Shame: In Defence of an Essential Moral Emotion

Shame: In Defence of an Essential Moral Emotion

By Daniel James Turnbull

A thesis submitted for the degree of Doctor of Philosophy (Philosophy)

Birkbeck College, University of London

I hereby declare that all the work presented in this thesis is my own, except where otherwise indicated.

Daniel James Turnbull

ABSTRACT

I argue that shame is an essential moral emotion, and that the capacity to feel shame is vital to allow us to pick out certain types of moral value. I do this by sketching out a general role for moral emotions, distinguishing shame from other moral emotions, notably guilt, and then arguing that shame has a distinctive role to play as a moral emotion that cannot be played by guilt.

By marking out a key role for the emotions in moral life, I am able to address two key concerns about moral judgement. First, I am able to explain how we overcome frame problems, allowing us to notice and appropriately conceptualise moral concerns, against a backdrop of everyday life. Second, I can give an account of the apparent intrinsically-motivating nature of moral judgements.

Shame, in central cases, is based on self-assessments of inadequacy; we judge ourselves to be less than we should be. This is contrasted with guilt, which centres on judgements of transgression against moral norms. It is also contrasted with embarrassment and humiliation, neither of which are primarily moral emotions.

Shame has a distinctive role to play as a moral emotion. It is capable of picking out cases of moral value that guilt cannot; in particular, supererogatory value and cases of wrongdoing by collectives, in the absence of individual culpability. Pace the claims of numerous psychologists and philosophers, shame is not necessarily a dangerous emotion; rather, only certain types of shame have the potential to do damage to those experiencing them. Situationist arguments threaten the role of shame as a moral emotion, by suggesting that there are no robust character traits; these claims are mistaken. Therefore, I am able to sustain the conclusion that shame has a vital role to play in moral life.

TABLE OF CONTENTS

PREFACE

A work of this nature inevitably has many intellectual influences, many of which no doubt evident from the text. Iris Murdoch, Michael Stocker, David Wiggins, John McDowell, Martha Nussbaum, Richard Wollheim and, in particular, Ronald de Sousa are among the philosophers whose published work has had the most profound impact on the shape of this thesis. However, while interaction with these thinkers has influenced the content of my work, it is membership of a living, breathing philosophical community that has provided its animating spirit.

The central manifestation of this effect has been through my primary supervisor, Miranda Fricker. Over the past five years she has read numerous drafts, and offered insightful comments and suggestions. Her patience, enthusiasm and judiciousness have improved this thesis in every way; quite simply, it would not exist in anything like its current form but for her guidance, and I am extremely grateful for it.

Thanks are also due to my temporary supervisors, Jen Hornsby and Sam Guttenplan, as well as my secondary supervisors, Sue James and Michael Garnett, all of whom made welcome and helpful comments, to the benefit of the final result. The wider postgraduate philosophy community at Birkbeck have also been incredibly supportive, not least through their probing questions at our graduate seminars and more informal discussions.

Going further back, I must also thank Peter Goldie, who sadly passed away recently, and Rai Gaita for their unique and valuable contributions to my intellectual life while studying for my MPhil at King's College London. I am also grateful to John Tasioulas for stimulating and encouraging my initial interest in moral psychology as an undergraduate in Oxford. Finally, my deepest gratitude goes to my wife, Maggie Boyle, for her support and understanding through the long process of writing this thesis.

The shortcomings displayed in this work are, of course, entirely my own responsibility. Without the influence of the people acknowledged here, however, they would be much more numerous and serious. I offer each of them my sincere respect, thanks and good wishes.

INTRODUCTION: PLACING SHAME IN MORAL LIFE

Shame is an emotion that provokes widely differing reactions. On the one hand, it is seen as a deeply damaging emotion, rooted in a profound sense of inferiority, which can lead to self-disgust, or even depression and withdrawal from others. Shame also has a troubling history of being used as a psychological weapon, particularly against out-groups in society. This view of shame has led some to suggest that the emotion ought to be eradicated from our emotional lives, usually at the expense of its more focused and 'civilised' cousin: guilt. At the same time, however, an accusation of 'shamelessness' indicates a significant character failing, and some have even argued that a rediscovery of personal shame is just what is required to combat some of the pervasive social ills of our time. These varying responses seem to have resulted in a deep ambivalence about the place of shame in both our individual lives and our society.

The aim of this thesis is to argue that shame deserves its place in our emotional repertoire, because of its vital role as a moral emotion. Shame can reveal moral values to us that would be invisible in its absence, and provide us with the motivation to honour and pursue them. While it is true that some forms of shame can be pathological and destructive, I shall argue that susceptibility to appropriate, well-focused shame remains essential to a well-developed moral sensibility. Despite its dark side, a world without shame would be one in which people have a diminished ability to grasp the full range of moral value. While shame may not be necessary in a world of angels, even an angel in our morally imperfect world would sometimes have need of the emotion.

To make this argument, I begin Chapter 1 by picking out some questions that any persuasive moral theory must seek to answer. Drawing on the work of Lawrence Blum, Iris Murdoch and Ronald de Sousa, I set out some important questions in moral epistemology, asking how we are able to pick out, and accord the appropriate level of salience to, all and only the morally relevant facets of a situation and how moral judgements are tied to emotion. One way of answering these questions, I suggest, is to integrate the emotions into our moral systems. I describe how this is achieved by the neo-sentimentalist approach, exemplified in its cognitivist variety by John McDowell and David Wiggins, and in its non-cognitivist flavour by Alan Gibbard and Simon Blackburn. After defending this approach from a fundamental critique offered by Shaun Nichols, I set out a key task that must be met by any proponent of a

neo-sentimentalist view: providing an answer to the 'conflation problem' by developing a framework that allows us to discern the appropriateness or otherwise of emotional episodes.

I broach this problem in Chapter 2, rejecting a traditional cognitivist view of the emotions, which sees them as (at least partly) constituted by beliefs, adopting instead a notion of emotions that involves seeing analogies between occurrent situations and certain paradigm scenarios, based on arguments from de Sousa. This yields a notion of emotional appropriateness as a matter of analogousness to a relevant paradigm scenario. With this picture of the emotions in place, I am then able to explain how creating space in our moral framework for the emotions allows us to address the epistemological and motivational questions set out in the preceding chapter.

In Chapter 3, I attempt to define the contours of an acceptable paradigm scenario for shame. First, I reject Richard Wollheim's view of shame as a 'radically heteronomous' emotion, arguing instead that in the central, developed case, shame is an autonomous emotion, based on the content of the subject's own self-directed judgements. Second, I argue that shame need not be 'globalised' so that the subject is ashamed of the entirety of his self; on the contrary, shame can be 'localised', and so focused on failings only in some specific aspect of the self. Finally, drawing on the work of John Rawls and John Deigh, I examine the content of the judgement that is central to shame, concluding that a broadly-drawn specification is in order: a judgement that the self is somehow substandard is the required content for a case of shame.

I go on in Chapter 4 to examine three emotions that are closely related to shame: embarrassment, humiliation and guilt. The first of these is, at root, a reaction to appearing in a way one does not wish to appear, while the second is a related emotion, with the added constraint that some other person is perceived as cruelly and deliberately putting the subject in this position. Finally, I examine the potential points of difference between shame and guilt that are proposed in the literature: social versus individual, different types of eliciting behaviour and action versus status of the self. I conclude that guilt is based on the perception of transgressions of moral norms through one's actions.

Chapter 5 sets out my central argument for shame as a moral emotion. I first set out a two-pronged case against shame playing this role. First, I consider arguments, mainly from the psychological literature, and particularly the work of Helen Block Lewis and Sandra Bartky, in which shame is painted as having significant dangerous and/or counterproductive effects on its subject. Second, I relate the argument put forward by Gibbard that shame is

superfluous as a moral emotion; appropriately felt guilt is sufficient on its own to pick out all cases of moral value. Taken together, these form a powerful case against accepting shame as a moral emotion, presenting the emotion as unnecessary and dangerous. Against these views, I argue that, pace Gibbard, shame can reveal to us areas of moral value that guilt cannot, notably in the spheres of supererogatory action and collective moral responsibility. I also argue that claims of the potentially damaging effects of shame are overstated, only properly applying to certain degenerate forms of shame, rather than the emotion in general. I conclude that we can and must accord shame the status of a moral emotion.

In the final chapter, I consider a powerful generic challenge to shame as a moral emotion: the situationism of Gilbert Harman and John Doris. By attacking the very notion of a general character trait, they potentially undercut the possibility of shame is a moral emotion; if shame involves the self-ascription of negative character traits, then if there are no such character traits then shame can never be appropriate. However, I argue, firstly, that Harman's more extreme form of situationism is unsustainable, and so poses no threat and secondly that Doris's more moderate situationism, insofar as it is acceptable, is insufficiently revisionary of our customary beliefs about character traits to rule out shame as a moral emotion – at least, on a proper understanding of shame.

Mistaken notions of what shame is or must be like have too often resulted in misconceived normative conclusions about that emotion; that it is a damaging or otherwise negative emotion, which we would be best advised to banish from our psychological repertoire. By dispelling some of these notions about the nature of shame, I am able to present a case for allowing shame to play a role as a moral emotion. Crucially, I argue that it is only through a properly functioning sense of shame that we can fully access certain types of moral value, and carve out an essential role for shame in moral life. In establishing these two conclusions, I hope to vindicate shame as a central and indispensable moral emotion.

CHAPTER ONE: THE EMOTIONS AND MORAL JUDGEMENT

Shame is an essential moral emotion; a world without shame would be a morally impoverished one, unless it also lacked the instances of moral disvalue that shame alone can allow us to perceive. In order to make the argument for this claim, I will essentially take it as read that the emotions in general have a central role to play in moral life. In this chapter and the next I perform the necessary groundwork of setting out what that role might be, and how the emotions have the necessary characteristics to play that role. This will provide me with the required basis for moving to the central concern of the thesis: characterising shame, distinguishing it from related emotions, and setting out a distinctive role for it in moral life. The purpose of these first two chapters is to set out a well-defined picture of the place of emotions in moral life, which will then provide the basis for investigating whether shame has an essential role within this. My aim here is not to convince the sceptical reader of the merits of this picture, but rather to stake out the ground on which the real debate over the particular role of shame can be played out in the latter part of the thesis.

The structure of this chapter is as follows. First I set out two key questions that any ethical theory must address, which I think can be resolved by integrating a necessary role for the emotions into our ethical outlook. Second, I discuss a type of ethical theory that can integrate the emotions in such a way that they are able to play the role I claim for them; these theories may be described as "neosentimentalist". Third, I examine, and ultimately reject, an argument from Shaun Nichols, which claims that all neosentimentalist theories are fundamentally flawed. Finally, I pose a question that all neosentimentalist theories must answer: what are the characteristics of emotions that allow us to distinguish appropriate emotional responses from inappropriate ones. This sets the scene for my second chapter, where I will attempt to give an account of the nature of the emotions that will provide us with resources to answer that very question.

A Role for the Emotions in Moral Judgement

The conceptual space for my central argument arises from the proposition that there is a significant and necessary role for the emotions in moral life, in explaining how we are (sometimes) able to successfully apply our moral beliefs, commitments and principles to

situations that the world presents to us. The reason why there is this space is that simply having moral commitments and the will to live by them is not enough to ensure morally good action; rather there are important lacunae between holding the relevant moral principle and being moved by some feature of the world to act in accordance with the dictates of that principle. Pointing out these lacunae naturally raises the question of how we are able to bridge them; my contention is that the emotions have an important role to play in our capacity to do so. My explanation of precisely how they do so will need to wait until the next chapter, where I will give my view of how we should characterise the emotions and how this characterisation renders them suited to perform the functions my view requires of them.

By a moral belief, commitment or principle, I am referring to a general commitment that certain types of action or outcome are of positive or negative moral valence. The notion is supposed to be general enough so that virtue theorists, for example, can subscribe to it (a virtue-theoretical principle might be "one should be courageous"). This definition is supposed to be sufficiently broad that the only people who would be unable to accept this notion of moral principles are thoroughgoing moral particularists, who believe that no characteristic of a situation (including thick concepts) is morally univalent (i.e. always counts positively or always counts negatively in the assessment of a situation)[1]. While I will not explicitly consider the particularist view, my arguments for the necessary role of the emotions apply, if anything, even more strongly in a principle-free context (but would, of course, have to be made out in a slightly different way).

Before I begin on this argument, however, it is important to be clear about how I think the moral case differs from the application of rules and principles more generally. As Wittgenstein's comments on rule-following indicate, it is a mistake to imagine that the process of subsuming a particular instance under a rule can be carried out 'mechanically' without any need for further judgement. With even the simplest rules (Wittgenstein's famous example is "add two") working out (a) that a particular rule applies in circumstances C, and (b) what would constitute following the rule in circumstances C, requires a certain amount of background knowledge and the ability to bring that to bear on the current circumstances in the appropriate way. In order to successfully argue that there is a *distinctive* role for the emotions in moral judgement, then, it will be necessary to show that moving from moral

---

[1] See, for example, Jonathan Dancy, *Ethics without Principles*, (Oxford: Oxford University Press, 2004) for a leading exposition of such a view.

principles to action requires a deeper explanation than the bog-standard case of applying a rule.

One important way in which the application of moral principles differs from the non-moral case is that one of the things we expect the morally good person to do – in fact, part of what it is to fully adopt a moral principle – is to notice when the circumstances around her are such that a moral principle is in play, and so to apply that principle, whereas this is not the case with non-moral principles. For instance, we might say that we learn to tell the time by adopting the time-telling principle. I have internalised this principle, and so now I can tell the time; the fact that I don't always act on the principle, and so tell the time, whenever there is a clock within view does not in any way count against my ability to tell the time. By contrast, if I claim to have adopted, and so to live by, a principle of helping those in need, a flagrant failure to notice opportunities for helping the needy would be evidence against my claim.

In order to apply a moral principle, then, it seems that we must first accurately recognise a situation's features and recognise that they are of moral salience. If we are unable to do this, then no matter how comprehensive and nuanced the set of moral principles we hold, and no matter how committed we are to applying them, we will still be unable to do so, due to an inability to see when they ought to "bite" on the world.

Lawrence Blum gives the example of two people, John and Joan, sitting on a crowded tube train. A woman, obviously exhausted, gets onto the train with heavy bags, and has to stand for lack of available seats. John is aware that the woman is there, but thinks nothing of it. Joan, on the other hand, perceives that the woman is uncomfortable and needs to sit down.

[Third party material excised][2]

As Blum rightly argues, the difference between John and Joan's reaction to the situation is morally significant, both in terms of the morally relevant action it puts Joan, but not John, in a position to do, but also in itself.

The question of whether or not the morally relevant aspects of a situation comes to our attention – whether we are more like John or Joan - is not, itself, morally neutral. Rather, it is a mark of a morally sensitive person that her attention is drawn to such aspects. This gives rise to a puzzle, however: *how* do morally relevant factors of a situation become salient to our

---

[2] Lawrence Blum, 'Moral Perception and Particularity' in *Moral Perception and Particularity* (Cambridge: Cambridge University Press, 1994), 32.

attention at all, given the huge number of different aspects of a situation to which we could equally be attending? I will call this the "attentive frame problem". Given that the ability to overcome the attentive frame problem is necessary to being a morally sensitive person, the question of how we are able to do so carries a certain urgency.

The attentive frame problem may be seen as a specifically moral version of a more general "frame problem"[3]: how are we to pick out when morally relevant factors are present in a situation, given the huge number of facts about any given situation which may turn out to have morally relevant consequences, due to the web of other facts about that situation. We simply do not have the cognitive resources to examine every facet of every situation for potential moral relevance. As de Sousa puts it:

> [Third party material excised][4]

Clearly, the adoption of firm rules for what we ought to accord salience in our reasoning will not help us here, as to use them we would need to have already noticed something before applying the rule to decide whether we ought to notice it or not. The attentive frame problem will then simply bite again at the next level.

Returning to the tube train case, I have already noted that it is morally significant that Joan notices that the moral salience of the woman's discomfort. Yet it is implausible to suggest that Joan goes around actively noticing things and considering whether they might be morally salient, for two reasons. First, there are simply too many features of the situation which could potentially be morally salient for her to check each one for moral salience. If she were required to do this, the weight of the requirement would be paralysing – she would never be able to reach a conclusion to act, due to the need to consider such a huge number of factors that may have a possible bearing on her decision. Second, even if she were somehow able to do this, it would be such a drain on her finite powers of cognition that she would not be able to concentrate on anything else. We can easily envisage, however, a situation where Joan is thinking about something else entirely, perhaps reading a book, but she notices the woman standing there and immediately comes to see this as morally relevant. This raises the question of how morally salient factors are able to break in and seize our attention.

---

[3] See Daniel C. Dennett, *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge [Mass.]: MIT Press, 1978), 125, for his formulation of the general problem.
[4] Ronald de Sousa, *The Rationality of Emotion* (Cambridge [Mass.]: MIT Press, 1987), 194.

The attentive frame problem, however, is not the only puzzle in this space. Apart from merely attending to morally salient factors, we must also be in a position to interpret these factors in the right way, so as to grasp their particular moral salience. Thus the question becomes not merely one of coming to pay attention to a situation, but also of how we conceive of that situation once we are attending to it. I will call this the "conceptual frame problem".

This problem has been indicated by Iris Murdoch in terms of what she calls "loving attention". She gives the example of a mother, M, who regards her daughter-in-law, D, in a negative light, perhaps believing that her son has married beneath him. In spite of her negative perception, M always behaves impeccably towards D. Over time, however, and through an effort of moral willpower, she attends to different aspects of D's personality and behaviour, so coming to see different things about her as salient, and so regarding her in a new light. [Third party material excised][5] In this situation, Murdoch argues, it is natural and correct to regard M as making moral progress, as in the light of a loving attention she comes to see D in both a more accurate and a more loving way.

The move that M makes from seeing D in one way rather than another is not simply a case of looking more closely, and so noticing new facts. Rather she is taking the same facts and construing them in a different way. It is common to have a kind of confirmation bias about other people; when we do not like someone we take examples of their bad character to be more salient than examples of their goodness, and vice versa for those whom we do like. Through an effort to reassess the salience of individual factors, however, it may be possible to effect a shift in our overall attitude, as M does in her stance towards D.

Of course, this is not just an issue in regard to the way we view other people; it can also come into play in the way we see a certain type of situation. Blum[6] gives the example of Tim, a white man, who sees a taxi drive past a black woman attempting to flag it down, only to pick Tim up instead. Tim could conceptualise this situation as a case where the taxi driver has made an honest mistake; alternatively, he could see it as an example of racial discrimination on the part of the driver (among a large number of other scenarios of greater or lesser likelihood). Obviously, conceptualising the situation one way or the other will bring very different moral considerations into play for Tim. It is important to note that my point here is about how Tim understands the situation in front of him – whether as a case of a mistake or

---

[5] Iris Murdoch, *The Sovereignty of Good*, 2nd edn (London: Routledge, 2001), 17.
[6] Blum, 36-37.

as one of racial discrimination – rather than about how he conceives of the moral disvalue of racial discrimination itself. My comments in this section are very much aimed at this everyday level of moral reasoning, rather than the more high-level questions of what it is that makes certain things morally valued or disvalued

Again, the way we conceive of a situation is not only instrumentally important, in terms of the action it can lead us to take, but is itself of intrinsic moral importance. As Nussbaum puts this point: [Third party material excised][7]

Given the wide array of possible ways in which we could conceptualise a single situation, we need to ask how the morally sensitive person is able to conceptualise situations in such a way as to render apparent their moral import. Again, it seems implausible, both in principle and in terms of accounting for the phenomenology, to suppose that one goes through a full range of possible ways of conceptualising a situation (whatever such a list might look like), before deciding which one is the best fit. Rather, it seems that the situation will simply strike the morally sensitive person in the right way – while we may come to view a situation in a different light as a result of argument from others, in general terms the way in which we conceptualise situations does not seem to be rule- or argument-bound at all, from occasion to occasion.

In order to sustain the claim that this is a distinctive issue in moral reasoning, however, it is necessary to argue that there is something peculiarly difficult about the conceptual frame problem in moral cases. One may argue, for instance, that even in non-moral cases, the question will present itself of how to appropriately conceptualise characteristics of a situation – any definitional rule will require the exercise of judgement to decide whether a given case properly falls under it.

The difference in the moral case is the "open-endedness" of moral concepts. In the case of most non-moral concepts, we would generally expect a high level of convergence between competent speakers of the language on whether the relevant term applies, outside of vague concepts. In moral concepts, however, this is much less clear – reasonable people can disagree over whether thick moral terms like "unjust" or "dishonest" apply in a particular situation in a way that goes beyond a mere acceptance of vagueness. In the case of deciding

---

[7] Martha Nussbaum, 'The Discernment of Perception: An Aristotelian Conception of Private and Public Morality', in *Love's Knowledge: Essays on Philosophy and Literature*, (Oxford: Oxford University Press, 1990), 84 [emphasis added]. The embedded quote is from Henry James.

whether a particular action or attitude is racist, for example, we may invoke a whole range of social, historical and moral factors to support our judgements. While we may wish to impugn the *moral* character of somebody who applies a moral term in a way that is different to us, in many cases it would not make sense to accuse them of a lack of linguistic acuity. The role of judgement in moral cases is not merely in working out how to apply a definitional rule, but in filling out the meaning and application conditions of a rule which inevitably remains quite schematic.

