one’s mind to the extent that one is not really conscious of what one is doing. It is not the all-or-nothing state as was suggested by Morgan’s argument above. The case law has been quite clear about this. For example, Lord Diplock in Philips v R. [1968] stated that:

> Before their Lordships, counsel for the appellant contended, not as a matter of construction but as one of logic, that once a reasonable man had lost his self-control his actions ceased to be those of a reasonable man and that accordingly he was no longer fully responsible in law for them whatever he did. This argument is based on the premise that loss of self-control is not a matter of degree but is absolute: there is no intermediate stage between icy detachment and going berserk. This premise, unless the argument is purely semantic, must be based upon human experience and is, in their Lordships’ view, false. The average man reacts to provocation according to its degree with angry words, with a blow of the hand, possibly, if the provocation is gross and there is a dangerous weapon to hand, with that weapon (at 135).

It is not that once a person has been provoked, she then loses control completely, being no longer master of anything that she does. That might well sometimes happen, but that is not what is required for the jury to find that the defendant lost control in the relevant sense. A person might lose self-control in a heated argument to the extent that they would throw something, but not to the extent that they would throw the thing at the person with whom they’re arguing; a man might react to a perceived gross insult from his spouse to the extent that he would slap her, but not to the extent that he would strangle her; additional taunting might lead to a loss of control to the extent that he would strangle her, but not to the extent that he would strangle her child sleeping in the next room. The defendant loses control, not to the extent that she cannot but perform *any* action, but just to the extent that she cannot but perform the *particular* action that she does, namely

---

98 The requirement of the loss of control being ‘sudden and temporary’ has been omitted from the new defence, partly to counteract the arguably male-biased conception of anger of having to be so, allowing instead for ‘slow-burn’ anger, which is said to be more characteristic of women, in particular those women who have been subjected to domestic abuse and kill their tormentors as a result (Baron, 2004, p. 359ff).
killing the victim. Thus it is quite clear that, far from being vague as to the extent to which control is lost, the defence requires a loss of control to a specific extent: that to which the defendant could not control himself in regards to the specific act of the killing of her victim.

### 2.2 Loss of control does not imply not knowing what one is doing

In *R v Richens* [1993] the defendant had killed a student who had raped his girlfriend two weeks previously. The student taunted the defendant about the rape, constituting provococation. The defendant killed the student in response to this taunting, and then the defendant and his girlfriend buried the body and lied to the police when questioned, which was held to constitute evidence that he knew what he was doing. The trial judge directed the jury that the loss of self-control required that the defendant *did not know* what he was doing:

> The law obviously does not excuse someone who may kill someone because he has lost his temper. All of us lose our tempers, I suspect, quite frequently. It is not dealing with that sort of loss of control at all. It is dealing with a complete loss of control, to the extent where you really do not know what you are doing ... In ordinary language, we talk of 'seeing red', or 'going berserk'. It is that sort of loss of control that we are talking about (italics added).

Because there was evidence that the defendant did know what he was doing, this direction led the jury to reject the defence. However, the Court of Appeal held that the trial judge’s description of the question of loss of control was a misdirection, and that, as per Duffy, provocative conduct must lead to a loss of self-control which resulted in the defendant being able to *restrain himself from doing what he did*, but *not* so far as not *knowing* what he was doing. Thus loss of control does *not* imply that the defendant did not know what she was doing, and the conviction for murder was quashed and substituted with manslaughter.

### 2.3 The agent is still an intentional agent
So far we have seen that case law does shed light on the concept of loss of control, in that the extent to which the defendant lost control is to the extent that she will perform, not just *any* action, but the action of killing the victim, and that the loss of control is in regards to what the defendant *does* and not what she *knows* she is doing. Further, case law dictates that the loss of control is in respect to the action of killing the victim, not in respect to the *means* by which the victim is killed. In *R v Thornton* [1992], for example, that the defendant went into the kitchen to arm herself with a carving knife which she spent half an hour sharpening, was held to not be evidence that counted against her having lost self-control. Thus, as Richard Holton and Stephen Shute maintain,

> [w]ithout self-control we can still have an intentional agent, one who will, for instance, make reasoned choices about the means of effecting the killing: she will be more likely to pick up a hammer than a cushion to use as a weapon, more likely to strike the provoker than an innocent passer-by (2007, pp. 61, italics added).