So, we have two versions of the frame problem to overcome in applying moral rules. The attentive frame problem asks how we are able to pick out morally salient factors of situations, from among a welter of other information. The conceptual frame problem asks how we are able to perceive of situations in the right way to reflect their proper moral import. While it has been convenient for me to present these as two distinct problems, in reality they form two aspects of a single, complex problem: if we conceptualise situations as involving a particular moral concept, this will draw our attention to them further, which leads us to further consider their moral import, thus forming a self-reinforcing cycle.

The key question relating to this process is how we solve the two varieties of the frame problem, attentive and conceptual. Given the wide variety of factors that may have a bearing on the questions under consideration in a given situation, how are we able to pick out the factors that are relevant and then how do we accord these factors the appropriate level of salience to ensure that we come to the correct conclusion?

Nussbaum writes that:

> [Third party material excised][8]

I think we need to go further than this, however – relying on an algorithm is not merely insufficient, immature or weak; rather, as the foregoing argument suggests, it is an incoherent fantasy. The sheer complexity of the world means that any recognisable moral deliberation could not possibly be based on an algorithmic method, due to the impact of the attentive and conceptual frame problems.

One possible response to this suggestion may be, however, to argue that the moral world simply is not that complex so, in reality, there are no framing problems. One example of a moral outlook that could encourage this type of approach might be individualistic hedonism.

---

[8] Nussbaum, *Love's Knowledge*, 74.

In such a case everything that is of moral value to an agent (i.e. her own pleasure) would be immediately apparent to her, or at least only subject to normal epistemological difficulties, and so no special framing problems present themselves.

While I would accept that in very extreme cases of simple moral outlooks, this type of approach may be plausible, it does not require the addition of a great deal of complexity to bring framing problems back into play. For instance, if we were to move to the "universal hedonism"of hedonistic act utilitarianism, one would again face the attentive frame problem of how we are able to notice when an action one is considering may impact on the happiness of another person. Similarly, if we moved to saying that certain pleasures, say one's which involved cruelty to others, were of no moral value, the conceptual frame problem would kick in as we asked ourself whether we were faced with a pleasure arising out of, or despite of, the negative effects that bringing it about would have on others.

Thus, while it seems that a very simple moral outlook could avoid the frame problems I have been discussing, once we introduce a modicum of complexity they become ubiquitous. I do not intend to argue here for the unattractiveness of simplistic moral theories such as individualistic hedonism; indeed, any reader who is seriously attracted to such a view is unlikely to find much of value in this work.

Once one has attended to and conceptualised the moral import of a situation, however, the work of moral agency is still not complete; one is still required to move to action. If Tim from the example above were simply to note that the taxi driver had racially discriminated against the woman, but feel no inclination to do anything about it, we would rightly count this against him. It is a commonplace that when we make a moral judgement about something, we will (absent weakness of the will) be motivated to act. If a moral judgement is a type of cognition, it is difficult to see how it can be tied to motivation in this way. If a moral judgement is a type of conation, on the other hand, it then becomes difficult to see how weakness of the will is possible. Adopting one of these two choices seems to make the tie between moral judgement and motivation either too weak or too strong.

Bennett Helm characterises the motivation problem as follows:

> [Third party material excised][9]

---

[9] Bennett W. Helm, *Emotional Reason: Deliberation, Motivation and the Nature of Value* (Cambridge: Cambridge University Press, 2001), 125.

While Helm describes the pull towards asserting the essential connection between desire/motivation and evaluative judgement as being based on the need to assert rational control over our actions, I would locate it more in our notion of what it is to make an evaluative judgement; we find it hard to make sense of someone who claims both to regard something as, say, extremely bad, but to feel no motivation to stop it. It is this difficulty that makes weakness of the will a puzzle in the first place, after all. I am in total agreement with Helm, however, that this is an important question, to the extent that any contender moral theory must give us the resources to deal with it.

Thus we face two huge questions in understanding how we move from holding a set of moral principles to acting in accordance with them. First, how are we able to attend to and conceptualise situations so as to understand their moral import, so we can understand that one of our moral principles is 'in play'? Second, how does this recognition motivate us to act?

In the next chapter I will argue that integrating the emotions into our account of moral judgement can give us resources to answer these questions. In order to provide a background against which this argument can be made out, however, I will first consider a leading contender for explaining how the emotions can be integrated in this way: neosentimentalism. Proponents of neosentimentalist theories propose a close conceptual relationship between emotion and value. While I will not seek to argue for its central contention, I intend to make use of it. If we suppose that neosentimenalism is broadly correct, then we are in a much stronger position to explain how we are able to resolve the problems I have set out.

Neosentimentalist Ethical Theories

In this section I will describe two types of ethical theory that integrate the emotions, both of which can be described as examples of neosentimentalism. This position, as defined by D'Arms and Jacobson, is one in which the instantiation of evaluative concepts is seen as being picked out by the experience of a relevant type of appropriate emotional response. Hence they argue that a neo-sentimentalist outlook is defined by acceptance of the Response-Dependency Thesis (RDT), which they characterise as follows:

[Third party material excised][10]

The two types of theory I will discuss differ significantly, but nonetheless they both accept the RDT.

The first type of theory I will consider is that expounded by John McDowell and David Wiggins, whereby we say that perception of a moral property elicits an emotional response, which both allows us to judge that (or, perhaps, constitutes judging that) the moral property in question applies, and motivates us in the required way in relation to this moral property. Of course, our emotions might misfire, for any number of reasons (depression, distraction, bad upbringing etc.) and so give us both false positives and false negatives of moral property instantiation. For this reason our emotions must be appropriately sensitive to the world, so we have the relevant emotions in response to all and only cases of moral property instantiation.

The classic response to this kind of claim, however, is that offered by J. M. Mackie in his "argument from queerness"[11]. Mackie argued that this kind of view requires that there be a class of moral properties, such that the recognition of these properties necessarily means that the agent should respond to them in a certain way. This, Mackie argues, would make them a strange class of properties, quite unlike any other in the universe.

John McDowell's response to this is that Mackie's criticisms only hit the mark if they are directed against a notion of moral epistemology as perception of primary qualities. First, McDowell concedes to Mackie that [Third party material excised][12]

McDowell then attempts to evade Mackie's argument from queerness by arguing that the picture we should be invoking is that of secondary, rather than primary, qualities. Secondary qualities are those which can only be defined by reference to a capacity the bearer of the property has to elicit a certain effect in human beings (given standard conditions). The classic example is that of colour; something is red if it has the capacity to create the type of experience we know as seeing red in observers, in the appropriate conditions. This means that there is scope for a secondary quality to have a non-standard effect if the conditions are not appropriate; we know from experience that just because something *is* red, this does not guarantee that it will always *appear* red, and vice versa.

---

[10] Justin D'Arms and Daniel Jacobson, 'Sentiment and Value', *Ethics* 110 (2000), 722-48 (729)

[11] John L. Mackie, *Ethics: Inventing Right and Wrong* (London: Penguin, 1977), 38-42.

[12] John McDowell, 'Values and Secondary Qualities' in *Mind, Value & Reality* (Cambridge [Mass.]: Harvard University Press, 2001), 132.

McDowell argues that value epistemology, properly conceived, involves something analogous to the perception not of primary qualities, but of secondary ones. The problem Mackie raises is that of how something that is brutely there in the world, independent of us, can intrinsically elicit an emotional (and hence recognitional and motivational) response from us. McDowell's answer is that, if we conceive of moral values as analogous to secondary qualities, they can only be ascribed by reference to their very ability to elicit these responses from us, in the appropriate circumstances; we do not need to ask how they do so, because this is definitional of them. As he writes:

> [Third party material excised][13]

Wiggins builds on this theme in his essay "A Sensible Subjectivism". He begins by giving an account of how valuational properties get off the ground. First, we decide we want to talk about objects that elicit certain responses from us – things that we find funny, say, or annoying. We group together all the responses that cause us to have a certain kind of reaction, and give them [Third party material excised]. As Wiggins notes, and as numerous unsatisfactory naturalistic analyses of the humorous bear out, [Third party material excised].[14]

So far, however, this seems less like a secondary-quality cognitivist model than an expressivist, non-cognitivist model. What places Wiggins' theory with that of McDowell, rather than with the non-cognitivism of Blackburn and Gibbard (of which more later), is his insistence that [Third party material excised].[15] Thus while it is our reactions that allow us to get a grip on the property, once we have that grip we can reflect on the nature of the property to assess our relevant reactions. Whilst it is highly improbable that we will ever have a reductive definition of such properties in natural terms, a reflection on the kind of things that instantiate this property can clarify and deepen our understanding of it, and inform our notion of the appropriateness of the relevant responses.

We are thus left with properties that can only be defined by their tendency to evoke a particular response from us, but of which the extension is not entirely at the mercy of whatever our responses happen to be. Once the property/response pair is set up, each can inform the other in an ongoing process of clarification. As Wiggins explains:

---

[13] McDowell, 146.
[14] David Wiggins, "A Sensible Subjectivism" in *Needs, Values, Truth* 3rd edn (Oxford: Clarendon Press, 1998), 195.
[15] Wiggins, 195.

[Third party material excised][16]

Applying this type of analysis to moral properties, such as cruelty, we arrive at a cognitivist variety of the neo-sentimentalist outlook. So, for example, on this view there are certain kinds of action/situation/attitude to which we experience a certain attitude of disapprobation, and we refer to these as cruel. On observing a number of cases of cruelty, we start to notice certain "marks" of cruelty such as a deliberate intent to do harm to another person, and so in cases where we feel the relevant sentiment of disapprobation, but the intent to do harm is not present, we might decide that our sentiment is inappropriate, and so this is not truly a case of cruelty. Of course, we might decide to go with the sentiment, and instead question whether intent to harm is a necessary and/or sufficient "mark of cruelty"; this is all part of our ongoing, holistic process of clarifying and refining our moral concepts.

Thus McDowell and Wiggins argue for a position that claims that evaluative properties, including moral ones, can only be picked out by reference to the emotional reactions that they evoke in us. Moral properties are not picked out by the mere tendency to evoke a particular kind of response, but rather by the fact that they *merit* such a response, i.e. such a response to them is *appropriate*. Thus we can 'work from both ends' refining our notion of the nature of the relevant property by reference to the things that appropriately evoke the relevant reaction and refining our notion of when the relevant reaction is appropriate by reference to the nature of the relevant property.

An alternative form of neosentimentalism takes the non-cognitivist route offered by Gibbard and Blackburn.[17] The crux of this view is the claim that while it appears that when we make moral judgements we are making claims about objectively existing moral properties (albeit, if we are convinced by the Mackie/Wiggins arguments, ones that can only be picked out with essential reference to our reactions), in reality this is an illusion. In fact there are no such objective moral properties, and when we make moral judgements we are actually *expressing* our attitude towards the object in question, rather than *describing* that object.

The question immediately arises, however, of what kind of attitude it is that we express when we make a moral judgement. The simplest model would be to take the subjectivist route and say that we simply express our approval or disapproval of the object in question, so when

---

[16] Wiggins, 198.

[17] See Allan Gibbard, *Wise Choices, Apt Feelings* (Oxford: Clarendon Press, 1990) and Simon Blackburn, *Ruling Passions* (Oxford: Clarendon Press, 1998).

someone says that X is wrong (or cruel, or unjust), this simply means that they disapprove of it. Of course one problem with this type of view is that it fails to do justice to the possibility of moral disagreement; with mere taste there can be no arguing. On this notion, the statement 'stem cell research is wrong/permissible' is to be identified with the assertion 'I like/dislike stem cell research'. Thus when people make conflicting moral judgements about stem cell research, they are no more disagreeing than they would be if one said she liked avocado and the other said he did not. This really does not fit in with our notion of what moral discourse is like; we are inclined to believe that something real is at stake in these disputes, and so are unlikely to be impressed by a theory that makes this impossible by definition.

The next move might be to add a commendatory or persuasion-oriented element to the attitudes. So on this kind of theory when we make moral judgements we are not merely expressing our attitude towards a particular object, but also commending that attitude to others: "I don't like stem cell research, and nor should you!' This is an improvement on the preceding theory in that it allows for the possibility of disagreement. When two people make contradictory moral judgements, on this view, they are recommending that others make moral judgements that are contrary to each other. The improvement is marginal, however, as the type of disagreement it allows for is very crude, merely a clash of recommendations. It seems, contrary to this view, that we can discuss conflicting moral judgements rationally, and make a reasoned choice between them on the basis of this. On this theory, however, this is a mistake; when we think we are talking about the reasons to prefer one moral judgement over another, this is really just hot air. To the extent that we think that we *can* have better and worse reasons for accepting particular moral judgements, then, this theory is unattractive.

Moving beyond these ideas, we arrive at 'quasi-realism', a family of views within which we can place the theories of Gibbard and Blackburn. Quasi-realism remains a variety of non-cognitivism, and so maintains that when we make moral judgements we are expressing an attitude, rather than making a claim about whether or not an objective moral property is instantiated by an object. However, these theories seek to maintain the possibility of a recognisable kind of moral discourse, allowing that there are genuinely good reasons for saying that one (non-cognitive) moral judgement is better than another. Indeed, as Blackburn in particular is keen to argue, we are even able to speak of moral properties and true or false moral judgements, although of course these must be understood in an appropriately non-cognitivist sense.

How is such a view possible? If there are no properties genuinely out there for our reactions to aim at (and, when they are appropriate, latch on to), how can we justify talk of property ascription, truth and the rest? According to Gibbard and Blackburn, there are norms governing the various ways in which we can legitimately criticise particular evaluative attitudes, be they those of others or our own. Obviously there can be no standpoint external to all our own evaluative attitudes from which we can criticise them, but it is still possible to perform a piecemeal process of examining their fitness.

Blackburn writes that:

> [Third party material excised][18]

According to Blackburn, we 'earn' talk of truth when we have an attitude which we cannot improve upon in terms of any of these possible flaws.

Once again, then, we arrive at a neo-sentimentalist position, where our emotional reactions pick out moral value. Unlike the McDowell-Wiggins position, this is not a realist outlook, so there is no sense in which our reactions are latching on to even a response-dependent property that can be said to truly exist. However, as this is a quasi-realist position it is not the case that anything goes, so any emotional reaction can pick out an attendant moral value in the world (e.g. the fact that I feel guilty about eating meat on Fridays does not mean that eating meat on Fridays is morally wrong). Rather we have standards of appropriateness, partly socially-created, that govern our emotional reactions, and it is only when our reactions meet these standards that the moral judgements they constitute can be said to be true.

Thus it seems that we have two types of view with differing ideas about moral ontology, but which converge on a single position about moral epistemology, by adopting the RDT. As D'Arms and Jacobson argue, [Third party material excised].[19] For my purposes, which at this stage extend only as far as arguing that there is a contender moral theory that integrates the emotions in the way that I require, the points of convergence between these two views are of far greater importance than their areas of divergence.

Before moving on, however, I must briefly address an objection from Jesse Prinz, who argues for a relativist form of neosentimentalism, whereby [Third party material excised][20]. As Prinz

---

[18] Blackburn, 313.

[19] Justin D'Arms and Daniel Jacobson, "Sensibility Theory and Projectivism", in *The Oxford Handbook of Ethical Theory*, ed. by David Copp (Oxford: Oxford University Press, 2005), 212.

[20] Jesse Prinz, *The Emotional Construction of Morals* (Oxford: Oxford University Press, 2007), 112.

argues, this is an advance on a very simple causal model, as it says that something can be wrong even in the absence of an occurrent emotion, so long as we have a sentiment (i.e. a long-term disposition to respond under standard circumstances) of the relevant kind toward it.

Prinz claims that his view is an advance on the "metacognitive" model because it avoids problems around the definition of appropriateness. An emotion may be appropriate in a number of different ways, and this allegedly causes problems for the view I am proposing.

[Third party material excised][21]

According to Prinz, then , any form of neosentimentalism that depends on a notion of appropriateness is at risk of introducing a vicious circularity.

I am not persuaded by Prinz's objection. First, it does not seem to give us the resources to deal with recalcitrant emotions. I will discuss these in more detail in the next chapter, but in short they are cases where we feel an emotion but do not endorse it; for instance, I might reliably be afraid of flying, but not believe that flying in modern passenger jets is dangerous. Here we would seem to have a sentiment in Prinz's sense, but we would not want to say that the property in question applies. Prinz's theory seems unable to deal with this, but the form of neosentimentalism I favour can do so fairly easily.

Prinz's objection to the standard neosentimentalist position, however, deserves to be taken seriously. While I am confident that we can come up with a non-question-begging notion of emotional appropriateness, it is useful to be reminded of the necessity and requirements of this task. I will take up this challenge in a general sense in Chapter 2, and then for some specific emotions in Chapters 3 and 4.

Nichols' Sentimental Rules Objection

---

[21] Prinz, 112.

Having set out the neosentimentalist route for creating a moral theory which gives a role for the emotions in moral judgement, I intend to move swiftly to explaining how taking this route can give us resources to answer the questions about moral judgement discussed above. Before I do so, however, it is necessary to provide assurances that neosentimentalism really is a candidate moral theory; it will be no use to argue that the emotions can help us in our moral judgements if I have been unable to provide a plausible way in which the emotions can be integrated into the moral sphere. In this section, I shall therefore consider what I take to be the most thoroughgoing critique of the coherence of a neosentimentalist approach; that offered by Shaun Nichols in his book *Sentimental Rules*.[22]

In the book, Nichols attempts to undermine the case for neosentimentalism, relying heavily on experimental findings relating to the capacity for moral judgement in young children, people with autism, and psychopaths. He argues that while neosentimentalism fares well in meeting three important conditions on an adequate account of moral judgement, it is ultimately unacceptable as it fails to accord with conclusions drawn from experimental data. This paves the way, he argues, for him, to set up an alternative account in its place, which he calls the "sentimental rules" account.

Looking at the criticisms that tend to be made of traditional emotivist theories, Nichols identifies three criteria that an adequate account of moral judgement must meet. First, it must allow that emotion plays a crucial role in moral judgement. Second, it must allow that a person can judge something wrong even if he has lost all feeling about it, for instance because he has become depressed or disillusioned. Third it must allow that reason plays a crucial role in moral judgement, for instance when we argue over moral disagreements, or make inferences from conditional statements about moral matters[23].

Nichols agrees that neosentimentalism fares well in these regards:

> [Third party material excised][24]

Nonetheless Nichols goes on to say that [Third party material excised][25]

---

[22] Shaun Nichols, *Sentimental Rules* (New York: Oxford University Press, 2004).
[23] Nichols, 84-87.
[24] Nichols, 88.
[25] Nichols, 88.

The thrust of his argument against neo-sentimentalist accounts centres on what he calls the "dissociation problem". If moral judgement is to be identified with judgements about the appropriateness of certain emotions, then there should be no circumstances in which the capacity to make these two types of judgement comes apart. According to Nichols, however, experimental evidence suggests that a dissociation of the two types of capacity is possible, most notably in the case of young children. For this reason he argues that we are unable to identify moral judgement with judgements of the appropriateness of emotional reactions.

From the experimental evidence he reproduces, Nichols concludes that children of three and four years old are able to make "core moral judgements". These are judgements of whether a given type of transgression is of a moral, as opposed to merely conventional, rule (for example, pulling another child's hair versus talking during story time, respectively). The findings he cites do seem to indicate that children of this age are capable of making such judgements, at least in simple cases.

Equally plausibly, he goes on to argue that young children are not in a position to understand the complex emotions that are usually invoked by neo-sentimentalists, such as guilt, shame and remorse, much less make judgements of their normative appropriateness. Thus it seems that children of this age can make "core moral judgements" without being capable of making judgements of emotional appropriateness. Nichols produces experimental evidence, again apparently compelling, to suggest that a similar dissociation of capabilities is evident in older children with autism:

> [Third party material excised][26]

Having argued that neosentimentalism does not fit the known facts, Nichols offers his alternative, the Sentimental Rules account. The basic claim behind this view is that [Third party material excised][27] This would explain how the children were able to make core moral judgements despite being unable to make judgements regarding the appropriateness of guilt (or shame, or remorse) in the situation in question. According to this theory, the children have internalised a range of norms, both moral (e.g. against hurting others) and conventional (e.g. about not talking during story time), but only the former are backed by the relevant affective mechanism, and so have the moral status conferred on them.

---

[26] Nichols, 92.
[27] Nichols, 3.

Nichols writes that his theory is open as to *how* this process takes place and, in particular, the form the affective mechanism it invokes will take. At one end of the scale, we have an off-line account, according to which the relevant norms must have the relevant emotional backing for a relatively limited, but crucial period in the child's development. [Third party material excised][28] Alternatively, we could look to an on-line account, according to which the relevant emotional backing needs to be present on an ongoing basis. [Third party material excised][29] As Nichols notes, however, there is no need to take either extreme position; he argues that the truth is perhaps somewhere in the middle, but the existing research does not give us the resources to say so with conviction.

Nichols's commitment to experimentalism, and insistence that any theory we put forward answers to the facts on the ground, is admirable. However, we ought not to accept his conclusion that neosentimentalism is unsustainable in light of these facts. First, one might argue that Nichols's characterisation of his target is faulty. His criticism of neosentimentalism is based on the idea that *making* a moral judgement is said to require an ability to assess the normative appropriateness of an emotional response. In fact, however, according to neosentimentalism it is only *assessing* a moral judgement that requires this. It seems open to the neosentimentalist, therefore, to say that children do make moral judgements, but what they lack is the ability to ascend to the higher level to assess those moral judgements. Indeed, it seems that adults make moral judgements directly all the time, only questioning whether those judgements are appropriate in difficult or novel cases.

Against this, however, Nichols could argue that without at least the capacity for sometimes assessing whether their judgements are appropriate or not, young children cannot truly be said to be making moral judgements at all. The analogy here would be with an archer who cannot see the target – merely by firing off arrows at random, which he is unable to see whether they hit or miss, we could not say that the archer is aiming for the target. It seems reasonable, on this basis, to say that merely experiencing emotions, without any possibility of testing them for appropriateness, is not enough to constitute genuine moral judgement.

There is, however, a more robust critique of Nichols' position, which he anticipates:

> [Third party material excised][30]

---

[28] Nichols, 181.
[29] Nichols, 181.
[30] Nichols, 93.

As will be appreciated from my argument in the early part of this chapter, I see moral judgement as a much more wide-ranging phenomenon than the mere recognition and acceptance of some moral rules or principles. In particular, moral judgement must include a means of resolving the attentive and conceptual frame problems both at the stage of recognising a morally salient situation and deciding how one ought to react to that situation. Nichols' "core moral judgement" represents, at best, an ability to resolve part of the conceptual frame problem at the recognition stage, by deciding whether a particular situation represents an infringement of a moral principle, as opposed to a merely conventional one. It completely fails to address how we are able to appropriately pick out and conceptualise the relevant situation in the first place, or how we are able to pick out and conceptualise the salient factors to guide our response to that situation. Given that it leaves out such important aspects of the process of moral judgement, it is hard to see why Nichols' "core moral judgement" should be regarded as the core of moral judgement, as opposed to one, relatively minor, component of it[31].

The experimental evidence that Nichols cites does not involve observing children in the presence of violations of both moral and conventional rules, and observing how they react; rather, the experimenters explain a situation to the child and question them on it. Thus the question of whether the child is capable of attending to the salient features of a situation and conceptualising those features in the appropriate way does not come up; the relevant aspects of the situation are, in effect, pre-noticed and pre-conceptualised for the children. Similarly, the children do not demonstrate how they would choose to respond to the situation, and so the experiment does nothing to examine whether they are able to appropriately take into account factors of the situation that may have a bearing on the correct response.

At base, I am suggesting that the argument between the neosentimentalists and Nichols is analogous to the following: I (in the role of the neosentimentalist) argue that in order to successfully bake a cake, one must follow the recipe. Nichols argues that research shows that children under four are incapable of following the recipe, but are capable of sifting flour. According to Nichols, sifting flour is absolutely central to baking a cake; in fact, one might call it "core baking". Children can manage core baking, but they cannot follow recipes; therefore following a recipe cannot be essential to baking. Regardless of whether you agree with the contention that recipe-following is essential to successful cake-baking, it seems clear

---

[31] It is true that I have not yet explained how neosentimentalism can help us do better on this explanatory task; I will turn to this question in the next chapter.

that the argument form I attribute to Nichols is invalid. In addition, it seems equally clear that where this argument goes wrong is in picking out flour-sifting as core baking. In Nichols' actual argument, the error is similarly in his conceptualisation of the ability to distinguish between moral and conventional principles as core moral judgement.

Nichols offers several retorts to this anticipated type of counterargument. First, he argues that [Third party material excised][32]. This argument, however, is not persuasive. My contention is not that adult moral judgement is radically *different* from so-called core moral judgement, but rather that it is radically *richer* than that. Far from merely making "an empirical assumption" that adult moral judgement involves more than Nichols' core moral judgement, I have spent the early part of this chapter offering an argument to this effect - an argument that I believe is both persuasive and fairly uncontroversial. I would therefore turn Nichols' charge around, and say that it is his account that fails to explain "everyday normative life".

Nichols' second counterargument comes out of the idea that while there is good evidence that "core moral judgement" can be observed cross-culturally, there is a danger that a more advanced type of moral reasoning may not be, so by focusing on anything more demanding than mere core moral judgement we run the risk of engaging in "a project in ethnometaethics"[33]. According to Nichols:

> [[Third party material excised][34]

In this line of argument, Nichols seems to be summoning up a phantom to give pause to the neosentimentalist: the phantom that if she proceeds with her line of argument, the neosentimentalist will have to swallow the very sobering conclusion that all the adults of some cultures are incapable of fully-fledged moral judgement. This line of argument invites two questions: (1) is it very likely that this conclusion will in fact be compelled by the neosentimentalist line, and (2) even if this conclusion is likely to be compelled, does that fact constitute a good reason to abandon a neosentimentalist line of argument? The answer to both of these questions, I submit, is "no".

First, aside any qualms we may have over the applicability or usefulness of Kohlberg's stage approach, it seems orthogonal to the main point at issue here. The stage approach is about the content of our moral judgements, whereas my concern, and that of Nichols' main argument, is

---

[32] Nichols, 93.
[33] Nichols, 94.
[34] Nichols, 93.

with *how* we are able to make moral judgements. The Kohlbergian stage at which adults in different societies are able to reason says nothing about their capacity to assess the appropriateness of their emotions, or lack thereof; rather the Kohlbergian considerations are liable to bear on the content of the reasons they will give for the appropriateness or inappropriateness of emotions.

Second, even if the neosentimentalist argument were to lead us to the conclusion that some societies are more advanced in the capacity for moral reasoning than others, is this necessarily so terrible? While any such conclusion must naturally be approached with caution (both in its formulation and in any upshot that one derives from it) there seems no reason to rule it out in advance. It is even less plausible to suggest that the mere possibility that a given line of argument may lead one to this type of conclusion is in itself sufficient reason to abandon that line of argument. As is suggested by my response to Nichols' arguments about core moral judgement, it is likely that fully-fledged moral judgement comes in a chronologically piecemeal style on the individual level; why should we not posit such a developmental account on the societal level too, rather than adopting an all-or-nothing approach? Far from being the reductio ad absurdum that Nichols' seems to suggest, then, this counterargument seems to be merely an attempt to ward off development of an alternative view through an unsupported claim that it may lead us into uncomfortable territory; such arguments should be resisted.

Finally, Nichols argues that a neosentimentalist approach cannot explain the moral capacities of young children, and that such an explanation is required.

> [Third party material excised][35]

Again, however, this argument is unpersuasive. I do not doubt that, as Nichols argues, young children are able to master some of the skills required for good moral judgement - notably the ability to distinguish some conventional rules from some moral ones. I am also happy to accept that Nichols' sentimental rules approach is a plausible account of how this process takes place. What I am not prepared to accept, however, is that this process is in any way a good approximation of fully-fledged, adult, moral judgement. The sentimental rules account may suffice to account for the morality of the nursery – a morality of simple rules about not hurting other children and waiting one's turn, played out within a closed and predictable environment. It is, however, insufficient for the full complexities of adult morality, which

---

[35] Nichols, 94.

involves appropriately attending to and conceptualising relevant factors from among a teeming mass of detail in the world. For instance, an adult seeing a child being pushed to stop him from touching a hot iron would be able to see that concerns for the child's safety justified the pushing, whereas we might expect a young child to merely apply the simple rule he has learnt and say that it was wrong to push the child.

Of course, none of this is to say that adults themselves never find themselves in situations where they are thrown back on the child-like way of approaching moral issues; that is, of applying apprently moral rules, without being able to give good reasons for why those rules should be applied. A good example of this is the hypothetical case of brother-sister incest when there is no chance of children being conceived, or of anybody else finding out about the incestuous nature of the relationship. A common response to this type of case, among adults, is to say that there is something morally wrong with such activity, and that it should be stopped, without quite being able to say quite what that wrongness consists in. My point is not that adults are never like this, but rather that they sometimes are not, whereas this is the default position for small children.

Nichols' attempt to undermine neosentimentalism fails because he conceives of the ability to distinguish moral rules from conceptual ones as the core of moral judgement, to which other aspects of moral judgement are mere refinements. As I have argued, however, this is incorrect. Just as central to fully-fledged moral judgement as the ability to pick out moral principles are the capacities to pick out and conceptualise situations so as to see that the principle in question applies, and to pick out relevant factors that bear on how one ought to respond to the application of this principle. Small children cannot do this consistently; as Aristotle pointed out, there are no moral prodigies and experience is required to attain good moral judgement. The reason for this, I contend, is that this experience is required to gain the ability to appreciate when an emotion is appropriate or not, which is the true core of moral judgement.

The Conflation Problem

While neo-sentimentalist theories seem capable of surviving Nichols' objection, they still face a significant challenge in the form of making out what it means for a moral emotion to be appropriate. Clearly, not every case of a person having an emotional reaction of a certain

type will correspond to a genuine instantiation of the associated evaluative property. For example, the fact that somebody genuinely feels offended does not mean that what was said was offensive – the person might just be oversensitive. As D'Arms and Jacobson point out, this is not just about how the man on the Clapham omnibus would react: [Third party material excised][36] For these reasons, the Response Dependency Thesis that is offered as a defining commitment of neosentimentalism does not say that an evaluative property is instantiated when someone, or most people, happen to feel an emotion of the relevant type. Rather, it states that it must be "appropriate" to feel the emotion if that emotion is to justify the ascription of the relevant evaluative property.

The problem for the neo-sentimentalist, then, is to give an account of how the emotion in question can be judged for appropriateness, which is still informative. Blackburn dramatises this as the problem of Scylla and Charybdis.

[Third party material excised][37]

We know that the bare fact that a response is elicited by an object is not sufficient to allow us to ascribe the relevant evaluative property to it, but if we say that the response must be *appropriate*, then we just seem to be adding another unexplained term. The task, of course, is to give an account of ways in which a response may be judged to be appropriate or inappropriate.

This leads us to what D'Arms and Jacobson call "the conflation problem". The problem is that the theories in question need to provide us with resources to differentiate those reasons for having the response that bear on the object's possession of the evaluative property in question, from those that do not. In order to do this, we need a theory of the emotions and how we are to judge their appropriateness. Unless we have good reason to think that such a means of judging the appropriateness – or inappropriateness – of emotions is open to us, then neosentimentalist theories will remain non-starters. A further constraint on an acceptable theory of appropriateness was pointed up by the objection from Prinz I discussed above: it must not invoke a notion of *moral* appropriateness that makes it viciously circular.

In the next chapter, I will set out my account of how the emotions are to be characterised, and how, in general terms, an appropriate emotion is to be distinguished from an inappropriate one. This will form the groundwork that will allow me to achieve two vital goals. First, it will

---

[36] D'Arms and Jacobson, "Sentiment and Value", 727.
[37] Blackburn, 116.

give me the resources to show that there is an answer to the conflation problem available to us, so providing the backing we need to allow neosentimentalist theories to remain viable. Second, by examining the nature of appropriate emotions, I will be able to explore the features of emotional responses that make them particularly well-suited to help us answer the questions posed in the early part of this chapter - namely how we are able to solve the attentive and conceptual frame problems to pick out and conceptualise the salient factors when we make moral judgements, and how such moral judgements are tied to motivation.

CHAPTER TWO: THE STRUCTURE OF EMOTIONS

In the previous chapter, I argued that neo-sentimentalist theories provide us with a means of integrating the emotions into our characterisation of mature moral judgement. In this chapter I must set out a plausible characterisation of the emotions in order to show that there are means to tell a justified or appropriate emotion from an inappropriate or unjustified one, as is required by neosentimentalist theories. Setting out my view of the emotions in this way will also give me the grounding I need to explain how they are able to contribute to our understanding of moral judgement, through addressing the attentive and conceptual frame problems, and explaining the motivating nature of moral judgements.

Obviously, the task of giving a complete account of the nature of emotions, and the conditions under which they are appropriate or inappropriate would be a mammoth one, and this entire work could hardly be sufficient to outline a position on this and defend it from even the most obvious lines of attack; indeed, the wider aim of my project precludes me from devoting more than a chapter to this. As a consequence, I will be unable to offer anything approaching a full defence of my view of the emotions. Instead, drawing on what I take to be the leading contributions to the field, I will outline the view of the emotions that I find most convincing, focusing on the aspects that will be most pertinent for my project.

Whilst I believe that the picture of the emotions I am presenting here is the most plausible one, I will not aim to convince those who would be inclined to disagree; rather I will set up the picture as clearly as possible in order to lay the ground for the more in-depth work I will undertake on specific emotions, most notably shame, in subsequent chapters.

Cognitivist Theories of Emotion

In approaching a theory of the emotions, a key starting-point is the idea that emotions, unlike other types of embodied phenomena, have intentional objects: there is something to which they are directed. If somebody tells us that they are, say, angry, it makes sense for us to ask them what it is that they are angry about. In the case of someone who tells us that they are in love, we would not ask what they were in love about, but we could very well ask with whom they are in love. These examples contrast with cases of physical drives such as hunger, and

raw sensations such as a soreness of the feet; it does not make sense to ask what these are about, they have no intentionality; these are certainly not cases of emotion.

Also contrasted, but in a slightly different way, are moods such as a general mood of melancholy. There are a few different ways in which we could treat such phenomena. On the one hand, we might say that they are simply emotions without intentional objects: emotions about nothing. On this view, while the central case is that emotions have intentional objects, there are non-standard ones that do not, of which moods are cases in point. This would leave us with the task of explaining how, then, we are to draw the distinction between emotions and drives/sensations, but perhaps this is not an insuperable obstacle.

A second alternative would be to say that moods are pseudo-emotions (where the physical sensations normally associated with an emotion are felt, without the corresponding intentionality), coupled with a heightened tendency to form genuine emotions of the relevant kind; as, for example, when one is in an irritable mood and so is more likely to become angry at minor things that would be regarded with equanimity in the absence of the mood.

Finally, one might argue that moods *are* emotions and they *do* have intentional objects, but these intentional objects are extremely general; one might be angry about "everything and nothing", "the world" or "life in general". Of course, one might reply to this by asking what an intentional object even amounts to if we allow it to be as general as this. My own inclination, however, is that this position is, on balance, the most attractive. Not much hangs on this for my discussion, however, so I will not dwell on it.

A second point to note is that not just anything can intelligibly serve as the intentional object of an emotion of a particular type (e.g. a case of fear, or jealousy). This idea may be expressed by saying that each type of emotion has a particular *formal object*. To take a very simple example, the formal object of fear is something like *the dangerous* or *the threatening*. We would have a hard time making sense of somebody who claimed to fear something but claimed not to regard it as dangerous or threatening[38]. While it does not need to be *true* that something is dangerous or threatening for fear about it to be intelligible, it does need to be apparent how somebody could regard that thing as dangerous; we can make sense of somebody's fear of eating cheese, if we also understand that they have a (mistaken) belief

---

[38] That is not to say that such cases are impossible – see my discussion of 'recalcitrant emotions', below.

that cheese is often poisoned. If they proffer no such reasoning, however, we are likely to find their fear of eating cheese unintelligible – a pathological case, or phobia.

Moving to slightly more complex cases, Anthony Kenny gives the example of somebody telling us that he was envious of his own fruit trees (under that description)[39]. We would be unable to make sense of that statement, however, because the formal object of envy (i.e. the enviable) must be something that we do not, at the present time, possess (or, at least, that we *believe* that we do not possess).

In a similar vein, we might think of somebody telling us that she is ashamed of being so generous. On the face of it, we could not make sense of this; while, as we shall see in the next chapter, the nature of what can qualify as the formal object of shame is contentious, if we are ashamed of some quality we attribute to ourselves, it certainly seems that that quality must be seen in a negative light. So to make sense of this statement, we would need to see how the subject is able to see her generosity in a bad light. Perhaps she does not view her 'generosity' as a true virtue, but rather a desperate need to please others (or some particular person), and it is that that she is ashamed of. Absent some story, however, we would likely be at a loss to make sense of her statement.

Considerations about the necessity of an emotion being directed at some object, which must itself fit some formal criteria, have led some philosophers to adopt a cognitivist view of the emotions. On this picture, an emotion is said to consist, either wholly or in part, of a belief in some proposition. For instance, to take a very simple example, a cognitivist might identify feeling fear with holding the belief that something is dangerous, in the sense of posing a credible threat to the survival or well-being of oneself or something/somebody one values. This type of view is taken by Robert Solomon, who regards the emotions as identical with certain types of judgement.[40] Alternatively, a cognitivist might take a more moderate view; instead of regarding the emotion as identical with a judgement or belief, she might see it as consisting of a belief plus something else (motivation of a certain sort, or certain types of physical feeling, for example). On this latter view, the presence of the belief is necessary but not sufficient for the emotion to be present.

---

[39] Anthony Kenny, *Action, Emotion and Will* (London: Routledge, 1963), 193.
[40] Robert Solomon, *The Passions: Emotions and the Meaning of Life* (Indianapolis: Hackett Publishing Company, 1993).

Adopting a moderate cognitivist viewpoint, Gabriele Taylor talks of the identificatory belief involved in an emotion. She writes that:

> [Third party material excised][41]

Taylor goes on to explain that when one feels fear, this is normally because one believes the prevailing situation to be dangerous. This belief is held because of some further belief that there is some particular characteristic of the situation in virtue of which it is potentially harmful. Her example is my fear of a snake, which is felt in virtue of beliefs that its bite is poisonous and poison would harm me.

On this picture, then, the formal object of an emotion picks out what can count as an identificatory belief for a given emotion. In fear, the identificatory belief is that something is dangerous. In anger, the identificatory belief might be that somebody has deliberately treated us in a way in which we ought not to have been treated. The content of these identificatory beliefs is what distinguishes one type of emotion from another. It is clear why this is called a cognitive view of the emotions: a particular emotion type is defined by the nature of the belief one must hold to be experiencing it.

On this view, an emotion will be intelligible to others as a case of a particular emotion-type if the person having the emotion holds both the identificatory belief and some further belief(s) that they take as supporting the identificatory belief. For instance, if somebody said they were afraid of eating cheese, they might say that eating cheese can be dangerous (identificatory belief) because cheese is often poisoned (supporting belief); this would then be a formally intelligible case of fear.

Where appropriateness comes in, over and above formal intelligibility, however, is in the justification of these beliefs. If one is justified in holding both the supporting belief(s) and the identificatory belief (and the truth of the identificatory belief follows from the truth of the justificatory belief(s) in an epistemically respectable way), then the emotion is appropriate. In the example given above, while we can say that the fear of eating cheese is, in this case, intelligible, it is not appropriate, as the subject is not justified in her belief that cheese is often poisoned (unless she knows something that we do not). Of course, there are well-documented puzzles about how we are to justify beliefs, but the cognitivist approach to emotions allows

---

[41] Gabriele Taylor, *Pride, Shame and Guilt: Emotions of Self-Assessment* (Oxford: Oxford University Press, 1985), 2.

us to reduce the question of emotional justification to the more readily understood problem of belief justification.

How are we to account for an emotion that is intelligible (qua emotion), but not justified, on the cognitivist story? If, from the third-person perspective, we do not see the supporting reasons offered by the subject as justifying the relevant identificatory belief, we might look for non-justificatory reasons to explain why the person is having this emotion. For instance, we might point to some emotion or mood that the person is already feeling, as when someone becomes angry at you over a trivial matter and somebody else explains to you that he has been in a foul mood all day. We might point to some other externally caused kind of cognitive derangement such as drunkenness or the side-effects of medication. Alternatively, we might point to some tendency to have certain emotional reactions, along with some explanation for this: "he has always been prone to angry outbursts – his father was a violent drunk". While this type of explanation might explain why the putative justificatory reasons are seen by the subject as justifying the identificatory belief, and hence the emotion, they are not themselves justificatory, nor need they be believed (or be believed to pull any explanatory weight) by the subject of the emotion.

As an account of the emotions, a cognitivist theory of the sort I have sketched would clearly be inadequate on its own, as it fails to say anything about the affective or feeling element of the emotions. As such, it has not yet provided the materials to allow us to distinguish between an affectless judgement and a fully-fledged emotion. There are numerous ways in which one might attempt to tack on or integrate the feeling aspect with the core cognitivist theory; however, I do not propose to pursue any of them now. The reason for this is that there are significant objections to any theory that argues that emotions are constituted, in whole or in part, by beliefs, meaning that any cognitive picture of the emotions is unsustainable. In the next section, I set out what I take to be the main challenges faced by a cognitivist view.

Problems for the Cognitivist View

To recap, the defining claim of a cognitivist view of the emotions is that in order to have some emotion, X, one must make a judgement, or hold a belief, that Y. Some cognitivist theories would maintain that the holding of Y is both necessary and sufficient to have X because the judgement is to be identified with the emotion, while others would argue that Y is necessary, but not sufficient, to produce X, as the emotion is constituted partly by the judgement and partly by something else (e.g. desire, eudaimonistic concern, etc.). In this section, I will first attack the view that emotions are identical with judgements by arguing that there are cases where we can have the relevant judgement without also having the emotion in question. Second, I will challenge the view that judgements are necessary for the experiencing of emotions by arguing that there are cases where we can have the emotion without having the relevant judgement. I will supplement this argument with some reflections on the phenomenology of emotional experience, suggesting that sometimes emotions just do not feel as if they can incorporate judgements in the way that the cognitivist requires.