The defendant is still capable of “rational forward-planning”, but this is not, as Jeremy Horder thinks, “inconsistent with the spontaneity of response at the heart of a plea of provocation” (2005, p. 125), because only mean-ends planning as regards to the killing of the victim is held to be consistent with losing self-control. Planning the means to other ends, such as avoiding being caught, is not. For example, in *R v Ibrams and Gregory* [1981] it was held that the defendant’s putting on gloves to avoid leaving finger prints was evidence that counted against his having lost control. The defence recognizes, as Amanda Clough puts it, “human frailty, as passion aroused in the provoked killer takes away his ability to reason” (2010, p. 118), but it does not require that his ability to reason *per se* has been taken away, not even that his ability to reason *about* the killing that has been taken away. What has been taken away is his ability to reason about *whether* to kill, not *how* to kill.

### 2.4 Loss of control rather than lack of control

It is important to the rationale of the defence that the agent *lost* control rather than *lacked* control in the first place, for, as Joshua Dressler (1988) argues, the point of the defence is that we partially excuse those who had the *capacity* to control
themselves, yet lacked the opportunity to do so because of the provocation. The defence is not supposed to be available to defendants who ought to have been able to control themselves; it is a “concession to human frailty” (R v Smith, 2000), not to the frailty of the individual, and the objective requirement of the defence ensures this. The cause of the defendant not being able to control himself needs to be the provocation, not the agent’s own lack of character; not only must the provocation cause the inclinations that motivate the killing, it must also cause the agent to not be able to control himself in regard to those inclinations.

Holton and Shute put this in terms of whether the defendant’s self-control was ‘undermined’ or merely ‘overwhelmed’. If self-control is merely overwhelmed by a provocation then the defendant,

lacked sufficient self-control to handle the provocation: it simply gave rise to violent inclinations that her self-control was unable to restrain, inclinations that therefore moved her to perform the violent act (Holton & Shute, 2007, pp. 58, italics added).

Such a defendant would not be able to rely on the defence, since the defence requires that a ‘person of the defendant’s sex and age, with a normal degree of tolerance and self-restraint’ might have acted in the same way; thus the defence rules out the defendant who kills as a result of a lack of self-control. However, if self-control is undermined by a provocation then the agent did initially have sufficient self-control to handle the violent inclinations and the provocation acted in two ways: as well as giving rise to the inclinations, it also undermined the self-control that would otherwise have restrained them. Without the latter, undermining, effect, the agent’s self-control would have been sufficient to prevent the violent response; with it, it was not (Holton & Shute, 2007, p. 57ff). The provocation was therefore responsible for the absence of self-control. This is consistent with the rationale for the defence being a concession to normal human frailty, unavailable to irascible agents.

3 Loss of self-control as the question of whether or not to kill having been settled in the affirmative, out of the reach of reason
From case law we can see that loss of control is not entire, in that it does not imply loss of physical control over one’s body, nor lack of conscious knowledge of what one is doing. The agent is still intentional in regards to how to kill, but has lost control with regards to her ability to refrain from acting upon the specific inclinations aroused by the provocation: to kill. Further, the agent must have had sufficient control in the first place such that the provocation caused not just the inclinations, but was also responsible for the fact that the agent’s usual self-control was undermined. This all supports my proposal that what it is to have lost control in the context of the defence is for the question of whether or not to kill the victim to have been settled in the affirmative, out of the reach of reason. I shall now argue that this definition, far from being immune to scientific explanation, is consistent with the evidence from cognitive science that I have appealed to throughout my thesis as supporting a reason-responsiveness conception of practical agency.

That the provocation gave rise to violent inclinations is consistent with my picture of emotions as tracking reasons. In Chapter 5 I argued that the reasons that emotions track are the relevant emotion-proper properties. The relevant emotion-proper properties in provocation cases will include those sorts of reasons which, in the seventeenth century, were seen as quite clearly things which would justify the sort of anger that would give rise to a desire to retaliate in a violent fashion, such as an affront to one’s honour, and in particular to sexual infidelity. The reformed defence states that ‘qualifying triggers’ are things done or said which constitute circumstances of an extremely grave character and caused the defendant to have a justifiable sense of being seriously wronged, thus the sorts of things which would satisfy these requirements will be the sorts of relevant emotion-proper properties tracked by the defendant’s anger. These reasons will

99 The defendant in fact need not have an intent to kill for a charge of murder to be brought; she need only to have intended grievous bodily harm. However, for the sake of simplicity I summarise this as an intent to kill.

100 The reformed defence looked at first to specifically exclude sexual infidelity as a qualifying trigger: §55.6(c) states that “the fact that a thing done or said constituted sexual infidelity is to be disregarded” (CJA 2009). However, already in R v Clinton [2012] this has been qualified such that if other factors count as qualifying triggers then sexual infidelity may be taken into account in assessing whether things said and done amounted to the “circumstance of an extremely grave character and gave [the defendant] a justifiable sense of being wronged” as per §55.4.
justify the defendant’s anger, and thereby justify her *desire* to violently retaliate, though of course this does not entail that her *acting* upon that desire is justified, for all of the reasons that count against killing outweigh those that count in favour of it.\(^{101}\)

What is more controversial is the claim that the provocation *undermined* the defendant’s self-control. My proposal is that what it is for the provocation to have undermined the defendant’s self-control is for it to have given rise, not just to *inclinations* towards killing the provoker, but to an *intention* to kill the provoker, an intention that is then *immune* to the reflective self-monitoring capacities of the agent, such that she is not acting as a reason-responder but as a reason-tracker. The provocation undermines her capacity for rational guidance and the ‘control’ that is lost is the regulative guidance control that is characteristic of rational agency.

### 3.1 The provocation gives rise to an intention, not just an inclination

The provocation defence is not available to a defendant just because she had violent inclinations, even violent inclinations that were in the circumstances understandable. As was said in *R v Richens*, the law does not excuse a killing just because the defendant lost her temper, for we all lose our tempers; that the victim had done something that provoked a loss of temper is not sufficient to make available the defence of provocation. The defendant has to do more than just demonstrate that she was provoked and that the provocation gave rise to violent inclinations, for the law expects that we ought to restrain ourselves in the face of anger and passion “which a man ought to keep under and govern” (Reg. v. Oneby, ‘727), for, as J. Coleridge ruled in *Reg v. Kirkham* [1837],

> [t]hough the law condescends to human frailty, it will not indulge human ferocity. It considers man to be a rational being, and requires that he should exercise a reasonable control over his passions (115, 119)

---

\(^{101}\) If it was held that her killing was justified then she would not be appealing to this defence, but to that which would reflect the reasons for which her killing was not unlawful, for example Self-Defence.
Thus merely that the provoker’s conduct instigated violent inclinations in the
defendant is not enough to warrant an appeal to the defence. Citizens are expected to control their behaviour and not act upon every inclination that they have. The defendant who lacked sufficient control to handle violent inclinations cannot be excused by the defence, for she should have acted reasonably – i.e. non violently – despite her desire. Thus the provocation must have given rise to more than just mere inclinations. Inclinations do not compel us to act upon them, we can still choose how to act; we can form an intention to act counter to these inclinations, and that is what the law expects of us.

However, I propose that in cases of Loss of Control the defendant cannot choose to act counter to her violent inclinations because she cannot form an intention to do so. The provocation has given rise not just to an inclination towards violence, but to an intention to act violently, and this intention has already settled the question of whether or not to attack the provoker, thus the question of whether or not to kill is not up for deliberation. The provocation gives rise to the intention before the defendant is even aware that there is a decision to make, thus she never has the opportunity to decide whether or not be ought to kill.

This is consistent with the case law as outlined above: the loss of control is in relation to the specific action of violently assaulting the provoker because the intention that it has given rise to is to violently assault the provoker, but control remains as to how to commit the assault. The defendant hasn’t lost the ability to reason, but does not reason about whether or not to attack the provoker. She can engage in conscious deliberation about how, and she can know what she is doing, but she cannot engage in conscious deliberation about whether or not to attack. Her intention, therefore, is, rather than that formed as a result of conscious deliberation, a direct result of the provocation, which has bypassed reasoning. This is consistent with the recent ruling in R v Clinton [2012], one of the few cases as yet to involve the new defence of Loss of Control, that “the greater the level of deliberation, the less likely it will be that the killing followed a true loss of self-control” (at §10), and the explicit exclusion in the statute of the defendant acting out of a “considered desire for revenge” (CJD §54.4, italics added). The desire
has not been ‘considered’; deliberation about whether or not to act upon it has not occurred, and, as I shall argue in the section 3.2, cannot occur.