Turning first to the strong variety of cognitivism that regards the emotions as identical with their central judgements, I would argue that this view cannot be squared with the observable facts; it seems that we can clearly conceive of cases where someone makes the judgement that is supposedly constitutive of a particular emotion, but does not have an episode of the emotion in question. For instance, think of somebody who is wrapped up in a strong emotion already, such as a person grieving for a loved one who has recently died. If this person is flagrantly and unjustifiably mistreated by somebody else, it seems plausible that he may make whatever judgement we take to be central to anger but, because of the grief in which he is wrapped up, does not dwell on the offence, or feel any desire to punish or get revenge on the offender and so on – in short, that he does not feel anger. If this case is coherent, and if it generalises over all emotion-types, then it seems that emotions cannot be identified with any particular judgements, as it is possible to make the judgement without having the emotion.

Clearly, someone who wants to defend the strong cognitivist thesis will need to deny the coherence of this type of example; this could be achieved in one of two ways. The first is to argue that the person in question is in fact feeling anger. The second is to deny that the person in question is really making the judgement that is (supposedly) constitutive of anger. However, neither of these approaches seem particularly promising.

Considering the first option, anger seems to be a phenomenon that is fairly easy to recognise and well understood by adults in our society, by reference to a set of paradigmatic indicators (shaky voice, red face, desire to punish or seek recompense etc.). To argue that, in spite of the absence of all of these indicators, anger is being felt would appear to have no possible basis other than the claim that anger must be being felt because the judgement that is constitutive of it is being made. This would clearly beg the very question we are attempting to answer - whether the judgement really is constitutive of the emotion.

If the second route is taken, again we have a number of signs by which we judge whether somebody really is making a particular judgement – first among which is the tendency to claim they are making that judgement in the absence of any strong reason to disbelieve them. Of course, this is not foolproof, as the myriad forms of other- and self-deception bear out. The onus is very much on the proponent of strong cognitivism, however, to come up with a non-question-begging reason why we should believe that in situations where one is not feeling a particular emotion, one cannot, in spite of appearances, be making the relevant judgement.

Of course, this argument is only relevant to the strong form of cognitivism, which claims that emotions are identical to judgements. It would be possible to accept the coherence of the scenario I have set out while maintaining that a judgement is necessary for emotions by claiming that it is not sufficient and that some additional element is required. In order to argue against this more moderate cognitivist position, it is necessary to present a case where the emotion is present but not the judgement.

An example of this type of case is provided by so-called "recalcitrant emotions". To illustrate this phenomenon, Michael Stocker suggests the case of feeling afraid of flying, despite knowing that flying in large jets is not dangerous, and in fact carries a much lower risk than other activities the subject would undertake without feeling fear at all.[42]

It seems possible that the recalcitrance can make an appearance at either of two levels. First, one might be afraid of flying, but deny that flying is dangerous; picture the person whose knuckles are whitened by their tight grip on the armrest, who is breathing fast and grinding their teeth, but says "I have seen the statistics, and I know that flying in a commercial airliner is not dangerous, or at least no more dangerous than driving a private car, which I regularly

---

[42] Michael Stocker with Elizabeth Hegeman, *Valuing Emotions* (Cambridge [UK]: Cambridge University Press, 1996).

do without feeling any fear or anxiety". Second, one might be afraid of flying and maintain that flying is dangerous, but be unable to offer any cogent reason for this belief. Such a person, when pushed, might say that they "just think it isn't safe", or that "it isn't natural for people to fly", or offer some similar non-reason. In such cases, people often accept that their reasoning, such as it is, is weak, but still maintain their claims of dangerousness.

Of course, a defender of the cognitivist view might allege that I have begged the question by presenting these cases as ones where we have a genuine case of the emotion but lack a genuine case of the attendant judgement. They may maintain that, despite what I have claimed, this situation is in fact impossible. In the hypothetical situations I have put forward, their argument would go, the subject must either not be experiencing a genuine emotion, or must actually hold the judgement in question. In the particular case under consideration the defenders of simple cognitivism have two possible responses. First, they could claim that, despite appearances, the subject was not really feeling fear; in fact, they might argue, his situation was instead one of phobia, or some similar pathological state, that does not work in the same way as a normal emotion, and so does not require the relevant judgement to be held. Second, they may suggest that the subject is being deceptive (or self-deceptive) about the judgements he is making; it must be the case that deep down, he really does think that aeroplane flight is dangerous, no matter what he says.

Once again, however, we should not be persuaded by these arguments; both strategies seem to have the tail of the argument wagging the dog. The situations I have set out seem possible, and indeed familiar; it is question-begging to simply maintain that they cannot be what they seem. The onus is on the proponent of cognitivism to argue that such situations as I have described cannot be realised. In the absence of any convincing argument to this effect (and I do not know of one), we should take these cases at face value.

Taken together, then, the cases of unemotional judgements and recalcitrant emotions respectively provide apparent counterexamples to the cognitivist claims that particular judgements are sufficient or necessary for emotions to be experienced. While I do not have space here to consider in detail all possible cognitivist responses to these putative counterexamples, my sense is that the prospects for a successful cognitivist reply are poor.

While these counterexamples, if coherent, are enough to render the cognitivist position unsustainable, they may be supplemented by a further consideration. The central point of this line of thought is that the cognitivist view of the emotions does not seem capable of

accurately reflecting the phenomenology of emotional experience. As Cheshire Calhoun suggests, the cognitivist picture fails to account for [Third party material excised][43] The cognitivist picture suggests that we will observe a situation, make a judgement about it, and thereby experience an emotion (either inevitably or if some other condition is met as well). There are occasions, however, when an emotion just seems to hit us as a situation is presented to us, without the need for any intervening "judgement" step. For instance, stepping out in front of a bus I might suddenly feel fear and leap back onto the pavement without having had time to formulate the belief that the bus posed a danger to me. Similarly, coming out of the library to see a thief trying to steal my bicycle I might feel anger immediately, seemingly without having time to formulate any belief about whether his actions constitute a deliberate attempt to treat me as I should not be treated (or any other belief that a cognitivist may claim to be necessary for anger to be felt). This would again tend to lead us away from the idea that having some identificatory belief, in anything but a heavily reconstructed way, is essential to having the emotion.

Admittedly, we ought to be somewhat wary of arguments from introspective examination of phenomenology, such as this one; there are many cases where the way something seems to us is in fact misleading. Whilst we ought not to rely on this line of thinking by itself, it is however a useful addition to the central argument from counterexample against cognitivism. At the very least it should motivate us to give those more formal arguments a fair hearing.

One way of responding to the arguments I have raised against cognitivism is to argue that the notion of belief with which I have been operating is too tight. For instance, one could argue that while emotions require the presence of a particular belief, such a belief does not need to be fully "formulated"; some mental state that falls short of being fully clear and articulated is sufficient. This would open the way for an explanation of recalcitrant emotions as based on a conflict of more versus less well articulated beliefs, while the bus case can be thought of as one whrere a not fully articulated belief is present and doing the work. The constraints of this thesis will not stretch to considering these arguments in full, but my suspicion is that any conception of "belief" which is loose enough to accommodate the proposed counterexamples to cognitivism would be unlikely to be deserving of the name. My purpose at this point, however, is merely to give an account of why we might be motivated to adopt the view I discuss in the next section, rather than to argue that it is the only acceptable approach.

---

[43] Cheshire Calhourn "Cognitive Emotions?" in *What is an Emotion? Classic and Contemporary Readings* ed. by Robert C. Solomon 2nd edn. (Oxford: Oxford University Press, 2003), 236-47 (247).

If we were to accept that the arguments I have put forward will ultimately render cognitivism unsustainable, we would be left with something of a dilemma. On the one hand, to give us an account of the emotions which is both plausible and which would allow the neosentimentalist view to stand up, we need emotions to be capable of being responsive to features of the world. On the other hand, however, we have seen that the simplest way of doing this – adopting the cognitivist view of saying that each emotion is (at least partly) constituted by a particular type of belief – is beset with difficulties. What we need, therefore, is some way of steering a middle path; the emotions must be capable of tracking reality so that they may be judged for their appropriateness, but in some way that is more flexible and holistic than by necessarily incorporating a single, particular belief.

An Alternative to Emotional Cognitivism

My strategy for achieving this will be to adopt a view of the emotions as not constituted by beliefs, but rather by distinctive ways of seeing their objects, which are suffused with concern. The type of theory I am thinking of has been offered by a number of philosophers in recent years, including Amelie Rorty, Cheshire Calhoun, Ronald de Sousa and Robert C. Roberts.[44] The central contention of these theories is expressed with admirable clarity by Calhoun:

> [Third party material excised][45]

Considering first, then, the version of the theory put forward by Roberts, he argues that emotions are "concern-based construals".[46] On this account, when we experience an emotion we, in an appropriately concerned way, see a situation as a case where the "defining proposition" of an emotion is satisfied. The defining proposition plays a similar role as the identificatory belief plays for the cognitivist; namely it identifies/defines the emotion as an emotion of a particular kind. Roberts' view differs from a cognitivist position in two main ways.

First, the "concern-based" part of Roberts' account means that, for him, emotions are [Third party material excised].[47] He goes on to explain that he does not see the emotions as construals plus concern, but rather the concern is, in part, constitutive of the construal: [Third party material excised][48]As a result, when setting out the "defining proposition" for an emotion he tends to give an appraisal-type element, which is like the identificatory belief a cognitivist may posit, and a concern-type element. For instance, for anger he suggests the following defining proposition: "[Third party material excised][49] Everything up to the last sub-clause could equally well serve as a cognitivist's "identifying belief" for anger. What the final sub-clause adds is a distinctively conative, "concern"-type element.

The second difference from cognitivism is the epistemic relation in which the subject must stand to the defining proposition/identifying belief in order to be having the emotion. Rather

---

[44] Amelie Oksenberg Rorty, "Explaining Emotions" in *Explaining Emotions* ed. by Amelie Oksenberg Rorty (Berkeley: University of California Press, 1980); Calhoun *op. cit.*; de Sousa *op. cit.*; Robert C. Roberts, *Emotions: An Essay in Aid of Moral Psychology* (Cambridge: Cambridge University Press, 2003).
[45] Calhoun, 246.
[46]Roberts, 64.
[47] Roberts, 79
[48] Roberts, 178
[49] Roberts, 204

than believing or judging that the situation is as specified in the identifying belief, as the cognitivist would claim, Roberts suggests that the subject of the emotion only needs to see the situation as a case falling under the defining proposition. While in the standard case the belief/judgement will follow the construal, this need not always be the case.

The notion of 'seeing as' in use here is a figurative one, but it has much in common with a literal seeing of one thing *as* another. For instance, to use a Wittgensteinian example, when we see a picture of a lion as a lion it is not that we believe that the picture *is* a real lion, or necessarily because we believe that there are certain similarities between the picture and a real lion. Rather it is because we see certain analogies between the picture and a lion, which does not seem to require having any beliefs about particular similarities between the two (although it may entail that we must be able to crystalise some such beliefs when challenged to do so, on pain of dropping the belief that the two are analogous).

This point becomes clearer when we consider ambiguous figures, such as the famous duck-rabbit picture. By hypothesis, we are able to see the picture as a duck or as a rabbit, but not as both a duck and a rabbit at once. Furthermore, we can usually shift which animal we see the figure as by an action of will. But clearly this ability does not depend on what beliefs we have about the similarities between the figure and either of the two animals in question, as we are able to move between seeing it as one thing or another by an effort of will, without any change in our relevant beliefs. It may be argued that while this shows that it is not sufficient to see an x as a y that we have a belief that says x is similar to y in z respects, it does not refute the contention that it is necessary that we do so. This is true, and I have no knock-down argument against the necessity thesis, save to say that examination of the phenomenology makes it seem absurd to make this claim, just as it seems absurd to say that I must hold the belief that a particular configuration of browns, greens and yellows is probably best explained by the presence of a tree, whenever I see a tree, qua tree.

Applying this framework to the case of emotions we can hypothesise that, in the same way, we do not need to have any full-blown *belief* that a situation is a case of a defining proposition in order to view the two as analogous, and so to regard the situation as an instance of the defining proposition. Nor need we be aware that we are viewing the situation

in terms of the defining proposition in question. As Roberts argues, in some cases at least, construals [Third party material excised][50]

In both of these respects, Roberts' account is preferable to that of the cognitivists. The requirement that emotions be *concern-based* construals accounts for the possibility of somebody making the relevant construal (and possibly holding the associated belief) without being bothered about it. This allows us to account for the case where somebody apparently holds the identifying belief/judges that the defining proposition is satisfied but does not feel the emotion in question. So in the case of the grieving person who judges he has been culpably wronged but is not angry, we could say that he construes the situation in the right way for anger, but his grief prevents that construal from being a sufficiently concerned one to constitute anger.

I am also attracted to Roberts' proposal of construal/seeing as, rather than belief/judgement as the required epistemic relation for emotion, because this allows us to overcome the problem of recalcitrant emotions. On the construal view, there is no problem with a person believing that x is not dangerous, but still feeling afraid of x, for example, because on the current model no particular beliefs are required to feel an emotion. What is happening here is that how a person sees a situation, and what they believe about it, have come apart. In the usual, cognitively penetrable, case, the beliefs a person has about a situation will determine the way they construe that situation. In other cases, however, this relation breaks down. When we have learned about the relative danger level of flying on a plane, this may lead us to view it as a situation that is perfectly safe and so feel no fear about it. In other cases, however, it is possible that this knowledge is unable to shake our tendency to view it as very dangerous, and so we keep feeling fear. Of course, in reality, one may also vacillate between these two ways of regarding the situation, like the person going back and forth between seeing the two animal aspects in the duck-rabbit picture.

I also made a supplementary argument against cognitivist theories of emotions, based around phenomenological factors. First, I argued that often, when we experience an emotion, it hits us all in a flash. In such cases, the traditional cognitivist story that we are forming a full-blown belief which constitutes/results in us feeling the emotion, simply did not seem to fit. This problem does not have the same bite on the type of theory now under discussion, however. Like seeing the tree outside the window as a tree, rather than as a patch of brown,

---

[50] Roberts, 178

green and yellow, seeing a situation as an instance of a paradigm scenario seems much more likely to be capable of happening instantaneously and passively than does forming a full-blown belief.

It seems, then, that Roberts' theory of the emotions as concern-based construals has significant advantages over the cognitivist point of view. By building in the notion of concern, it means that we can make sense of the proposed counterexample to the cognitivist view whereby someone can make the required judgement but not feel the relevant emotion because she simply is not that concerned by the situation. By referring to construal, rather than judgement, it means that we can both make sense of cases of recalcitrant emotions and better account for the phenomenology of "unmediated" emotional responses.

Where I wish to go further than Roberts, however, is in rejecting the notion that it is in terms of some "defining proposition" that we must see a situation in order to feel a particular emotion in response to it. Instead, I would prefer to adopt Ronald de Sousa's notion of a paradigm scenario as the focus of our emotional experience. According to de Sousa:

> [Third party material excised][51]

De Sousa goes on to claim that [Third party material excised][52].He thus gives us the notion that their different paradigm scenarios *define* or *type* the different kinds of emotion. How is this linked to assessments of the appropriateness of individual emotions? De Sousa goes on to argue that:

> [Third party material excised][53]

Thus, for de Sousa, paradigm scenarios play a dual role, both typing our emotions and providing the standard by which we are to judge their appropriateness.

These paradigm scenarios can be more or less culturally influenced, on both the object and the response side. For instance, on the object side, fear of spiders and snakes appears to require very little, if any, cultural input, apparently being observed in very young infants and cross-culturally, whereas emotions like shame and jealousy would appear much more infused with cultural norms, appearing only in more mature subjects and subject to much more

---

[51] de Sousa, 182.
[52] de Sousa, 184.
[53] De Sousa, 188. The reference to "minimal" appropriateness is meant to account for cases where an appropriate emotion is felt, but the action that results, while consonant with an acceptable paradigm scenario, is inappropriate given wider considerations.

cultural variation in terms of what is regarded as shaming or jealousy-inducing. As de Sousa notes, some of the more biologically-determined paradigm scenarios display a remarkable tenacity and cognitive impenetrability, even in emotionally and intellectually mature subjects. This line of thinking, of course, will provide an opening for explaining recalcitrant emotions, a point to which I will return.

Similarly, on the response side, the appropriate responses to humiliation, say, would seem to be very different depending on the cultural setting. For example, a challenge to a duel is no longer deemed an appropriate response to male-on-male humiliation in mainstream Western culture, although when Aaron Burr challenged Alexander Hamilton it (just about) still was[54]. On a more basic level, Paul Ekman records differences in suppression of facial expressions of emotions such as fear and disgust between American and Japanese groups of subjects, despite strong evidence that the emotions are felt in the same circumstances and with the same intensity.[55]

One might think that an emotion such as grief is fairly insensitive to cultural concerns in terms of response, but the different types of behaviour that are characteristic of mourners at funerals in the Middle East, the UK and Japan would suggest that this is not the case. Also, even within funerals in the UK there are cultural norms regarding the acceptable emotional responses of different categories of mourner; for instance, it would not be appropriate at the funeral of a colleague to appear more upset than the colleague's partner.

A paradigm scenario must fall within some boundaries that have intra-group acceptance in order to count as a paradigm scenario for that particular emotion-type at all – something like Roberts' defining proposition for each emotion will set the boundaries for what is an intelligible paradigm scenario for an emotion. For instance, it would clearly not be intelligible for me to have a paradigm scenario for anger which consists of someone engaging contentedly in a cooperative activity with me in a way that pleases me. To be intelligible as a paradigm scenario for anger, I would venture, the notion of harm (in some suitably wide sense) being unjustly inflicted is required.

To say that paradigm scenarios must be anchored by some definitional constraints, however, is not to make the excessively conservative claim that these constraints must be eternal and

---

[54] For an account of this incident, and its wider context, see Joseph J. Ellis, *Founding Brothers: The Revolutionary Generation* (London: Faber and Faber, 2002), 20-47.

[55] David Matsumoto and Paul Ekman, "American Japanese Cultural Differences in Intensity Ratings of Facial Expressions of Emotion", *Motivation and Emotion* 13 (1989), 143-157.

unchanging. Writing in the different, but related, context of paradigms for the interpretation of social customs and institutions (like courtesy, or law), Ronald Dworkin notes that cases in which we say something conforms to a certain paradigm are very different from those where we say that something is true 'by definition' (e.g. that a bachelor is an unmarried man). [Third party material excised][56] The example Dworkin uses is that men standing when a woman entered the room may once have been thought the height of good manners, but may now be regarded, by some, as unacceptable chauvinism. In the sphere of morality and the emotions, we might point to appropriate emotions towards enslaved peoples as a case of a paradigm that has radically shifted over time.

How does de Sousa's view differ from Roberts'; that is, how does seeing a situation in terms of a paradigm scenario differ from seeing it in terms of a definitional proposition? The difference is that a paradigm scenario will be much richer, with more extraneous detail, than a defining proposition. While a paradigm scenario will need to conform with something like a defining proposition in order to be an acceptable paradigm scenario for the emotion, it will go beyond that, constituting a much more fleshed-out situation. Given this difference, we have two good reasons to prefer the paradigm scenario account to the definitional proposition one, drawn from considerations of emotional experience and interpretative charity, respectively.

On the argument from emotional experience, while Roberts' examples of defining propositions seem to provide a good attempt at providing a definition of particular emotions, they do not seem to be the type of thing we refer to when we seek to justify our emotions. Rather, I contend, we would tend to turn to cases where we agree that the emotion in question is justified and show how the current situation is similar. Our experience of emotional justification seems to be built up from a core of cases of clear application, which we then apply by analogy to cases which are less clear. The capacity of literature to shape our moral repertoires, from the Brothers Grimm to Tolstoy, seems to consist in just this positing of alternative paradigms, not in telling us anything we did not know about the definition of emotions.

In terms of interpretative charity, the paradigm scenarios view is also to be preferred as it allows us to make sense of certain idiosyncratic emotional responses which we would otherwise have to write off as unintelligible. Under Roberts' view, we could only understand someone's emotion as anger if we were able to understand how they could be seeing it as

---

[56] Ronald Dworkin, *Law's Empire* (Cambridge [Mass.]: Harvard University Press, 1986), 72.

falling under the defining proposition of anger. This is an advance on the cognitivist view, under which we would need to see how someone could hold the identifying belief before we could understand them as experiencing the relevant emotion. On the paradigm scenario view, however, we would be able to make sense of people feeling emotions in relation to situations in virtue of seeing them in terms of a paradigm scenario, even if they are not seeing them in terms of the defining proposition that makes that paradigm scenario. This will enable us to class such emotions as inappropriate (as they are not appropriately analogous to an appropriately-rooted paradigm scenario) but still intelligible. Indeed, this seems to be a common tactic in psychoanalytic practice, whereby an emotion is addressed not in terms of merely how we are seeing the current situation, but how it relates to previous situations in a person's emotional biography.

I therefore propose to conceptualise emotions as follows: we experience an emotion when we see a situation in terms of a relevant paradigm scenario. While this accounts for the appraisal side of the emotion, however, any acceptable account of the emotions must also give some account of their feeling aspect. As Stocker argues, [Third party material excised][57] Not only would an account of the emotions that leaves out their feeling aspect be incomplete as a description of the phenomena in question; as we shall see, it would deny my theory the necessary resources to explain how the emotions may be supposed to play the role in moral psychology that I wish to stake out for them.