My proposal is also consistent with the evidence from cognitive science that I have referred to throughout my thesis. Recall that Dual Processing accounts propose that humans have two distinct mechanisms for thought processing that operate largely independently of each other. System 1 is fast, automatic and non-conscious; System 2 is slow, controlled and conscious, with emotion being associated with System 1 and deliberation with System 2. Both System 1 and System 2 generate solutions to problems, but System 1 does this more quickly than System 2. My proposal is that System 1 generates the intention to kill the provoker before System 2 is activated, in the same way that System 1, as I described in Chapter 5, generates the intentional action of running away when confronted with a threat before the agent has conscious awareness of it. This intention can be either an ‘intention in action’ in that it is acted upon immediately, or it can be a ‘future-directed intention’ in that it is the sort of ‘stable’ and ‘controlling’ intention to act when the time to do so comes that was discussed in the last chapter in relation to weakness of will. System 2 does not have a chance to be activated regarding the question of whether or not to kill the provoker, because the matter has already been settled with an intention to do so generated by System 1, thus the question of whether or not to kill the provoked isn’t up for deliberation. System 2 is, Dual Processing Theories hold, “mobilized when a question arises for which System 1 does not have an answer” (Kahneman, 2011, p. 24), but in these cases of Loss of Control, System 1 does have an answer, at least to the question of whether or not to respond violently to the provocation (i.e. ‘yes’) so System 2 is not activated. System 1 may not have an answer to the question of how one might best kill the provoker, and this is where System 2 would come in, as in R v Thornton, where the defendant sharpened the knife in the kitchen for half an hour. But in relation to the question

---

102 That it can be a future-directed action as well as an intention-in-action is a feature that distinguishes the new defence of Loss of Control from the old defence of Provocation, since, as discussed earlier in this chapter, the original defence included the requirement that the loss of self-control be ‘sudden’, a requirement which has been deliberately removed in the reformation of the defence.
of whether or not to kill the provoker System 2 is not mobilized and, as I shall now argue, it remains immobile in relation to such because, as Holton and Shute also recognize, anger can undermine the very thing by which it is supposed to be controlled (2007, p. 58).

3.2 The intention is out of reach of reason

As Dressler argued, a provocation excuses the defendant because it denied her the opportunity to exercise her capacity to control herself. Not only does the provocation generate an intention to kill the provoker, but this intention is put out of the reach of reason: System 2 is not activated as regards the intention, so the question of whether or not to kill remains outside the reach of the defendant’s usual powers of rational guidance control.

I have already argued that System 1 has generated the intention to kill the provoker, but of course this does not necessarily mean that the agent is then bound to follow through with the intention. As I argued in the previous chapter, intentions are designed to be stable and controlling but they should not be completely immune to reassessment. If the agent is given good grounds to reconsider her intention then she ought to do so. I propose that what happens in cases of Loss of Control is that the intention is immune to revision in the light of good reasons, because the provocation not only gives rise to the intention to kill, but it also distorts the agent’s perspective such that what she would ordinarily realise as good reason to reconsider her intention to kill, i.e. reasons that she has that count against killing the provoker, are not seen as such. To borrow Peter Goldie’s phrase, her anger skews her epistemic landscape (Goldie, 2008). Evidence which she might otherwise, through calm and deliberative thinking, take to count against her killing her provoker she now ignores, or even takes to be confirmation of the justification of her desire to violently retaliate. It is a similar effect to that by which the xenophobe ignores the reliable data and latches on to the stories in his copy of the Daily Mail:
We might suggest that the xenophobe buy a newspaper which we consider to be more reliable, but ... so far as he can tell, it is only the Daily Mail that has the courage to say how bad things are, and our telling him otherwise just shows how little we know (Goldie, 2008, p. 160).

By skewing the defendant’s epistemic landscape, the provocation affects her ability to respond to her reasons. What would usually be seen as reasons are either overlooked or severely undervalued in terms of their normative force, thus she cannot respond, or respond properly, to these reasons. The provocation thus undermines her self-control is by undermining her usual ability to respond to her reasons.

Recall that according to Dual Processing theories System 2 is activated when “an event is detected that violates the model of the world that System 1 maintains” (Kahneman, 2011, p. 24). The effect of the provocation is to skew the defendant’s epistemic landscape such that events (that is reasons tracked by the agent’s reason-tracking mechanisms) that would usually be interpreted as conflicting with the current view of the world (the view in which killing the provoker is what the defendant ought to do) are not detected as violating the world view. That is because the emotion to which the provocation gives rise – i.e. anger – has a powerful distorting effect upon how the agent sees the world. The distorting effect is so powerful that it is immune to pure reason. Only a similarly powerful effect can correct the distortion.