The feeling aspect of the emotions may be divided into two categories: physical feelings and psychological feelings. To take first physical feelings; these refer to the subject's awareness of bodily changes that arise as a result of an emotion-inducing event or situation. This would include all the familiar visceral sensations that typically go along with emotions: include the visceral aspects of emotion such as quickened pulse and breathing, sweating and the feeling of light-headedness. While William James famously argued that an emotion was simply to be identified with the awareness of these bodily changes, we have no reason to accept this – bodily changes do not seem to be necessary or sufficient for feeling an emotion. We can understand someone feeling a cold, obsessive anger that focuses on seeking revenge, without any awareness of bodily feelings at all, which suggests a case where this is not necessary for an emotion. Similarly, we can imagine the type of physical feelings we would usually expect in a case of extreme fear (sweats, nausea, trembling) being occasioned by an illness –

---

[57] Stocker, 26.

apparently a case where the feelings are not sufficient for an emotion (because, of course, they lack a formal object).

While it may be true that one can have an emotion without having any bodily feelings at all (in the case of some types of emotion, at least), I suggested above that it is impossible to have an emotion without having *any* associated feelings; the missing element is psychological feeling, which I take to be present in all cases of emotion. Stocker gives the example of moving from seeing the dangers of a patch of ice to, having fallen on it, feeling these dangers which, on Goldie's gloss, becomes moving from seeing the ice as dangerous to feeling fear towards the ice.[58]

The most important aspects of psychological feelings are attention effects and desires. As indicated by my description of the person who feels a cold anger without any bodily feelings, what makes this plausible as a case of anger, rather than a bloodless judgement that he has been wronged, is the person's obsessive focus on the source of the emotion. When we experience an emotion, our attention is typically held (or dragged back to) both the particular 'situation' (person/action/fact) that is the object of the emotion, and the aspects of that emotion that enable it to be seen as an instance of the relevant paradigm situation. Think of Othello, who, in the grip of jealousy, cannot tear his thoughts way from images of Cassio's intimacy with his wife. These attention effects seem to be necessary aspects of an emotion – if somebody told us that they were afraid of flying, but they never thought about the supposed dangerousness of flights when they were on a plane, we would find this very puzzling.

A similar story holds for desires. In my view, certain types of desire are a vital and defining aspect of individual emotions. Their role, I would argue, is two-fold. First, certain paradigm scenarios will involve the presence of particular types of desires, which will form part of the formal object of the emotion. For instance, unless I desire to possess something (or someone) then I cannot be jealous when I note that somebody else possesses that very thing. Equally, however, desires can be a required "output" of emotions. For instance, if somebody said they were afraid of something, but did not want to prevent what they were afraid of from happening (if this were possible), then this would be difficult to understand. This is not to say that we could not construct such a story – for example, of someone who wants to do something they are afraid of, under that description, to prove to themselves that they can master their fears – but it does require a special explanation of this kind.

---

[58] Peter Goldie, *The Emotions: A Philosophical Explanation*, (Oxford: Oxford University Press, 2000), 59.

Where we ought to be careful, however, is in adopting what Goldie refers to as an 'add-on view' of the emotions. In the context of the paradigm scenario situation, this would involve saying that what really constitutes the emotion is the judgement that a particular situation is analogous to the relevant paradigm scenario, and the feeling aspect is simply caused by this. This approach risks reducing the true core of the emotion to a cold judgement, with the feeling aspect just a contingent effect. As I have argued, however, seeing a situation as analogous to a paradigm situation is not to be identified with coming to hold a belief that the two are analogous. Rather, we should say that experiencing the relevant psychological feelings are an integral part of seeing this analogy; someone who does not have the relevant feelings, cannot truly be said to have seen the analogy in the appropriately concerned way for this to constitute having the emotion.

While, on this interpretation, it is necessary to have some feelings and desires in order to be having an emotion (if not necessarily physical ones) this does not mean that we must "feel" them, in the sense of being consciously aware that we are having them, and that they are part of the relevant emotion. In order to feel our emotions, we must recognise that we are having some feelings and conceptualise them as part of the emotion in question. This opens up space for both unfelt emotions, where we either do not notice the relevant feelings and desires or do not attribute them to any emotion, as well as false emotions, where we attribute the feelings and desires to some emotion or other state that we are not in fact feeling ("I am not angry with you, I am just disappointed"). Of course, there may be good strategic reasons why we might want to deceive ourselves in this way; alternatively, unfelt or false emotions could simply be down to a lack of emotional self-awareness.

Thus, I would argue that feelings, of the psychological variety, are a crucial element of what it is to see a situation in terms of a paradigm scenario; if no psychological feelings result, then I would want to say that the subject is not genuinely seeing the situation in terms of the paradigm scenario, and therefore not genuinely feeling the emotion. This allows us to draw a contrast with a cognitivist version of the paradigm scenario theory, where one could simply have the background knowledge of the paradigm scenario in mind, and then judge that a situation conforms with that scenario, without any attendant feelings. On the view under discussion, feelings operate according to a kind of feedback loop – we begin to see a situation as similar to a paradigm scenario, which generates feelings in us, which in turn lead us to concentrate on the relevant aspects of the scenario in more detail, which leads to stronger feelings. Not only is this role for feelings true to the phenomenology, it will prove to be a

vital element in the account I will go on to give of how the emotions can help us solve the varieties of frame problem I have set out, in a way which would simply not be open to a proponent of a more cognitivist account.

Emotions and Moral Judgement

Having set out my view of the nature of the emotions I can now return to the questions I posed at the beginning of this chapter: (1) is there some means available to us to judge whether an emotion is appropriate or inappropriate and (2) how can the emotions help address the questions about moral judgement (the attentive and conceptual frame problems and the motivational tie) that I expounded in the previous chapter?

Turning first to the question of appropriateness, given the structure that the paradigm scenario model proposes for the emotions, it seems that there are two ways in which an emotion can be fundamentally unjustified. First, the subject of the emotion might be operating with an unacceptable paradigm scenario in mind. Second, the subject may be seeing a false analogy between the situation and the paradigm scenario.

The emotions seem to be quintessential folk psychological concepts. While each individual will have unique paradigm scenarios for their emotions, based on their particular history, character and emotional repertoire, the broad strokes of how these must look, and what situations can rightly be regarded as analogous to them, are set by our shared cultural understanding of the emotion in question. The boundaries of these issues are extremely blurred, and we can expect these scenarios to be in constant flux, both at the individual and the societal level.

Take a person who becomes extremely angry at a minor perceived slight, such as somebody being mistakenly served out of turn at a bar. This is clearly an unjustified emotional response, because the situation cannot reasonably be seen as analogous to an acceptable paradigm scenario for anger. What we cannot tell without further investigation, however, is whether his error is in terms of having a faulty paradigm scenario for anger (whereby anger is considered the proper response to negligible slights) or drawing a faulty analogy to an acceptable paradigm (e.g. by believing that the present situation is just like getting a slap in the face).

It may be objected that the account I have just given tells us very little about how to decide whether an emotion is justified or not; faced with a real life emotional reaction and asked whether or not it is justified, someone armed only with what I have said on the subject may be tempted to throw up their hands in despair. I have some sympathy with this point of view, but I do not take it as a criticism; what I have said so far, inadequate though it is, is just about all I feel able to say at this level of generality. Given the fact that each emotion has its own formal object, in order to set out what constitutes an acceptable paradigm scenario it is necessary to come down from this position of considering the emotions in general to look at how individual emotions work. My aim here is merely to show that there is a plausible framework by which we can decide on the appropriateness or otherwise of emotional responses, as the existence of such a framework is central to the viability of neosentimentalist moral theories. In the next two chapters I will seek a more specific answer to the question of what constitutes appropriateness in the case of shame and some related emotions. We should note, however, that this notion of appropriateness seems capable of avoiding the vicious circularity that Prinz thought inevitable (as I discussed in the preceding chapter).

Now I can move to the question of how the emotions can help us make sense of how we overcome the attentive and conceptual frame problems in making moral judgements, as well as how those moral judgements are tied to motivation.

First, the attentive frame problem: how are we able to pick out and attend to factors of moral salience in a situation from amongst the multitude of other factors to which we could be attending? The model of the emotions I have set out in this chapter provides us with the resources to answer this question. When we see situations as instances of paradigm scenarios, this provides us with a conceptual frame through which to view the situation. The psychological feelings that form part of the emotion focus our attention by picking out for us as particularly salient those aspects of the situation that drive our perception of it in terms of the paradigm scenario in question – that is, those aspects of the current situation that correspond to the paradigm scenario.

As de Sousa notes, [Third party material excised][59] As soon as we cotton on to the idea that an instance of one of these paradigm scenarios is playing out in front of us, our attention is gripped on the details we take to be relevant. Of course, we still need to notice that a particular paradigm scenario may be relevant to something that is going on around us; once

---

[59] de Sousa, 195.

we make this initial recognition, however, the emotional response drives our attention to focus relentlessly on the salient factors, blocking out the 'noise' of extraneous detail. Thus in the case of moral emotions, we have an explanation of how they are able to make instantiations of moral value stand out from the crowd.

The claim that emotions have this role in directing our attention is a natural one, with much introspective support. There is also empirical support for this. For instance, in Stroop tests, where the subject must name the colour of written words, people tend to perform more poorly (i.e. take longer to correctly name the colour) when the words are emotionally affecting, rather than neutral. The leading theory to explain this is that attention is directed towards the emotional relevance of the words, so leaving less attentional resources to focus on the task of naming the colour.[60]

Next we have the conceptual frame problem: once we are attending to the salient factors, how are we able to conceive of them properly, in the way that appropriately represents their moral import? Again, the model of the emotions I am presenting points to a potential answer. The notion of a paradigm scenario provides us with a ready-made framework through which we are able to conceptualise a situation, and the effects on our attention, described above, lead us to focus on those things which may have a bearing on whether that framework is appropriate or not. As the model with which I am operating here is one of *analogy with* a paradigm, rather than *satisfaction of* some fully-worked out definition, this makes it much easier to see how we are able to conceptualise novel situations in terms of our existing knowledge and experience.

Of course, I do not wish to suggest that we will always get this conceptualisation right; as discussed above, one may be drawing analogies with a poorly conceived or inappropriate paradigm scenario, or mistaken analogies between an occurrent situation and an appropriate paradigm. This is as it should be: good moral judgement is achieved over time, as Aristotle's comment about the lack of moral prodigies suggests. Nonetheless, what it does do is to provide an explanation of how this aspect of moral judgement is even possible. As I argued in the previous chapter, the notion that this is achieved by applying some formal definition is implausible. Instead, the notion of seeing analogies with a paradigm is flexible enough to

---

[60] Donald G. MacKay, Meredith Shafto, Jennifer K. Taylor, Diane E. Marian, Lise Abrams and Jennifer R. Dyer, "Relations between Emotion, Memory and Attention*:* Evidence from Taboo Stroop, Lexical Decision, and Immediate Memory Tasks", *Memory and Cognition* 32 (2004), 474-488.

provide a plausible means of allowing us to conceptualise novel cases, while being substantive enough to avoid the anything goes game of "I see it this way, you see it that way, let's agree to disagree".

Now we can move to the question of motivation: what is it about seeing a situation as being of moral import that motivates us to act in response to such an appraisal? As Prinz points out, [Third party material excised].[61] The psychological feelings involved in seeing a situation as analogous to a paradigm scenario will include particular types of desire – in the case of dangerous situations, for example, this will typically be (in broad terms) to flee or to fight. In ordinary cases we do what we take to be the thing to do (or one among the range of reasonable things to do), but there are cases where this breaks down. Indeed, part of what constitutes a paradigm scenario will be that it is the type of situation that merits a response of a certain type. Thus, the paradigm scenario theory is able to explain why, in standard cases, the experience of an emotion (and so, in the case of moral emotions, the recognition of moral significance in a situation) motivates us.

This, of course, is given an extra kick by the physical feelings that go along with an emotion, which can motivate us to get rid of them (if they are negative) or prolong them (if positive). These physical feelings can also be experienced in an anticipatory way, motivating us to act or avoid acting in ways that would make us feel good or bad, respectively. As Patricia Greenspan argues in the case of guilt, [Third party material excised].[62] While I would not wish to limit this phenomenon to guilt, I strongly agree with Greenspan's claim.

There is good empirical evidence for the motivating power of emotion. For instance, Prinz refers to [Third party material excised].[63] One plausible explanation is that emotions are essential in human motivation, so when these patients lose the ability to experience emotions, they are no longer motivated to act.

The remaining issue I must address is how we are able to judge the appropriate way to act in such a situation, given the many potential factors that could bear on whether a particular course of action is appropriate to a particular situation or not. While the paradigm scenario and attendant emotion will suggest some broad course (or courses) of action that are appropriate to that type of situation in general, particular facets of the situation in question

---

[61] Prinz, 17.
[62] P. Greenspan, *Practical Guilt: Moral Dilemmas, Emotions and Social Norms* (Oxford: Oxford University Press, 1995), 67.
[63] Prinz, 17.

may mean that this course of action is *not* appropriate in this case (or, where two possible courses of action are suggested, one is more appropriate than the other).

This, really, is a recurrence of the conceptual frame problem. By focusing our attention on the details that are most pertinent to the paradigm scenario in question, emotions can help us attend to the factors that are relevant to determining how we ought to act, and help us conceptualise these factors so that they impact on our decision of how to act in the appropriate manner. Of course, people will sometimes go wrong in their attempts to do this, but by bringing the emotions into the picture we can get a sense of how they are able to get it right, at least some of the time.

One potential challenge to the account I am offering would be to say that it is unclear how emotions are able to solve the conceptual and attentive frame problems, Could we not simply run a cognitivist account, arguing that the individual notices that an occurrent situation is analogous to a particular paradigm scenario, which then tells us how we should direct our attention and conceive of a situation? This type of account would be cognitivist 'all the way down' and so would remove the need for an emotion-based account of moral epistemology.

My response to this is that such a cognitivist account simply cannot be correct; reason alone cannot do the job of solving the conceptual and attentive frame problems. As de Sousa notes, [Third party material excised][64]. It is true that we may, without emotion playing a role, notice some factors of a situation that are analogous to a given situation. However, one of the myriad other factors about the situation may mean that the situation is *not* appropriately analogous to the paradigm situation, and therefore that the response we should adopt is *not* the one that that paradigm scenario would suggest. Given the sheer range of factors that could play this role, reason alone simply cannot help us focus on the relevant ones.

Where emotion is different is that it is a dynamic phenomenon which focuses our attention in a process of iteration with the paradigm scenario. Again, to quote de Sousa, [Third party material excised][65]. Thus when one notices that a situation is apparently analogous to a paradigm scenario, resulting in an emotional reaction, the psychological feelings evoked lead us to focus on a particular set of potentially relevant features of the situation. This gives us a more refined understanding of the situation at hand, which can then be compared to the paradigm scenario in question. In this way, the paradigm scenario and the attention-focusing

---

[64] De Sousa, 191.
[65] De Sousa, 195

effects of the emotion go hand in hand in helping us gain an appropriate conceptualisation of a situation and its moral import. Just as the desire-infused feelings of an emotion help create in us the motivation to act in accordance with our moral assessment of a situation, so at this earlier stage they generate the desire, and the ability, to properly consider and conceptualise the moral meaning of a situation.

In this chapter I have given an account of the emotions as being based on responses to analogies between occurrent situations and relevant paradigm scenarios. This allowed me to avoid the traditional difficulties faces by a cognitivist account of the emotions, while maintaining a connection between the justification of emotions and facts about the world. With this notion of the emotions in hand, I have been able to both give an indication of how one would go about distinguishing an inappropriate emotion from an appropriate one, and of how the emotions can help us solve the frame problems in moral judgement I set out in the previous chapter. Together, this adds up to a defence of the idea that the emotions can play an important role in moral judgement; my task for the rest of this work will be to argue that shame ought to be among the emotions playing this role in a person of good moral judgement. I begin this task by considering the emotion of shame, and the nature of the paradigm scenario at its centre, in the next chapter.

CHAPTER THREE: SHAME

In the previous chapter I argued for a conception of the emotions as constituted by a subject seeing the situation at the centre of the emotion as analogous to some acceptable paradigm case or cases. This led me to accept a view of the appropriateness of emotions that holds that an emotion is appropriate when the occurrent situation at its centre is genuinely analogous to an acceptable paradigm scenario. Once we conceive of emotions in this way, I argued, we can see how assigning them a key role in our moral theory can help explain how we overcome the challenges for accounts of moral judgement I set out in the first chapter.

In order to advance my argument that shame is one of the emotions that ought to play a key role in moral judgement, I need to get a firmer grip on what appropriate shame looks like. As my account of the emotions in the previous chapter suggests, the way to achieve this will be to examine the limits of what constitutes an acceptable paradigm scenario for shame. This will give us a sense of the range of situations to which shame can be an appropriate response; in the following chapter I can compare this with similar results for other emotions to support my argument that shame is uniquely well-placed to allow us to respond to certain types of important moral considerations.

The importance of setting out the central paradigm scenario for shame is thrown into stark relief when we consider some more-or-less popular conceptions of the emotion which cast it as being very much from the dark side of human emotionality, alongside the likes of envy, schadenfreude, jealousy and rage. Bernard Williams neatly sums up this approach to shame:

> [Third party material excised][66]

It seems clear that it would count against a contender for the role of moral emotion if it were inherently superficial, heteronomous or egoistic; if shame necessarily manifests all three of these characteristics then it cannot be a contender for that role. Williams goes on to argue, however, [Third party material excised][67] It is vital for my defence of shame as a moral emotion that Williams' assertion is true; this chapter is dedicated to showing that, and how, it is.

---

[66] Bernard Williams, *Shame and Necessity* (Berkeley: University of California Press, 1993), 77.
[67] Williams, *Shame and Necessity*, 78.

Shame as an "Emotion of Self-assessment"

At the centre of my picture of the paradigm situation that is central to shame is a negative judgement about the person who is to feel the shame (i.e. the 'subject' of the shame). This is a vital claim, though surely an uncontroversial one. At the outset the claim is as minimal as possible: it simply says that the paradigm situation relevant to shame involves somebody passing a negative judgement on the subject. It does not say anything about the identity of the person passing the judgement, or the content of the judgement beyond the fact that it is a negative one.[68]

One might take issue with my claim that a *negative* judgement of the subject is required. Reflection on the familiar case of feeling ashamed at being seen naked is particularly likely to evoke this kind of question, as it does not seem as if it is necessary to have anyone passing a negative judgement; shame might seem to be quite possible even if onlookers passed no negative judgement. My answer to this will be fleshed out more by what I have to say later in this chapter, but in order to head off the immediate objection it is important to give an indication of the shape my answer will take. I will interpret "exposure" type cases as causing shame, when they do, through the subject's self-directed negative judgement, based on inability to control how he or she is presented to others. This formulation allows us to make sense of such cases, while accommodating them within the broader model.

In saying that shame involves the passing of judgements of a certain kind about the subject, I am placing shame in a distinct category of emotions, which Gabriele Taylor, in her book of the same title, refers to as "emotions of self-assessment". This places shame alongside emotions such as pride and guilt as emotions that essentially involve a judgement of some kind about the subject. Taylor writes that [Third party material excised][69]

---

[68] It may be argued that in talking about judgement here I am in danger of falling into the same trap as the traditional cognitivists I discussed in the previous chapter, and that I should instead reintroduce my weaker notion of seeing an analogy at the level of the paradigm scenario. I find this suggestion unattractive, however, for two reasons. First, I do not think it would work, as it would be the first step on an infinite regress as we try to answer the ever-repeatable question "analogous to what?" Second, I do not think it necessary; the paradigm scenario represents a clear, central case of the emotion at hand, rather than replicating all the features of every case of the emotion. In effect, the notion of judgement is playing a very different role in the two cases. In the cognitivist theory it is the formation of a belief which constitutes the relevant emotion. In the case of an emotion of self-reflection, it is the judgement the making of which is central to the identity of the emotion itself, which we then must adopt a further attitude towards if we are to experience that emotion. As such, it seems sensible to talk about judgement at the paradigm scenario level, and the weaker notion of seeing an analogy at the level of the relation of the paradigm scenario to the real-world instance.

[69] Taylor, 1.

Emotions of self-assessment can be contrasted with other emotions, which do not include an element of self-assessment. Take an emotion like anger. We might follow Aristotle, arguing that the situation at the centre of anger is that of being treated unjustly, of being harmed by someone who has no call to inflict such a harm[70]. On my account, experiencing anger would arise when we come to see a situation as relevantly analogous to some specific paradigm scenario in which one is treated unjustly. In this case, then, there is one kind of mental state *necessary* for a case of anger: seeing the situation as relevantly analogous to the paradigm case.

In emotions of self-assessment, by contrast, two types of judgement are involved: the relevant judgement about the subject in the present situation, and the subject's seeing the situation in which this judgement is made as relevantly analogous to the paradigm situation in question. The point is that, for emotions of self-assessment, the paradigm situation itself must involve a judgement about the subject of the emotion.

Shame then, like other emotions of self-assessment, has at its centre a paradigm situation in which a judgement is made about the subject. In order to define the limits of what this paradigm situation can look like, and so the types of situation in response to which shame is appropriately felt, I will need to be able to characterise this judgement.