That this ‘judicially invented’ concept was, in fact, a recognition of the distorting effect of anger can be seen in the history of the development of the provocation defence: it was grounded in the recognition that defendant had reasons to kill her victim but that anger “modifies perspective and so helps to explain how the defendant could have viewed a wholly inappropriate act of violence as broadly proportionate” (Gough, 1999, p. 481). This theory of the effects of anger finds support in cognitive science. The so-called cognitive approach to “hot” violence, that is, violence that is associated with anger, maintains that the violently hostile person has strong negative biases towards the victim that leads to their genuinely believing that they have a legitimate grievance that justifies their violence (Beck,
2002, pp. 211-2). The victim is seen as dangerous and as needing to be isolated, punished or eliminated, and this leads to the person seeing revenge for damage they believed has been sustained, or a pre-emptive strike to forestall such (*ibid.*, p.212). Aaron Beck, who endorses this model, argues that we have natural inhibitions against killing that provide restraints against even the strongest of urges to kill, that operate automatically unless they are attenuated by the agent’s beliefs. The strong negative biases generated by extreme anger can attenuate these restraints, because they support the belief that violence in certain situations is desirable, even noble (*ibid.*, p214).

This ‘system of idealized hate’, as Beck calls it, is certainly consistent with the history of the development of the provocation defence as excusing men who had killed as a result of believing their honour to have been violated (Morgan, 2013, p. 120), as a violent response was actually *expected* from a ‘man of honour’ in response to such an affront (Horder, 1992). In the seventeenth century such a response was seen as *reasonable*, that is, that there were good reasons that counted in favour of responding violently to a serious-enough provocation. We may no longer think that such reasons are good enough to justify killing, but our emotion of anger still *tracks* such reasons. Subjected to rational criticism we can see that provocation does not, in fact, justify a violent attack, but if our anger tracks such reasons and presents them to us as justifying violence, as the cognitive approach to hot violence maintains, in the ‘heat of passion’ they appear to be good enough. Unfortunately, anger can present such reasons as being so strong that they are immune to rational criticism. Certain provocations might very well count in favour of a violent response, but in our civilized society they are not strong enough to justify acting so because there are competing reasons that count against reacting violently which outweigh those that count in favour of doing so. The problem is that these competing reasons are either overlooked or undervalued by the provoked and angry agent because the strong negative biases that are generated *skew* the agent’s epistemic landscape such that reasons that count in favour of violence are given too much weight, facts are interpreted as fitting in to the system of idealized hate too readily and thus are mistakenly seen as lending additional normative force to the skewed view, and reasons that count
against fail to register with the agent because the reason-tracking mechanisms that would ordinarily latch on to them are disabled by being overwhelmed by the motivational force of the provoked anger.

### 3.2.1 Skewing the epistemic landscape

Conscious deliberation, as a System 2 process, is activated when it detects something that violates the model of the world that System 1, in this case the provoked anger, maintains. In cases of Loss of Control System 1 maintains its initial presentation of the world – that of the provoker deserving to be violently attacked – much more resiliently than it ought to, and thus System 2 is not activated as readily as it ordinarily would be. One way in which System 1 has done this is by jumping ahead of System 2 in producing an intention to act before System 2 has access to the reasons for which the intention was formed, and, importantly, it masks that it has done this by skewing the agent’s epistemic landscape such that she overlooks or mistakes as supporting those reasons which would ordinarily alert her to the fact that there was ever a choice to be made. The intention is formed, and then the reason-giving aspects of the situation are interpreted in whatever way makes them consistent with the situation as would justify that intention, before System 2 is activated, thus once System 2 is activated, it is presented with a coherent world view in which the intention to kill is unquestioned. The reasons that would ordinarily be taken to contradict this world view and the intention to kill have already been interpreted in such a way as to be consistent with the world view and intention to kill, thus there is no reason for the defendant to think twice about acting on the intention. As Kahneman puts it, a definite choice was made, but we did not know it; “[o]nly one interpretation came to mind, and you were never aware of the ambiguity” (2011, p. 80). In Loss of Control cases the interpretation of the situation of being one which called for a violent response came to the defendant’s mind, as this is the world view that would make sense of her extreme anger, without her being consciously aware of any other possible interpretations. Already, then, her anger has masked its effect by generating in a context with which her anger is consistent, thus avoiding the triggering of System 2 in response to a potential inconsistency.
Once this model of the world - as one in which her extreme anger is justified - is in place, it will only be questioned by System 2 if something is detected which violates it. As I argued in Chapter 6, this would happen if a conflict between the deliverances of two reason-tracking mechanisms were detected, as System 2 would be activated to resolve the conflict between them. For the intention to kill to be immune to reassessment then, such conflicts must not activate System 2. This is, I propose, precisely what happens when a person ‘loses self-control’ in the context of the defence. Her emotion of anger misleads the defendant by not only presenting a biased view of her reasons for action, but, and this is what makes her intention immune to reason, masking the fact that it has mislead her in this way. Not only does the emotion ‘by-pass’ conscious deliberation and give rise to the intention to kill the provoker, but that it has done this goes undetected by her deliberative faculties and so is immune to correction by reasons. System 2 should step in when it detects that the deliverances of System 1 are not in accordance with reason, but her emotion masks its misleading representation of her reasons, such that System 2 does not detect errors, and since System 2 does not question the ‘facts’ as presented by System 1, it does not step in to correct them. In misleading her about her reasons, and then disguising this from her self-monitoring mechanisms of System 2, her emotion undermines the defendant’s ability to act rationally.