One important question concerns who must or can be making the judgement in question. This corresponds to the debate in the literature over whether, and in what sense, shame is *autonomous* or *heteronomous*; that is, whether the relevant negative judgement is made by the subject herself, or by other people.

As well as picking out who must make the judgement at the centre of shame's paradigm scenario, I must also determine what the precise object of the judgement must or may be. I have already said that the object of shame is the self, but the precise meaning of this is still to be worked out. My exploration of this area will focus on the question of whether shame is *globalised* or *localised*; i.e. whether shame takes as its object the whole self, or just some localised aspect of it.

Once we have a sense of who is making the judgement about what, the final question to explore will be the content of that judgement; I must therefore examine the limits of the

---

[70] Aristotle, *Rhetoric*, Revised Oxford Translation, vol. 2, trans. by W. Rhys Roberts (Princeton: Princeton University Press, 1984), 1378a30.

actual content of the judgement that lies at the centre of acceptable paradigm scenarios for shame.

While these questions are important from a purely descriptive point of view, defining as they do the types of situation to which shame can be an appropriate response, they are not value-free. As we saw in the quote from Williams, critics of shame as a moral emotion often argue that it is necessarily heteronomous (and so reflects only "what the neighbours think"), globalised (and so deeply damaging to the psyche of its subject), and egotistical (so focusing only on what a situation means for the agent, rather than for others). For these reasons, they argue, shame is unfit to serve as a moral emotion. In this chapter I will argue that another type of shame is possible – shame can be autonomous, localised and non-egotistical, and so can be suited to serve as a moral emotion.

Heteronomy versus Autonomy

In the paradigm scenario that is at the centre of shame, whose judgement is it that is required? Is it the judgements of others - our family, friends, community or some other group - in which case we might say, following Wollheim[71], that shame is *heteronomous*? Or is it only our own judgements (which may, of course, be influenced by the judgements of others) that make the difference - a situation we could contrastingly describe by saying that shame is *autonomous*? While those who denigrate shame as a moral emotion sometimes suggest that it is necessarily heteronomous, I contend that an autonomous version of shame, with the mature agent passing judgement on herself, in accordance with her own standards, is also a possibility.

One might argue, however, that the very distinction between autonomous and heteronomous forms of shame is a false one. All our moral thinking is inevitably strongly influenced by those around us, so what is the real difference between autonomous and heteronomous ways of thinking in this area? Wollheim argues that we can distinguish between heteronomy and "radical heteronomy" when thinking about shame: [Third party material excised][72] It is Wollheim's "radical heteronomy" that we are interested in here – although I will continue to refer to it as simply heteronomy. In heteronomous shame not only must the relevant standards

---

[71] Richard Wollheim, *On the Emotions* (New Haven: Yale University Press, 1999), 152.
[72] Wollheim, 153.

come from outside, so must the judgement about how the subject stands in relation to them, and the subject's own attitude towards this judgement, even if she strongly disagrees with it, is irrelevant to whether shame is felt.

The idea that shame is an essentially heteronomous emotion may be thought to gain some support from the idea that some of the most central cases of shame seem to involve the necessary presence of an observer. One key example is being seen naked; of course, I am not usually ashamed to be naked alone in my own flat, but I may well be ashamed if I somehow found myself naked in a busy street. Alternatively, we might consider the famous example given by Sartre of the man who is looking through a keyhole, without feeling shame, but is discovered by another person and so comes to feel shame.[73]

In both these cases, we can unproblematically understand the description of the subject as feeling shame; they seem like fairly central examples of the emotion. The fact that both these examples involve the presence of an observer may therefore suggest a heteronomous account of shame. On this view, we become aware that somebody is observing us, and we gather that they are passing a negative judgement on us. It is this awareness (or it could be a mistaken belief) that some observer is passing a negative judgement on us that is central to the emotion of shame, through being a necessary aspect of an acceptable paradigm scenario for shame.

On closer examination, however, we may begin to doubt whether the presence of an observer is strictly necessary. First, it seems clear that there does not need to be a physical observer who passes a negative view of us, actually present at the time, in order for shame to be induced. The idea that there *does* have to be such an observer is what Bernard Williams calls "the silly mistake"[74]. Taking the example of Achilles' refusal to accept the treasure offered to him by Agamemnon's embassy, he writes:

> [Third party material excised][75]

Williams is surely correct to dismiss the idea that shame depends solely on being seen by a disapproving audience, or the fear of this happening. Suppose that Sartre's keyhole-peeper, knowing that there is no one else in the house except for himself and the person he is observing, nevertheless *imagines* somebody coming around the corner to find him looking through the keyhole. It seems plausible to posit that in this situation the man may be jolted

---

[73] Jean-Paul Sartre, *Being and Nothingness* trans. by Hazel E. Barnes (Abingdon: Routledge, 2005), 300.
[74] Williams, *Shame and* Necessity, 81.
[75] Williams, *Shame and* Necessity, 81.

out of his unselfconscious keyhole-peeping reverie, and come to feel shame about his actions. If this seems like a plausible story, then it suggests that shame does not require a physically-present observer.

Of course, there are some cases where a real observer does need to be physically present, and this is in cases where the fact of being observed transforms one's situation into something to be ashamed of, as in the case of being naked in certain public situations. As I have already indicated, one is unlikely to feel ashamed of being naked in private; it takes the presence of an observer to make this comprehensible as a potential occasion for shame. This contrasts with the keyhole-peeping case, where, as I have said, we can imagine the man coming to believe his action is shameful whether or not he is interrupted by somebody; if he merely imagines somebody catching him in the act, this may be enough to lead him to feel shame.

This dichotomy need not, however, make us reassess the heteronomy/autonomy question as the role of the observer in the two cases is quite different. In the appearing naked case, the presence of the observer is what makes the situation a comprehensible object of shame; what is potentially shaming is being naked *in front of people* (in a certain type of setting etc.). In the keyhole-peeping case, this is not so; we can see peeping through a keyhole as a shameful act, both when this action is observed and when it is unobserved.

This point is emphasised, when we consider the nature of the judgement of the people watching. If it is the negative judgement of an observer that is central to cases of shame, as the heteronomous view of shame would suggest, then *this* type of observer need not be physically present even in the appearing naked type cases. This is because we can imagine a situation where none of the observers have a negative attitude towards the person appearing naked, perhaps feeling admiration, amusement or bemusement, but the case still elicits a sense of shame. *If* some kind of "observer" making a negative judgement is required for shame, then in such a case *this* observer must be an imagined one. So it seems that even in cases like the appearing naked one, there need not be a physical observer making a negative judgement for shame to be an intelligible reaction.

This brings me to my second point, which is that in cases where there is an observer making some sort of judgement, one does not need to agree with that judgement in order to feel shame. This is clear in cases where the judgements of the observers are irrelevant to shame, such as the people observing the public nakedness with admiration or amusement. Quite

clearly one does not need to agree that one's appearing naked is admirable or amusing in order to feel shame; indeed, such an outlook would seem to preclude a shame reaction.

With regard to reactions of others which *are* relevant to shame, this idea of non-agreement cuts both ways. On the one hand, one may respond to the positive reactions of certain types of observer with shame. For example, imagine someone peeping through a keyhole unselfconsciously and without shame, when somebody comes around the corner. Instead of rebuking or being scandalised by the keyhole-peeper, however, the person finding him reacts with an expression of shared pleasure in the illicit action, perhaps giving the peeper a crooked smile and a thumbs-up signal. The keyhole-peeper does not take this in the spirit in which it is intended, however; he has always thought of the man who has discovered him as a seedy old lecher, and he takes no comfort in his approval. Instead, the positive reaction of this person actually causes him to snap out of his unselfconscious state and feel shame at his actions. If this story seems intelligible, then it seems that we have a case where a *positive* reaction from an observer can cause someone to feel shame.

Conversely, we can make sense of cases where a negative reaction from somebody fails to cause us to feel shame. For instance, think of the environmental protester on the airport runway who is told by a police officer that she should be ashamed of herself, ruining decent people's holidays like that. Given the protestor's perception of the relative importance of the issue of aviation's effects on the environment, versus the right to foreign holidays, the police officer's expressed judgement completely fails to make her feel ashamed. She feels that she knows better, and that her actions are 'nothing to be ashamed of'.

This type of situation seems very familiar. A negative reaction to our behaviour need not evoke shame (or even guilt/remorse); resentment, anger, contempt or indifference (to name a few possibilities) seem just as intelligible as possible reactions. Of course, there is a difference between negative reactions from strangers and those from people we love, or whose judgements we treat with particular respect for other reasons; in these cases we are liable to give the benefit of the doubt to judgements that differ from our own, due to the particular status the people in question have in our own lives. However, it is certainly not impossible for us to fail – or actively refuse - to feel shame even when those close to us pass negative judgements on us, especially over matters where we hold firm and considered convictions.

Could it be argued, however, that where shame is elicited despite the absence of any negative judgement from somebody else, or even in the face of a positive judgement from somebody else, this is always due to a negative reaction from some imagined observer? If this were the case, then we could argue that a negative reaction from some "observer" is after all a necessary aspect of an acceptable paradigm scenario for shame, but in the absence of a real-life observer, this reaction can come from an imagined one. In my view, however, this is not a plausible claim, for reasons analogous to those we have been discussing above in the case of physical observers.

First, it seems possible that the case of the lecher having a positive reaction to one looking through a keyhole, and so eliciting shame, could be repeated at the imagined level; simply imagining a positive reaction from the wrong type of person may be enough to generate a shame response. Second, for any figure whose responses we might imagine to different things that we do, we can imagine situations where they would have a negative reaction but no shame would ensue, as we believe differently. Finally, in any situation, we are free to imagine many different observers, having many different types of response to our behaviour/situation, some positive, some negative, some indifferent. Only some of these reactions, according to the view under discussion, can determine our emotional reaction, however, so we must ask why it is a particular imagined response in a particular situation that has this causal efficacy.

We are left, then with two fairly plausible interim conclusions. First, shame does not necessarily require a physically present observer. Second, whether one feels shame or not in a given situation does not seem to be perfectly aligned with the reactions of actual or imagined observers. Taken together, these ideas give us good reason to accept that shame can be autonomous. The defining characteristic of autonomous shame is that the negative judgement that is central to the relevant paradigm scenario is the judgement of the subject on herself. Of course, the reactions of others, actual or imagined, may have a facilitating role to play, in bringing the agent to self-consciously examine her own actions and character, but the judgement that is vital for shame is, in the end, the agent's alone.

Williams acknowledges the attractiveness of this thought, noting that [Third party material excised][76] Rather, he argues, [Third party material excised][77]

---

[76] Williams, *Shame and* Necessity, 83.
[77] Williams, *Shame and* Necessity, 84.

This leads Williams to pose the following question:

> [Third party material excised][78]

In other words, ought we not to conclude that shame is necessarily autonomous?

Adopting an autonomous view of shame would allow us to make sense of some of the examples I have discussed. When those around us, or imagined figures, pass negative judgements on us, but we do not feel shame, this is because we have not been persuaded to agree with those negative judgements. When we have positive judgements passed on us, again either in actuality or in imagination, but this leads us to feel shame rather than, say, pride, this is because the positive judgement coming from this particular quarter makes us reflect on our own conduct and/or character, and it is the resulting judgement of our own which is vital to bringing about the shame response. Purely heteronomous accounts of shame would have difficulty coping with such examples, but when we switch to an autonomous account we can handle them with ease.

We should be wary, however, of thinking that this resolves the question of the autonomy/heteronomy of shame. This point is made by Wollheim, who argues that the view of shame as necessarily autonomous [Third party material excised][79] He goes on to argue that

> [Third party material excised][80]

For Wollheim's argument against a thoroughgoing autonomous view of the emotions to stand up, there must be some cases where an agent feels shame about something, despite failing to make the relevant negative judgement about himself. Again, Wollheim writes:

> [Third party material excised][81]

Equally, Williams seems to be alive to this possibility, writing that:

> [Third party material excised][82]

Do we then need to take the possibility of heteronomous shame seriously after all?

---

[78] Williams, *Shame and* Necessity, 84.
[79] Wollheim, 151.
[80] Wollheim, 152.
[81] Wollheim, 152.
[82] Wiliams, 84.

It certainly seems that we can make sense of such cases. Consider the example of a man who feels shame when failing to use a butter knife to butter his toast because as a child his mother drummed into him the wrongness of such behaviour, despite the fact that as an adult he feels that there is nothing to be ashamed of in doing so. Such examples seem importantly different from cases of recalcitrant emotions discussed in the last chapter – there is no nagging sense that there is something not quite right about using the butter knife, as there is a nagging sense of danger in the person who is aware of all the safety statistics but is still afraid to fly. If we can make sense of such cases as genuine examples of shame - and it certainly seems that we can - we need to take Wollheim's claims seriously.

One possible move would be to deny that such apparent cases of heteronomous shame are really as described. For instance, one could argue that they are not genuine cases of shame, but rather some other, related emotion such as embarrassment, humiliation or fear. Alternatively, one could accept that such cases were examples of shame, but deny that they were genuinely heteronomous. We might argue that while the person claims to find nothing shameful in what they are doing, in reality they must, on some level, believe that this is not the case. Such cases are merely examples of internal cognitive conflict, so despite what the person might claim, we may posit that he actually *does* pass a negative judgement on himself, and so experiences autonomous shame.

Each of these responses seems to sell Wollheim's point short, however. Both responses seem to rely on our making it true by definition that shame is autonomous, responding to putative cases of heteronomous shame simply by claiming either that they are not genuinely heteronomous, or they are not genuinely shame. This risks winning a terminological point only to lose the philosophical battle; we may be left with cases that are phenomenologically identical to other cases of shame, but we refuse to call them shame as they are heteronomous. Instead, it may be better to accept that there can be genuine cases of heteronomous shame and examine the circumstances under which this is possible.

The relative attractiveness of taking this route is underlined if we want to maintain the commonsense idea that young children are capable of experiencing shame. This is certainly an idea that we should try to make room for; as John Deigh notes (albeit in a different context):

[Third party material excised][83]

Wollheim suggests that trying to keep hold of this claim while saying that shame is necessarily autonomous is incoherent. This is because young children do not have the independent sense of their own identity necessary to formulate the negative judgements about themselves that are required for autonomous shame; thus any shame that they do feel must be heteronomous.[84] Once again, we could react to this argument by claiming that while young children may appear to experience shame, they in fact do not, experiencing only related emotions which depend on recognition of the judgements of others. Again, however, such a response looks like it is making distinctions that are contrary to the phenomenology; it would also invite the question of how autonomous shame appears fully formed in maturity without any more basic form of shame in childhood from which it can develop.

The best strategy, then, is to accept a mixed picture, with shame as neither necessarily autonomous or necessarily heteronomous, but as falling into each camp on different occasions. This does not, however, mean that we must agree with Wollheim entirely. Wollheim argues that the heteronomous form of shame, is more foundational than the autonomous, and that the autonomous form comes about only as a modification of the heteronomous.

While taking on board the point that young children are only capable of experiencing the heteronomous form of shame, we may take a different conclusion with regard to mature agents, arguing that, contra Wollheim, autonomous shame is the primary sort. Heteronomous shame remains a possibility in maturity because few of us are absolutely certain in all our moral judgements. Indeed, to be so would be to exhibit worrying psychological tendencies.

Williams notes this point, referring to Glaucon's thought experiment in *The Republic*. We are asked to imagine a just man whom everyone else believes to be unjust, and it is suggested that this man will not be shaken from his course by the views of others, as he knows what is just. If we imagine this situation from the man's point of view, however, this conclusion does not necessarily seem to follow:

[Third party material excised][85]

---

[83] John Deigh, "Shame and Self-Esteem: A Critique", *Ethics* 93 (1983), 225-245 (233).

[84] Wollheim, 152.

[85] Williams, *Shame and Necessity*, 99.

Furthermore, as we become morally mature adults, we become aware of our own capacity for self-deception and self-serving backsliding. Condemnation from others, particularly those we trust, and whose moral outlook we tend to share, may indicate to us that we have engaged in one of these practices, or simply made a mistake (although we cannot, ourselves, see where we have gone wrong). This opens up the possibility of a morally mature, self-possessed type of heteronomous shame, where despite not seeing the justification of the relevant type of negative judgement himself, the agent comes to believe (or strongly suspect) that negative judgements from others are true, and so feels shame.

While Wollheim's arguments about children suggest that heteronomous shame is developmentally prior to the autonomous variety, we aim at instilling children with a sense of independent moral identity as they mature; we aim at moving children away from feeling shame based on the negative judgements of others, real or imagined, to a more autonomous sense of shame. As the child matures, and builds up her own views about what is morally important, she will begin to rebel against some negative judgements which would earlier have elicited heteronomous shame, no longer feeling shame at all, whilst feeling shame autonomously in other situations.

This gives us two ways in which shame may be heteronomous. In the morally immature, heteronomous shame is a necessary starting point for the development of autonomous shame. In the morally mature character, autonomous shame is the central case, with heteronomous shame acting as a prudential response to the possibility of error or self-deception. We should note that the notion of moral maturity is not a strictly chronological one; indeed some adults may never achieve it. Nor is it an all-or-nothing notion; one may achieve moral maturity with regard to shame in some cases, or aspects of one's life, while regressing to immaturity, in others. This links back to the notions of cognitive impenetrability and recalcitrant emotions I discussed in the last chapter; in particular, issues around human sexuality and frailty seem apt to support "immature" heteronomous shame even in adults[86], for instance, because of their upbringing.

Correspondingly, we are left with two possibilities for explaining cases of heteronomous shame, such as the man who remains ashamed of using the wrong type of knife, without having to deny that they are genuine cases of this type. On the one hand, it could be that the

---

[86] c.f. Martha Nussbaum's notion of "primitive shame" in *Hiding from Humanity: Disgust, Shame and the Law* (Princeton, Princeton University Press, 2004), 192.

man, despite genuinely believing that there is nothing wrong with using the "wrong" knife, on some level suspects that he may be wrong in this, and that there may be something in what his mother said (although this is unlikely in this particular case). Much more likely, is that in this area he is reacting in the immature manner, and so feeling heteronomous shame the way a young child does.

Thus my contention that shame can sometimes be autonomous, as well as heteronomous, appears to be vindicated. The primary form of shame is autonomous, but shame begins as heteronomous in young children (sometimes remaining that way, in whole or in part, in older people), and also a mature type of heteronomous shame exists, where we believe the relevant negative judgements of others may indicate a mistake in our own judgement, but our judgement remains the same. Autonomous shame is an achievement, reached only through a process of maturation – if at all. In its mature form, then, far from being merely a reflection of "what the neighbours think", shame represents a situation where an agent passes judgement on himself, according to his own standards. In such cases no observer, be it real, imagined, or even constructed from out of one's own moral values, as Williams suggests, is necessary for shame to be felt.

Localised versus Globalised

A further question about the nature of shame concerns the object of the judgement at its centre. Whilst I have already said that this object must be the self of the subject of the shame, there remains the question of whether the judgement must be passed on the self as a whole, or whether it can be restricted to just one aspect of the self. It appears clear to me that we understand both of these as possible forms of shame, almost to the extent that I am tempted to acknowledge this as a fact and move on. I am given pause from doing this, however, by a disagreement in the philosophical literature on shame, with some significant figures denying that the restricted type of shame is possible. As I noted above, the supposed impossibility of restricting shame to some limited aspect of the self is sometimes used to argue that shame is not fitted to be a moral emotion at all. It is therefore necessary to investigate why this disagreement has arisen.

This issue is raised by Michael Stocker and Elisabeth Hegeman, who consider the question of [Third party material excised][87] While Stocker and Hegeman believe that it is possible for

---

[87] Stocker, 221 (emphasis in the original).

shame to take either of these forms, they maintain that the consensus view on the subject, as expressed by many of the leading writers on shame, denies this possibility. Instead, they claim, these other commentators have held that shame can only be of the globalised variety. Putting this in the context of the paradigm scenario view, the question becomes: must an acceptable paradigm scenario for shame involve the subject making a negative judgement about his whole self, or just some aspect of himself?

Before we assess Stocker and Hegeman's arguments, it is important to get clear about what they take the distinction between localised and globalised shame to hang on. They write that the claim that shame is localised can be taken as:

> [Third party material excised][88]

Localised shame, they argue, is genuinely distinct from the globalised kind.

> [Third party material excised][89]

In addition, Stocker and Hegeman offer a different route to understanding their distinction between globalised and localised varieties of shame. This is to understand it as a reaction to the idea that shame is globalised by understanding the "entire self" that is involved in the claim that shame is about the entire self, as meaning "the self as a whole": [Third party material excised][90]

The thought here is that other emotions can be thought of as taking as their object either a partitionable object or a nonpartitionable one. In the first case, I might be bored with my job, but only be bored with the administrative aspects of it, not the research aspects – the job is a partitionable object of emotion. By contrast, if I am proud of my daughter, I am proud of all of her, the whole person. Stocker and Hegeman go on to argue that at least some emotions seem capable of taking either a partitionable or a nonpartitionable object. For instance:

> [Third party material excised][91]

Applying this to the claim that shame is always globalized, they write that they:

> [Third party material excised][92]

---

[88] Stocker, 222.
[89] Stocker, 222.
[90] Stocker, 224.
[91] Stocker, 223.
[92] Stocker, 223-24.

Stocker and Hegeman argue that there must be more to the point that their opponents are making than the conceptual claim that when we are ashamed of ourselves we are ashamed of something non-partitionable, and so the object of shame in such cases must be the whole self. They go on to write that:

> [Third party material excised][93]

Stocker and Hegeman never give a neat definition of globalised shame, but from the discussion that I have just summarised, it seems that there are two key points that they are attributing to other commentators, and which they want to deny. First is the idea that the *object* of shame is always and must be the whole self – which Stocker and Hegeman take to be in conflict with the claim that shame can be for a particular quality that we possess, or act we have done or failed to do. Second is the claim that this means that the *impact* of shame must be somehow spread throughout the whole self, rather than confined to one aspect of the self. This suggests that, for instance, a person feeling shame will be unhappy with every aspect of herself, rather than just some particular characteristic or failing.

Who are the figures who, according to Stocker and Hegeman, deny the possibility of localised shame? Some of the most prominent philosophers to have worked on shame have seemed to suggest that shame is always globalised. For instance, Martha Nussbaum writes: [Third party material excised][94] Wollheim also seems to make a case for a necessarily globalised view of shame:

> [Third party material excised][95]

Stocker and Hegeman, however, declare that they [Third party material excised][96]

Stocker and Hegeman's case for the existence of localised shame is convincing. I think it fits well with the phenomenology to see shame as, in many cases, focusing on a particular aspect of one's life, and only on one's life as a whole more extreme cases. Furthermore, it seems that we can make sense of cases where shame exists alongside pride, without positing any kind of cognitive failing. For instance, we might think of a woman who is ashamed of her failure to keep in touch with old friends, but proud of her recent performance in her job.

---

[93] Stocker, 224.
[94] Nussbaum, *Hiding*, 184.
[95] Wollheim, 182.
[96] Stocker, 225.

Positing the existence of localised shame makes it easy to explain such cases; this is much harder if we maintain that shame is always globalised in Stocker and Hegeman's sense.

If it is so straightforward to make out a case for the existence of localised shame, however, why do so many writers on the subject deny that this is possible? As we have seen, Stocker and Hegeman say they do not know. We should, however, feel uncomfortable with a view that sees so many other researchers in this field ascribed a view that appears patently absurd, for no apparent reason.

My answer to this point is that I do not believe that these philosophers and psychologists are arguing for the nonexistence of localised shame in the sense meant by Stocker and Hegeman, who have seen a point of disagreement where none really exists. This line of reasoning can be illustrated by looking at the work of Deonna and Terroni, who look carefully at the sense in which shame may be said to necessarily take the whole self as its object.

Deonna and Terroni describe shame as having both a primary and a secondary object. The primary object of shame is always oneself, whole and indivisible: [Third party material excised][97] In addition to this, however, shame has a secondary object, which is the trait or behaviour etc. for which one feels ashamed.

While I find the general direction that Deonna and Terroni suggest to be promising, their proposed discussion of primary and secondary object raises yet more questions. How can it be that an emotion is "about" two different things? In what senses do the primary and secondary objects in an episode of shame differ, qua objects of shame? To avoid some of these issues, I would instead propose that we say that shame always has the same object, the self, but that it can differ in its content. Whereas sometimes the content of the judgement at the centre of shame is very general, and so sees the whole of the object in a negative light, it can also be more specific, seeing only one aspect of the object negatively.

Once we have this structure in place, it is easy to see how Stocker and Hegeman may have ended up speaking at cross purposes to their putative opponents. Nussbaum and Wollheim, in the passages quoted, are making a point about the nature of the object of shame. One is always ashamed of oneself, which is indivisible, rather than somebody else[98] or some state of affairs totally unconnected to oneself. Stocker and Hegeman, by contrast are making a point

---

[97] Julien Deonna and Fabrice Terroni, "Distinguishing Shame from Guilt", *Consciousness and Cognition* 17 (2008), 725-40 (730).
[98] Although see my discussion of vicarious shame later in this chapter.

about the nature of the content of shame. They are arguing that the content can either be extremely general, as in globalised shame, or more specific, as in localised shame. Nothing that Nussbaum or Wollheim says contradicts this claim.

As we saw, the position that Stocker and Hegeman saw themselves as attacking involved two related claims: first, that the object of shame must be the whole self, rather than some property or aspect of the self, and second, that the impact of shame must be spread through the whole self. In the light of the foregoing analysis, however, it is clear that this is a view to which their putative opponents are not committed. On the first claim, they are committed only to saying that the object of shame must be the whole self, so leaving open the possibility that the content may be directed at a property or aspect of the self. On the second claim, they are not committed to saying that the effects of shame need be spread through the whole self; if the content of the judgmement is relatively specific, then the ashamed person need not think of himself as thoroughly bad, nor is he precluded from also being proud of himself, for something else.

Given that those philosophers who Stocker and Hegeman characterise as denying the possibility of localised shame are in fact doing no such thing, I feel confident in returning to my original, common-sense view that shame can come in both localised and globalised forms.

Before moving on from this discussion, it is necessary to make a further distinction: between whether shame is localised or globalised and whether an instance of shame is strong or weak. While there may be a correlation between these two aspects of shame (a possibility I will go on to discuss in Chapter 5), the two are not necessarily conceptually linked. Shame can be spread weakly over one's whole self. Equally, one might feel intense shame about one aspect of one's life or character, without feeling any shame at all about other aspects. The globalised/localised dimension of shame is solely about the generality of the object of the emotion, rather than its intensity. As Manion writes: [Third party material excised][99]

The Content of Shame Judgements

---

[99] Jennifer C. Manion, "The Moral Relevance of Shame" in *American Philosophical Quarterly 39(1)*, 2002, 73-90 (79).

In my discussion of whether shame is paradigmatically localised or globalised, I adopted the strategy of contrasting the object of shame with its content, arguing that while the object of shame is always the (whole) self, the content of shame can differ so that one is ashamed of oneself for more or less specific traits, dispositions, behaviour etc. I have not, however, said anything further about the content of the judgement that must be at the heart of a suitable paradigm scenario for shame.

The question of the judgement's content is not only of interest in setting out what shame is like; it is of central importance in making out shame's suitability to play the role of a moral emotion. Some have argued (or, more often, simply stated) that the judgement at the centre of shame is intrinsically egotistical, being focused entirely on how the subject feels about himself. As such, they sometimes argue, shame is not suited to be a moral emotion, as it takes our attention away from where it should be focused, i.e. on the person who is being harmed.

My view is that shame is not necessarily like this. On the contrary, shame is often deeply rooted in shared values, and we are not free to focus purely on what concerns ourselves. I will explore this view of shame through a consideration of the accounts offered by John Rawls and John Deigh. The former offers a view of shame that certainly is vulnerable to the "egoistic" characterisation. The latter points out some key weaknesses of this account, but in the process risks going too far in the other direction, resulting in a notion of shame that is too social. By forging a middle ground between these two accounts, we can build an account of shame that appropriately recognises both the egoistic and interpersonal aspects of the emotion that make it a unique and valuable moral emotion.

At the centre of Rawls' characterisation of shame is his notion of self-esteem: [Third party material excised][100] For Rawls, regret can focus on the loss of virtually any type of prized good, as for instance when somebody makes an ill-advised comment to his boss, and so scuppers his chances of promotion. Shame, however, is more intimately linked to our sense of self, as it always reflects damage done to one's self-esteem or self-respect. Rawls uses the terms 'self-esteem' and 'self-respect' interchangeably; I will henceforth restrict myself to using the former.

Given that Rawls links shame so closely with self-esteem, the next task is to establish how he develops this concept. He writes that we may

---

[100] John Rawls, *A Theory of Justice* (Cambridge [Mass]: Harvard University Press, 1971), 442.

[Third party material excised][101]

According to Rawls, then, one experiences shame when one comes to doubt, or outright disbelieve, either that one's plan in life is a good one, or that one possesses the necessary attributes to achieve it.

Rawls goes on to note, however, that the plan of life that is important in the experience of shame is the one that the subject herself endorses, and so we will only feel shame relative to our own plan of life, relating to what we do and with relation to the relevant group with whom we compare ourselves. For instance, he writes, [Third party material excised]This means that, [Third party material excised][102]

There are a few further subtleties to Rawls' view, including a division of shame into moral and non-moral varieties. I will address this point in the next chapter, when I come to discuss shame's role as a moral emotion. The central points of Rawls' view, however, are that (a) shame is related to damage to self-esteem and (b) one's self-esteem is damaged by realising either that one's goals are in some way inadequate, or that one lacks the means to achieve them.

One of Rawls' claims immediately leaps out as being contrary to the phenomenology of shame: his contention that we will not strive to achieve something, and feel shame at our failure to do so, if we completely lack the talent that would be required to be successful in the relevant field. While it makes sense that people do tend to avoid organising their life plans to include goals that they patently lack the means to achieve, it seems that there are cases, perhaps involving self-deception shored up by the well-meaning encouragement of others, where people do place a particular goal very centrally in their life plans, despite completely lacking the ability necessary to achieve it. Such unfortunate cases have recently become prime-time television fodder through a spate of televised "talent shows"; insofar as anybody can be said to completely lack musical ability, there are good candidates for this description among the contestants of these shows. This is no reason, however, to suspect that their emotional reactions to discovering that they are ill-equipped for a musical career are anything but real, nor to deny that shame might be among them.

This, however, is a relatively minor criticism of Rawls' view, which could probably be patched up by a slightly more circumspect expression of the relevant point. More seriously,

---

[101] Rawls, 440.
[102] Rawls, 444.

his account would leave some of the most central cases of shame mysterious. For instance, take Scheler's case of the artists' model, who feels shame when she realises that she is being looked upon lustfully by the artist.[103] Whilst it *could* be the case that the realisation that she is being looked upon in this way leads her to realise that her plan of life is wrong or that she lacks the means to fulfil it, it seems that we can make sense of this as a case of shame even when this is not so; indeed, it seems unlikely that this would be the case in such a situation.

This kind of criticism of Rawls' conceptualisation of shame is developed by John Deigh.[104] The essence of his argument is that whilst Rawls is correct to link shame with damage to one's self-esteem, his account of self-esteem, and what constitutes damage to it, is too thin to do the work required of it.

First, Deigh offers three apparent counterexamples, each of which is intended to show a case where shame and damage to self-esteem (as conceptualised by Rawls) come apart. If he is successful in achieving this, we will be forced to conclude either that there is more to shame than damage to one's self-esteem, or that we need a richer conception of self-esteem than Rawls gives us.

The first example is of a young tennis player who is extremely successful in local matches, beating other players with relative ease. The player starts to dream of becoming a professional, and comes to adopt this as one of his goals. His coach, perhaps with his critical faculties somewhat dimmed, perhaps through self-deception, believes that the player is one of the best junior players in the country, and tells him so. The player thus comes to believe that he has a reasonable chance of becoming a professional player, and this increases the seriousness with which he regards playing professional tennis as part of his life-plan. When the player attends his first regional championship, however, he is badly beaten. There is no excuse readily available, and he was not sick or injured or tired; he played close to his best and was simply outclassed by a far better performer. It thus becomes clear that the tennis player is not as good as either he or his coach had thought; certainly, professional tennis now looks like a very unlikely prospect. On Rawls' definition of self-esteem, then, the tennis player has suffered a knock to his self-esteem and so should experience shame.

Deigh suggests, plausibly in my view, that we can make sense of a situation where the tennis player does not feel shame about his lack of ability. Perhaps he feels a little sad at discovering

---

[103] Recounted in Taylor, 61.

[104] Deigh, *op. cit.*

that the professional tennis career he has dreamed of and planned for is unlikely to come to fruition; perhaps he feels angry at his coach for misleading him in this way. Whilst it seems pretty clear that shame would be one intelligible response to this situation, it is much less clear that shame must inevitably be felt. As Deigh remarks, [Third party material excised][105] This example seems to suggest that a blow to one's self-esteem, as defined by Rawls, is not always sufficient to bring about a case of shame.

The second example aims to show that such a blow is not necessary to bring about shame either. The example Deigh chooses is from a novel by Gide, which describes the shame felt by a young girl who, on her first day at school, is teased for her name (she is called Mlle Petérat, which may be translated, Deigh suggests, as Miss Fartwell). The girl had never previously suspected that there was anything funny or ridiculous about her name, and so is shocked by the teasing of the other children. There seems no reason to deny that she could be feeling shame in such a situation.

On Rawls' view, as we have seen, shame is invariably bound up with blows to one's self-esteem, which he tells us means the realisation either that one's life plan is bad, or that one is ill-equipped to achieve the goals that life-plan sets out. In Mlle Péterat's case, however, this does not seem to hold true, for how could it be that the sounds that happen to make up one's last name could have any bearing on the soundness of one's life-plan or capacity to fulfil it? Of course, one could say that Mlle Péterat had some relevant goal as part of her life plan such as avoiding being teased at school, and the discovery that others found her name ridiculous showed that she was not well equipped to achieve this. This seems very implausible, though, and threatens to stretch the concept of a life-plan to breaking point. There does not seem any antecedent reason to think it likely that this goal formed any part of Mlle Péterat's life-plan, or indeed that it had ever crossed her mind. The only reason for thinking that it might have done, is the fact that she felt shame in this instance. This strategy risks taking away the possibility of any independent criterion for some concern forming part of a person's life-plan, leaving us only capable of picking out such cases by reference to whether shame is felt in the relevant situations. The danger here is that the notion of a blow to self-esteem, that Rawls picks out, ceases to be a tool by which we can explain cases of shame, instead becoming a mere redescription of these phenomena.

---

[105] Deigh, 231.

The third case that Deigh describes is that of children. While, as I have argued above, it seems fair to say that young children are capable of feeling shame, and we often behave as though this is the case, it seems much less plausible that they are capable of experiencing shocks to their self-esteem of the kind that Rawls has claimed is necessary for shame. It is unlikely that young children would have a firm life-plan, and the capacity to assess either the worthiness of that life-plan, or one's capacity for fulfilling it, which Rawls claims to be necessary for experiencing a blow to one's self-esteem. Thus it seems that we have a case where shame is possible, but there is no possibility of the subject having experienced a blow to self-esteem, in the terms Rawls sets out.

With these three cases, Deigh provides us with plausible examples of shame and a blow to self-esteem, on Rawls' conception, coming apart, showing us both cases of damage to self-esteem without shame and shame without damage to self-esteem (indeed, shame being experienced even when damage to self-esteem is impossible). These counterexamples suggest that there must be something wrong with Rawls' characterisation of shame, but do little to suggest what that might be. The final type of case that Deigh introduces does, however, provide more of a sense of where the problem with Rawls' account might lie.

The type of case Deigh offers us revolves around a contrast between shame felt in response to an "achievement ethic", as opposed to an "aristocratic ethic". Deigh argues that Rawls' characterisation of shame works well as an account of the former type of shame. An achievement ethic focuses on what our own values are, and how we ourselves act to fulfil them. If this were all that is important in our self-conception, then Rawls' characterisation would work fine.

Where Rawls' conception falls down, according to Deigh, is in failing to account for shame felt in response to an aristocratic ethic. When this type of ethic is in play, [Third party material excised][106] It should be noted that Deigh's choice of the words 'aristocratic' and 'class' are not to be taken to suggest that this type of ethic only exists with reference to socio-economic grouping. Rather it can extend to ethnic, or occupational, or social group; in fact, to any grouping, of which the subject takes membership to be relevant to her self-conception. For instance, one might feel shame over behaving in a way that is unworthy of one's status as

---

[106] Deigh, 235.

a soldier, a caregiver or, as the example I discuss below suggests, a Mashpee Indian. Deigh goes on to argue that [Third party material excised].[107]

Perhaps the most striking example given by Deigh to support his claim that 'aristocratic' shame is a genuine phenomenon, is that recounted by Earl Mills, a Mashpee Indian. Mills recalls how, growing up, he took little interest in his Native American heritage, refusing to take part in tribal ceremonies, and neglecting to learn any of his people's traditional rites. Later, he was drafted into the US Army, and served in a Company which included two other Native Americans, from different tribes. One night, each of these two men performed a traditional tribal dance for the other men in the Company, and when they had finished all eyes turned to Mills, who of course did not know how to perform any of his tribe's traditional dances. Mills reports experiencing deep shame at having to admit that he did not know how to perform the dance in question.

Deigh writes that [Third party material excised][108] Whilst agreeing with Deigh's suggestion that these options seem unattractive, I think he unfairly restricts the options open to the Rawlsians; it would be open to them to argue that Mills had, in fact, on some level, seen the value of participating in Mashpee traditions for some time, but it took this event to make him realise that he had come to genuinely value them, and so feel shame. However, I agree with Deigh's contention that we can explain this case without relying on any such assumptions:

> [Third party material excised][109]

Deigh argues that the fault in Rawls' characterisation of shame lies in the adoption of what he refers to as an "auteur" theory of self-esteem. The central point of this view, he argues,

> [Third party material excised][110]

Deigh goes on to contrast this with the 'aristocratic' type of shame, where attributions of worth flow from one's membership of some grouping, or having some antecedent status. On this view, he writes,

> [Third party material excised][111]

---

[107] Deigh, 235.
[108] Deigh, 236.
[109] Deigh, 236.
[110] Deigh, 241.
[111] Deigh, 241.

When we experience shame of the 'aristocratic' type, Deigh argues, this is not to be understood as a case where we attribute less worth to ourselves because of some way in which we have behaved. Rather, the pertinent factor is the mismatch between appearance and reality. When one is in fact a proud Mashpee Indian, behaving in such a way as to justify people in assuming that one is not, such as by having to admit lacking knowledge of how to perform traditional Mashpee dances, this is an occasion for shame. Deigh claims that this gives us an alternative conception of shame: [Third party material excised][112]

What are we to make of Deigh's critique of Rawls' characterisation of shame? On the positive side, Deigh seems to have identified a genuine gap in Rawls' account. The examples he gives of cases where shame and Rawls' conception of self-esteem comes apart seem plausible, meaning that his claim that the Rawlsian analysis is inadequate is given an initial attractiveness. As regards his description of the aristocratic form of shame, the examples he gives, such as that of Earl Mills, seem once again quite plausible, and difficult to explain from the point of view of the Rawlsian characterisation. Deigh is convincing in his arguments that the notion of self-esteem that Rawls gives us is inadequate to account for all cases of shame.

Nonetheless, I think Deigh goes beyond these insights to make stronger claims than they entitle him to do so. First, consider Deigh's claim that shame is about the relation between appearance and reality. According to Deigh, when we feel shame it is because we have behaved in such a way as to make it appear that we lack some important status that we actually do have. This seems too weak to account for shame; indeed, it seems more attractive as a characterisation of embarrassment.

I will set out my own views on embarrassment in the next chapter, when I come to compare shame to neighbouring emotions, but for the moment it will suffice to say that many central cases of embarrassment seem to involve someone mistakenly viewing the subject as being a certain way, and the subject being unable to correct that impression. For instance, think back to the case of Earl Mills, only this time imagine that he knew how to perform the traditional dance, but was unable to do so as he had hurt his ankle that day. Imagine further that, for some reason, Mills is unable to tell his fellow soldiers about the injury; as a result they assume that Mills does not know how to do the dance, and Mills is unable to convince them otherwise. In this case, it seems to make much more sense to say that Mills would be

---

[112] Deigh, 242.

embarrassed, rather than ashamed of himself. Deigh's view, however, is unable to distinguish the real story of Mills from the amended version I have just told: in both cases, on his account, the emotion Mills feels should be shame.

Contrary to Deigh's claim, then, it seems that shame cannot just be about how we have caused ourselves to appear to others, and its relation to certain identities that we hold. Deigh is right that part of the point of the cases of shame he discusses is that we think that, despite our behaviour, we still retain the identity in question. However, it also seems necessary that we think the reactions of others are based in some justifying factor; that the behaviour is not what is to be expected of a person of the type in question. A central thought is that we have failed to be as a Mashpee Indian, or an aristocrat, or a soldier should. While this is certainly a different thought to the ones that Rawls identifies as central to shame, it is also a more complex one than merely the difference between appearance and reality.

A similar line may be taken against Deigh's view that his arguments show that shame is really a reaction to the threat of demeaning treatment. This would seem to follow if we think that shame is really about appearances, as the only thing to be worried about would be the consequences for us of appearing that way. However, I have argued that there is more to shame than mere appearances, and so I also disagree with this other claim. As I discussed earlier, there must be more to shame than mere fear of bad consequences; Williams is right to claim that a person who only refrained from doing shameful actions because of fear of how others might react would not really understand shame at all.[113] We can certainly imagine cases where a person feels shame of this sort despite no-one being around to witness the shameful action, and there being no real chance of anyone finding out; in such a case, the idea that shame is to do with fear of demeaning treatment is unsustainable. The important idea is that the subject thinks that he has failed to behave as the type of person he takes himself to be should behave.

So, where does this leave us? Taking away some of Deigh's stronger claims, we should still recognise his central insight that there are sources of self-esteem, and thus ways in which self-esteem can be damaged, that Rawls fails to recognise. We can still say that autonomous shame is a reaction to damage to our self-esteem (and, by corollary, heteronomous shame a reaction to damage to the esteem in which others hold us), but we must recognise that we need a wider conception of the form that this can take. While Rawls' auteur conception of

---

[113] See Williams, *Shame and Necessity*.

self-esteem is a valid one, it must be more carefully circumscribed, as shown by the case of the tennis player who apparently suffers a knock to his self-esteem, but feels no shame. We also need to incorporate Deigh's insight that another source of our self-esteem as well as what we achieve, or aim to achieve, is who we are. A failure to behave as the type of person we take ourselves to be ought to behave is a genuine source of shame. Far from the purely egotistical standards of Rawls' auteur conception, this is a non-egotistical basis for shame, rooted in shared values and the way we behave around and towards others.