3.2.2 Misleading presentation of our reasons

The defendant’s anger gives her a false representation of her reasons. Even if her emotion is justified, it may be giving her a false representation of the strength of what really are reasons, thus representing them as providing a great enough normative force to outweigh any competing reasons, or it may be preventing her from seeing her competing reasons, or their real normative force. Recall from Chapter 5 that emotional arousal reduces the range of information to which we are capable of paying attention. Emotions make salient those features of our environment that are emotion-proper properties, at the expense of non-emotional
features, which will lead to false-positives: because the decision is based upon minimal evidence, any evidence that runs contrary to this assessment will be ignored. System 1 has, prior to our becoming consciously aware, reduced the range of reasons to which we are capable of becoming consciously aware to those that are salient to the emotion. As Jonathan Haidt puts it, ‘the reasoning process is more like a lawyer defending a client than a judge or scientist seeking truth” (Haidt, 2001, p. 820) in that the only evidence that System 2 can consider is that which System 1 has already presented as being only that which needs to be considered. In Loss of Control cases that evidence is that which justifies the intention to kill the provoker, with the evidence that contradicts this view having already been discounted before the defendant is aware that this has happened.

The evidence that is ignored is that which counts against killing the provoker. Anything that violates the view of the provoker as presented by the agent’s anger has the potential to trigger System 2 into deliberating about whether or not the intention to kill ought to be revised, but, I propose, so long as the defendant is in the state of ‘lost control' this will not happen. If these reasons do trigger System 2 then the agent will ‘regain’ control. A number of things might do this, for example a photograph of the provoker’s child or a plea from the potential victim, but as an example I shall concentrate reasons for her own welfare that count against her killing her provoker, such has been declared to be evidence that counts against the finding that a defendant lost control; for example, as mentioned earlier, in the case of R v Ibrams, putting on gloves to avoid leading fingerprints.

If a defendant showed evidence of considering how to avoid being caught, as putting on gloves to avoid leaving fingerprints was held to be, then this is evidence of her having access to reasons that count against her killing her provoker. This would explain why something like the presence of a police officer would usually be enough to ‘enable’ a defendant to get control of himself and restrain from acting upon her intention: for a person to kill in the presence of a police officer, we might think, her loss of control would have to be extremely severe, but lower hurdles may also prove a bar to action. Why would these facts – the pleading, the photograph, the presence of a police officer – enable someone to restrain
himself once he’d already ‘lost control’? They provide the agent with access to her reasons to not kill. She already had these reasons, but until she perceived these things which made these reasons salient, she could not act for them, since to act for a reason an agent needs to have access to it.

Such access, as I argued in Chapter 1, does not need to be conscious, and indeed in cases of Loss of Control such is unlikely to be. The reasons that count against her killing are already ‘out there’, but her emotion of anger is preventing him from gaining access to them. What could give him access to them? There would need to be a counteracting effect to the prior distortion, and in the case of reasons to avoid being caught, a relevant effect would be the competing emotion of fear. Although the reasons that count against her killing are hidden from System 2, they can be tracked by the competing System 1 mechanism of a different emotion. The presence of a police officer would make salient the fact that there is a chance of being caught, prosecuted, and punished, which are reasons that counts against acting on the intention to kill. That one’s welfare is potentially threatened is a likely emotion-proper property of fear, thus fear will be triggered by the (non-conscious) perception of these reasons.

The competing emotion of fear might be strong enough to counteract the distorting effects of anger. Evidence from cognitive psychology suggests that whether or not a strong-enough emotional reaction will be incited will depend upon how vivid the reasons for fear are to the defendant. According to psychologists Richard Nisbett and Lee Ross, information is ‘vivid’ when it is,

likely to attract and hold our attention and to excite the imagination to the extent that it is (a) emotionally interesting, (b) concrete and imagery-provoking, and (c) proximate in a sensory, temporal or spatial way” (Nisbett and Ross 1980, 45 cited in (Goldie, 2008, p. 152)).

Nisbett and Ross argue that an event will be more ‘emotionally interesting’, and therefore more vivid, if one is directly involved, or if one already has strong feelings towards those who are involved. An event will be more ‘concrete and imagery-provoking’ the more emotional detail is available: close-up television
images of a natural disaster will be more vivid than a newspaper report. An event will be the more ‘proximate’ the more recently it occurred, and the more geographically close its location.

The sort of events that constitute provocation would be exactly the sort that will attract our attention and excite the imagination as would be predicted by such a theory. It would be hard for a defendant to have the required ‘justifiable sense of being seriously wronged’ by an event that involved neither himself nor someone for whom she already had feelings, was lacking emotionally-arousing detail in the way that an objective reporting of fact in a newspaper report might, and that had happened a long time ago on the other side of the world. However, the theory would also account for why facts such as the presence of a police officer would be likely to enable the defendant to regain control. The reasons that count against her acting on her intention are already ones that she has, and to which she would, ordinarily, be able to gain access to via conscious routes, but if they are represented ‘vividly’ enough they can counteract the distorting effects of anger by attracting and then holding the attention of the defendant, thereby getting round the distorting effect of her anger.

Emotions restriction of information to which the defendant can pay attention; her anger restricted her attention to those reasons made salient by her anger that supported the view of her victim as deserving of a violent retaliation; if vivid enough, the information that represented her reasons against killing could attract her attention away from the emotion-proper properties of her anger on to the emotion-proper properties of her fear. The appearance of a police officer would be a very vivid event indeed to an agent about to kill. The reason to which it gives her access – that of her own welfare – is one that she already has, and one to which other information already points, but because the other sources are not so vivid, they do not attract her attention. She already knows that police officers exist, but her other sources of this information lack the emotional interest, concrete, and proximal properties of the sudden appearance of a police officer in front of her. How vivid the information will need to be to attract her attention away from the focus of her anger will depend upon how strong the distorting effect of
her anger is. If it is strong enough, it might not be able to be overcome at all; if
great perhaps the officer will need to be directly in front of her, and would be
missed if in the periphery of her vision; if mild perhaps the photograph of an officer
on the front of a nearby newspaper would suffice, or even the glimpse of a record
of Sting’s band. Such information could correct for the dominance of the vivid
information by focussing the agent’s attention on the other reason-giving features
of her situation. It would, unconsciously, be following Peter Goldie’s advice that
we “increase the salience of the paler information, and thus to increase its
emotional import and motivating power” (2008, p. 161) in order to attempt to
counteract the epistemically-skewing effect of emotions, and might also enable
the defendant to follow Goldie’s other suggestion that we “stop and think” (2008,
p. 162), to force oneself to engage in further deliberation before acting. Both of
these methods enable the defendant to have access to more of her reasons, and
thus enable her to respond to them.

Fear is an especially relevant competing emotion in provocation cases, as it
plausibly explains the marked differences between the ways in which women and
men typically respond to provocations. Horder argues that the absence of fear
felt by men who violently attack women explains why such cases are so relatively
common. He argues that,

The taken-for-granted assumption that there will be no equal and opposite
retaliation, or otherwise threateningly negative, response from a (female)
provoker is what nourishes some men’s sense that they can ‘permit
themselves’ to lose self-control when angered (2005, p. 128).

I do not endorse any connotations of intentionality that might be implied by the
choice of the word ‘permit’, but it certainly seems plausible that had the provoker
been male there would have been more vivid reminders of the potential negative
consequences for the defendant’s welfare that would have attracted and held his
attention such that she was able to restrain himself. Horder further argues that
fear explains why women are far less likely to lose self-control in the face of
provocation. A woman’s anger is likely to be tempered by the emotion of fear of
provoking the potential victim’s own reaction, and the
“mixed character of the defendant’s emotional response to a taunt … is liable to ensure that she does not lose self-control in an expressive way – whether or not immediately – of any retaliatory impulse she may be experiencing” (2005, p. 128).

In the case of women who experience provocation grave enough to lead to retaliatory action, such is likely to be taken “only if and when the provoker is ‘off his guard’” (ibid., p.129). A provoker off his guard will present less threat and so incite less fear, and thus reasons for her own welfare that count against her retaliation will be less vivid and less likely to capture her attention away from that upon which her anger is focussed.

When an agent loses control she loses the ability to deliberate about whether or not to act upon the intention she has formed. In the defence of Loss of Control, this intention must have been formed without any considered deliberation in the first place, but we can apply the same to cases of agents who pursue intentions that were initially formed as a result of practical reasoning, even good reasoning, in the face of competing reasons to abandon their intention. Once an agent has lost control in the relevant sense, it takes more that ‘mere’ reason for her to regain it. The reasons that provide the good grounds for her to reconsider her intention need to be vividly represented, in order to capture her attention, and a competing emotion, such as fear, might be able to do this. Thus although emotion was responsible for her initial loss of control, emotion is also what enables her to regain control. Other methods that have been traditionally dismissed in the dominant philosophical literature as ‘irrational’, might also be effective in correcting misleading saliences and contributing to an agent’s being able to respond to her reasons such that she is practically rational. Rhetoric, for example, can, contra Bernard Williams (1980), enable correct deliberation by effecting a non-rational route of access to reasons an agent had but to which she was previously ‘deaf’.

103 John McDowell (1995) presents a very convincing argument in defence of ‘non-rational routes’ as a means to being brought to see reasons of which one was previously ignorant, against Williams’ claim that such a transition can be made only via a “sound deliberative route”, such that rhetoric could in fact enable an agent to “consider the matter aright”. For a specific example of the effect of emotions in rhetoric, see cognitive linguist George Lakoff’s (1993) arguments for the centrality of metaphor to human thinking. Lakoff argues that emotions are central to many metaphors, and that hundreds of emotionally-laden metaphorical links are made in childhood.
4 Conclusion

The reformed provocation defence of ‘Loss of Control’ retains the requirement that a defendant lost self-control, despite a general consensus in academic discussion of such, and a recommendation by the Law Commission, that such a concept was a ‘judicial invention’ that had no scientific basis. In this chapter I have argued that on the contrary, the concept of loss of control, far from being a ‘valiant but flawed’ invention by the judiciary, reflects a phenomenon that has a sound basis in both science and philosophy, the coherence of which can be seen in the light of my thesis of the role of emotion in practical rationality. What it is for a person to ‘lose self-control’ is for their action to be beyond the rational guidance of their agency as a self-regulative reason-responder. The rational agent guides her actions in line with her reasons, and thus acts for her reasons, because she has the capacity to step back from an impulse to action to reflect upon it, and thus guides her reason-tracking mechanisms in line with her reasons. When she loses control, she has not got this capacity and thus is not capable of guiding her action for her reasons. It is not merely that she fails to do what she has most reason to do for those reasons (it is not that she is practically irrational), but that she cannot guide her actions in accordance with her reasons, for her capacity to, as Karen Jones puts it, “reflect upon the deliverances of her reason-tracking mechanisms” has been disabled, in regards to the specific question of whether or not she ought to $\Phi$, by an emotion which, not only overwhelmed her capacity for self-control as reason-responsiveness, but undermined it by making emotion impervious to the very thing to which it was meant to be guided.

These ‘embodied’ metaphors are exploited in political rhetoric, yes, but they are crucial to human thinking such that to reject them as irrational on the basis that they are open to exploitation is to be wedded to a false idea of human rationality. “Real reason”, he argues, is inexplicably tied up with emotion; you cannot be rational without being emotional. False reason thinks that emotion is the enemy of reason, that it is unscrupulous to call on emotion. “Rational” decisions are based on a long history of emotional responses by oneself and others. Real reason requires emotion” (Lakoff, 2010).
Bibliography


Holton, R., n.d. Inverse Akrasia and Weakness of Will. _n.p._.


James, W., 1884. What is an Emotion?. _Mind_, April, 9(34), pp. 188-205.

James, W., 1979. _The Will to Believe and Other Essays in Popular Philosophy_. Harvard University Press.


*R v Duffy* (1949).


*Reg v. Kirkham* (1837).