Finally, then, we should say that the content of the judgement at the centre of shame is as follows: that the subject of the shame is not as he or she should be. While this in itself is fairly uninformative, my discussion of Rawls' and Deigh's notions of shame stands as an indication of how I think this thought ought to be fleshed out.

Shame: Not All Bad, Then

In this chapter, I have staked out some of the ground that will be covered by an acceptable paradigm scenario for shame. It will necessarily involve a negative judgement being made about the subject of the shame, but a number of further features of this judgement remain open to variation. The judgement can be either autonomous or heteronomous, and so made by either the subject herself or others around her. It can also be either globalised, referring to the whole self, or localised, and so focused primarily on some aspect of the self. Finally, while its content will be around the subject failing to be as he or she should be, this can be based on a failure to achieve our desires, but also on our failure to live up to our background or identity, through our behaviour among and towards others.

So how does shame stand up to the critique set out in the Williams quote at the outset of this chapter? Must we accede to a view of shame as necessarily superficial, heteronomous and egoistic, and so accept that shame is inherently unsuited to serve as a moral emotion? If my argument in this chapter has been convincing, it will be clear that the answer is "no".

First, consider the claim that shame is necessarily superficial. We might take this claim in two ways. On the one hand, the claim may be that shame is superficial in that it necessarily focuses on superficial matters. Alternatively, the claim might be that shame is always a light emotion, which is never deeply affecting as it has only a superficial impact. Of course, these two claims could be

combined in a single attack: shame is never deeply affecting for the very reason that it is never about anything deep or important.

I hope that I have done enough to show that both aspects of this claim are highly implausible. While shame can be about superficial matters, such as one's clothes, it can also be about much deeper things – one's failings as a parent, for example. A similar story applies to the claim that shame is never deeply affecting. While it is possible for individual cases of shame to be relatively 'light' and unaffecting, shame can also take deeply affecting forms; at the limit, an enduring experience of shame can have a profound effect on someone's entire life.

Given that these points seem obviously true, why might someone be led to make the types of claim that would require their denial? The most likely diagnosis of this is a failure to properly distinguish between the emotions of shame, embarrassment and guilt. As I will go on to argue in the next chapter, embarrassment *is* necessarily superficial (in one sense) as it relates to how we appear to others, rather than how we really are. Therefore, a failure to properly distinguish between these two emotions, coupled with a tendency to misclassify all 'deep' cases of shame as really cases of guilt, would encourage a conceptualisation of shame as necessarily superficial. As I intend to argue in the next chapter, however, such a picture of shame, embarrassment and guilt is not sustainable. First, we can clearly and consistently distinguish between embarrassment and shame as distinct emotions. Second, we can clearly make sense of cases of emotional experience that do go deep, but which it would be extremely counterintuitive to classify as guilt, rather than shame.

Now for the second charge against shame that Williams mentions: that it is necessarily heteronomous. I have dealt directly with this claim at some length above, with the following conclusion: while shame has its roots in a heteronomous emotional response, and some heteronomous forms of shame can survive into moral maturity, the central, mature form of shame is autonomous. This criticism of shame, then, also fails to stick.

The third claim that Williams mentions, that shame is egoistic, is more plausible than the other two, at least on some possible sharpenings of that claim. I will address one such sharpening in Chapter 5. To rule shame out of contention as a moral emotion in advance, however, the relevant sharpening would have to take the form of saying that one's shame can only focus on the needs, wants, interests, etc. of its subject. Again, from my discussion of the content of the judgement at the centre of shame, it seems clear that we can make sense of cases of shame that focus on our behaviour towards others, and harm that we have caused them ("I am ashamed of treating her so cruelly", for example). It is of course true that shame is always about the self, in that when someone feels

shame, the emotion necessarily takes as its object some aspect of the self, some way that the person is or has become. This is not, however, the type of egoistic self-concern that would rule shame out of contention as a moral emotion in advance.

The thoroughgoingly negative picture of shame, which, if true, would mean that shame was inherently unsuited to serve as a moral emotion, can therefore be exposed as groundless. I am not content, however, to show merely that shame cannot be ruled out as a moral emotion due to its inherent nature; rather, my aim is to show that shame is an essential moral emotion. In order to do this, there are two more key steps in my argument. First, I need to show the ways in which shame is a distinctive emotion by contrasting it with its closest neighbours: embarrassment, humiliation and guilt. This task is the focus of the next chapter. Second, I must argue that these distinctive features of shame allow it to play a distinctive role in moral life, one which cannot be played by any other emotion (and, in particular, not by the leading contender for that role, guilt); I will address this challenge in Chapter 5.

CHAPTER FOUR: EMOTIONS RELATED TO SHAME

Now that I have spent some time examining the nature of shame, it will be useful to place shame among other emotions that occupy the same area, insofar as they are also negative emotions of self-assessment. First, I will consider two emotions that are similar to shame, but may be thought to be of less moral importance: embarrassment and humiliation. Second, I will address the question of how we are to distinguish shame from one of its key competitors as a moral emotion: guilt. Having a sense of how these emotions divide up will be important when I come to consider the role of the moral emotions, and the place of shame among them, in the next chapter.

Embarrassment and Humiliation

Embarrassment and humiliation are related to shame in that they both involve some kind of negative assessment about the self. When we focus purely on the physical reaction connected with these emotions, it may seem that there is no pertinent distinction between them. Taking embarrassment first, Goffman, for instance, notes that embarrassment may be recognised in others by the following reactions: [Third party material excised] He goes on to note that embarrassment may also involve the following physical symptoms which are undetectable 'from the outside': [Third party material excised][114] It seems plausible to say that these physical sensations can also be present in episodes of shame, at least, in particularly acute ones.

This crossover of typical physical reactions leads Goffman to identify embarrassment with shame. He speaks of a sense of shame being involved when we feel embarrassed, and does not mention shame in any separate context. We must, however, retain some distinction between the concepts of embarrassment and shame, not least because the situations in which the two words are used are not interchangeable. Think of somebody who commits a faux pas at a dinner party. It would seem totally socially acceptable for them to laugh it off, saying, "Oh, how embarrassing!" I cannot, however, think of a situation where it would seem appropriate to say "Oh, how shameful!" in such a light-hearted way (unless one were speaking ironically). As Taylor points out, shame strikes us as a weightier, more shattering

---

[114] Erving Goffman, *Interaction Ritual*, 2nd edn (New Brunswick: Transaction Publishers, 2005), 97.

emotion than embarrassment[115]. Perhaps, then, we might take this as indicating that embarrassment is merely a name given to cases of shame that are relatively minor, perhaps relating to occasions when we show ourselves to be deficient in some way that we would rather not be deficient, but this is just a nice-to-have, rather than an essential. On this account, we would say that embarrassment is a subset of shame, felt over shortcomings that are more peripheral to our self-conception.

There are, however, difficulties with this characterisation, which run sufficiently deep to render it unacceptable. There are structural differences between the two emotions, for which we will be unable to account with this conception of them. Instead, I am going to argue that we should conceive of them as two distinct (although similar) types of emotion. My task here, therefore, will be to show the difference in the paradigm situations by reference to which we view cases as examples of embarrassment and shame, respectively. I will begin by noting some of the differences between the types of situations in which we typically describe our emotion as embarrassment, versus those where we would naturally describe it as shame.

First, the role of the audience seems to be different in shame and embarrassment. As I argued in the previous chapter, the audience in the central case of shame (the mature, autonomous kind) serves mainly a heuristic function, providing an external standpoint that we are able to adopt to view our actions in a more self-conscious way. No real or even imagined audience is necessary for shame to be felt. In embarrassment, by contrast, the audience seems to play a much more central role, to the extent that it seems impossible to truly feel embarrassment without one. When people do something potentially embarrassing in public, such as tripping and falling in the street while running, they will often quickly look around to see if anybody has seen them do it; embarrassment seems to be much more likely if somebody has indeed seen them trip. Whilst one may feel embarrassment in advance of knowing whether anybody has seen or not, this seems to be due to the uncertainty; once it has been established that in fact no one did see, embarrassment seems a much less intelligible response. Although it may be possible to feel some embarrassment over acts committed in private, this would seem to be a very weak form, involving an imagined audience; the central thoughts seem to be around how embarrassing it *would have been* if somebody *had* seen one do it.

Also, the composition of the relevant audience seems to differ in each case. In embarrassment, it seems that we would be more embarrassed when we do things in front of

---

[115] Taylor, 69.

strangers or casual acquaintances, than in front of close friends or family members. In shame, the opposite seems true. It is hearing, or imagining what the people closest to us say, or would say, about our behaviour that seems most likely to evoke shame; the views of strangers and casual acquaintances can often be shrugged off much more easily.

Second, we might consider the way in which embarrassment appears to be contagious, in a way that shame does not. In the previous chapter, I mentioned how shame can be felt over the behaviour of others, if there is a sufficiently close link between us and them, so that their behaviour results in us failing to be as we should. It seems, however, that we can be embarrassed over the behaviour of others, in a couple of different ways. If we see somebody become embarrassed this may lead us, too, to be embarrassed, without it necessarily being the case that the actions by which this person embarrassed himself reflect on us, in any way. Alternatively, we may feel embarrassed *for* somebody who does something we regard as embarrassing, but fails to feel embarrassed. Again, no special connection between us and them seems necessary. It is a fairly common phenomenon to feel embarrassed with or for characters in novels and films (and, in a particularly modern example, participants in television 'reality' shows), in cases where we certainly would not feel shame on learning of venal or wicked behaviour on their part. Embarrassment, then, can easily be catching whereas this seems not to be a feature of shame.

Third, as Helen Block Lewis has pointed out, our concept of embarrassment seems to involve the notion of loss of power to act in an appropriate way:

> [Third party material excised][116]

In my discussion of globalised shame in the previous chapter, I suggested that one possible symptom of this type of shame is being unable to form a plan of rectifying action, due to the shame's lack of specificity to a particular aspect of one's character or behaviour. However, this seems a contingent feature of shame, attaching only to one (pathological) way in which we may experience that emotion. In embarrassment, this feature seems much more central to the emotional experience.

Each of these differences seems to suggest that embarrassment is somehow more concerned with surface appearance than shame. Perhaps the central difference, then, is this: embarrassment is concerned with how we take ourselves to be appearing to others, shame

---

[116] Helen Block Lewis, *Shame and Guilt in Neurosis* (Madison Conn.: International Universities Press, 1971), 74.

with how we take ourselves (or suspect that we ought to take ourselves) to really be. Consider an example. I am at a dinner party, engaging in light-hearted chatter, when I make a passing remark which others construe as carrying racist or sexist undertones. While I myself believe that I am not racist or sexist, and that my comment did not convey racist or sexist convictions, I can see how my fellow guests may have interpreted it that way. In the central case, it seems unlikely that I would feel shame here, as I do not believe that I have expressed any underlying racism or sexism in my character. Of course, I might suspect that, despite not having conscious racist or sexist intent, my words betrayed some subconscious prejudice, and feel ashamed about that. Alternatively, I may feel shame for some other reason, such as my inability to express myself in a way that avoids upsetting, offending or embarrassing my hosts and fellow guests. These subtleties aside, however, the central message is that shame for my actions would be neither appropriate nor intelligible in such a situation.

The emotion that may well be appropriate, however, is embarrassment. This seems to be a reaction to cases where we have inadvertently presented ourselves to others in a way in which we do not wish to be seen. In the example I have just described, I may feel that due to my comments, people now see me, incorrectly, as harbouring sexist or racist attitudes. Embarrassment need not necessarily follow, however; I may be able to simply and effectively clarify my comments, and so convince others that the perception they took from me as a result of this was incorrect. Where embarrassment seems more likely is where I am unable to correct their perception of me in this manner. This may be because it would be socially unacceptable to do so; to take some other examples, perhaps I am attending a funeral service or a theatre performance, so speaking to explain myself is, for the time being, inappropriate. Or perhaps, as in the case of tripping and falling in the street, explaining myself would be socially awkward, and may even lead to more embarrassment: if after I trip I try to explain myself to onlookers by saying things like "I don't usually trip" or "I am actually quite well coordinated" or "Look! My shoelace was untied!" this might lead them to see me as slightly mad, as well as physically ungainly or ridiculous, so creating more scope for embarrassment.

In most cases where we embarrass ourselves by our words, however, we are embarrassed not because we cannot put our own case in order to correct the mistaken impression others have of us, due to situational or social norms, but because when we try to do so, we are unable to find the right words to convince the onlookers that their perception of us is mistaken. One familiar case is the person who gets himself 'into a hole' by saying something embarrassing, and then 'keeps digging' i.e. makes the situation worse by trying to explain himself, only

succeeding in embarrassing himself further. In part, this phenomenon is due to the nature of the emotion; by attempting to explain oneself, one is indicating a certain understanding of why one's actions could be taken to show one in the relevant unflattering light, so adding the potential for further embarrassment. Also, the feeling aspect, the physical and psychological effects of embarrassment, can make it difficult to focus on finding the right words to correct the mistaken impression of you, even when such words are available. The salience-altering characteristics of emotional experience, in this case the embarrassment, may also lead one to see one's attempts to extricate oneself from the embarrassing situation as more embarrassing than one would if one were not already embarrassed, so allowing for embarrassment to whip itself up.

Of course, embarrassment can also be felt by the revelation of things that are true, but which could lead others to form a mistaken view of the agent. Think of the teenage boy who sees himself as fairly independent, and wishes to be seen to be so by his peers. When it is revealed to the boy's friends that his mother takes him shopping and chooses all his clothes, the boy may feel either shame or embarrassment. If the reactions of the other boys lead him to see that he is not quite as independent as he had liked to think, he may feel shame because of this. If, however, he continues to believe that having his clothes bought for him is a minor thing, that does not really bear on how independent he is, then he is unlikely to feel shame, but may feel embarrassment, if he believes that his friends will, on the basis of this information, assume that in fact he is not very independent.

Sometimes people become embarrassed simply because they have attention focused on them. One common type of case is of people being embarrassed when they are singled out for praise, or given an award or a birthday card. My notion of embarrassment can cope with this sort of case, as it has to do with the knowledge that people's attention is focused on you, so they are liable to be forming, or operating with, a view about what you are like, but the situation is out of your control, and so the type of view they are forming of you is similarly out of your control; there is therefore room to suspect that they are forming a mistaken image of you, which you are unable to correct.

Of course, this is not to say that in all cases like this, the emotion that is being felt will necessarily be embarrassment, rather than shame. We can easily tell a story where public

praise leads one to view one's situation as fitting a paradigm scenario for shame, and so feel shame itself. Deonna et. al. give the following example [Third party material excised][117]

We should also note that embarrassment may be felt when people mistakenly hold a view of you that is neutral, or even positive, rather than negative. One example of this was an incident during the BBC's coverage of the Beijing Olympic Games, when a presenter, the former British swimmer Steve Parry, took a lifesize cardboard cut-out of the multiple gold-medal winning American swimmer Michael Phelps into Tiananmen Square to do a piece to camera. Parry was immediately mobbed by locals who believed him to be Phelps. Unable to effectively communicate to the crowd that he was not Phelps, Parry became visibly embarrassed. Presumably this was not because he thought being seen as (arguably) the world's greatest swimmer was a negative perception. We can, however, account for this by saying that Parry's embarrassment was not over the perception of him by the locals who thought he was Phelps, but that of other on-lookers, who knew who he was and may have suspected him of deliberately giving the impression that he was Phelps for some reason of his own.

These observations, then, suggest the kind of paradigm situation that is at the basis of our concept of embarrassment. It will involve the person experiencing the embarrassment believing, or reasonably strongly suspecting, that because of some action of themselves or others, other people have come to hold a view of them that is incorrect. It will also involve an inability to easily rectify this perception.[118]

This conception of embarrassment allows me to account for the differences with shame that I pointed out above. First, and most obviously, an audience is pretty much required in embarrassment, because that emotion is about how we appear to others. We can make sense of why embarrassment tends to be more acute in front of those we know less well by noting that our close friends and family have a better grasp on our characters, and so are less likely to mistakenly take our actions as indicating some serious mistaken aspect to our characters.

---

[117]Julien A. Deonna, Raffaele Rodogno and Fabrice Teroni, *In Defense of Shame: The Faces of an Emotion* (Oxford: Oxford University Press, 2011), 117.

[118] This characterisation of embarrassment is similar to that offered by J. David Velleman in "The Genesis of Shame" (*Philosophy and Public Affairs* 30:1 (2001), 27-52), where he argues that a central cause of shame is failures of self-presentation. I would argue that while these cases can cause shame as well, embarrassment is the more usual reaction, and in any case Velleman's conception of shame is too narrow to account for some significant central cases of that emotion.

Second, embarrassment seems contagious because it creates a demand for action; action to clarify the situation, to show the onlookers that their perception is unjustified or mistaken. When somebody embarrasses themselves, this demand for action falls not just on them, but it can fall on onlookers as well, and if we cannot think of any way to act appropriately as an onlooker, this might lead us to become embarrassed as well.

Third, the sense of inability to act as required that seems central to embarrassment is explained by the need to correct the mistaken perception that people have of us. As I discussed above, experiencing embarrassment makes it difficult to do this, as it distracts us from thinking about the most effective way of doing this, makes us see our attempts to do so as even more embarrassing, and (often) puts us in a situation where anything we do say only digs us deeper into the hole of embarrassment.

Obviously, while I have given a conception of embarrassment that sees it as a different emotion to shame, the two are very much related. Both are emotions about the self, and one's sense of what constitutes a positive and negative conception of oneself are relevant to each. As my example of the teenager whose mother buys his clothes suggests, there are cases in which both embarrassment and shame can be intelligible reactions to the same situation. In such situations, one might feel just one of these two emotions, or vacillate between the two, or feel both, over subtly different aspects of the same situation. Given the similarity of the two emotions both in physical manifestation and the mitigating action that they would render intelligible, it can be difficult to decipher whether somebody is feeling embarrassment or shame, even for the person who is having the emotional reaction. This latter point is emphasised when we consider that, due to the more personal, shattering nature of shame, one may have an incentive to redescribe or pass off one's shame as mere embarrassment.

I would argue that there is a clear conceptual distinction between shame and embarrassment, rooted in the idea that the former is based on how one actually is, whereas the latter focuses on mere appearances. However, it is clear that in practice there will often be a great deal of uncertainty and vagueness over which of the two emotions is appropriate, or is being felt, at any particular time, not least on the part of the agent himself. As Deonna et. al. note, [Third party material excised][119]. Though these two emotions are easily confused at the margins, we should not be seduced into the idea that they are really two cases of a single emotion.

---

[119] Deonna, Rodogno and Teroni, 117.

Turning now to humiliation, again the task is to see if I can build a plausible paradigm scenario for this emotion that is distinct from those I have posited for shame and embarrassment. If I can show that there is such a distinct paradigm scenario, this will suffice to show that humiliation is a distinct emotion from both shame and embarrassment.

Any purported distinction between shame and humiliation, however, is minimised by Nussbaum, who seems to view humiliation as nothing more than a subset of shame. She writes that she understands humiliation [Third party material excised][120] On this definition, however, humiliation comes out not as an emotion but as a state, albeit one that we are able to understand only if we already understand what it is to be in a position to have one kind of emotion: shame. One is humiliated, for Nussbaum, when one is put in a position that exposes one to shame.

Indeed, for Nussbaum, one may be humiliated without having any negative emotion about oneself at all: [Third party material excised][121] However, as well as *being* humiliated, Nussbaum acknowledges that we may speak of *feeling* humiliated. This is an emotion, she says, [Third party material excised][122] This seems to say that humiliation, as a feeling, is nothing more than a particular type of shame: that felt due to the deliberate act of another in exposing one to that shame.

While she never says so explicitly, the idea that Nussbaum sees humiliation as a subset of shame is given credence by her claim that [Third party material excised][123]

Gabriele Taylor, on the other hand, sees humiliation as a different, though related, emotion to shame. As in shame, she argues, [Third party material excised][124] In humiliation, unlike in shame, however, it is a diminishment of one's status in the eyes of others that is central, rather than one's newly lower status.

Taylor goes on to suggest that the central point in humiliation is to do with how one appears, rather than how one actually is. [Third party material excised][125] This would contrast with shame, at least of the central, morally mature variety, as there it is how one actually *is* that matters. One will only feel shame if one agrees with the negative assessment of oneself (or, at

---

[120]Nussbaum, *Hiding*, 203.
[121]Nussbaum, *Hiding*, 203-04.
[122] Nussbaum, *Hiding*, 204.
[123] Nussbaum, *Hiding*, 204
[124] Taylor, 67.
[125] Taylor, 67.

the very least, suspects that it may be correct). On Taylor's account of humiliation, however, one's own view is irrelevant; the centre of humiliation is a judgement by others that one's status is lower than they had previously taken it to be.

In addition to this, Taylor identifies another element that is central to humiliation, but not to shame: that one will be thought presumptuous by others for having pretended to the previous, higher position, whether or not one agrees with this assessment oneself:

> [Third party material excised][126]

Both Nussbaum's and Taylor's accounts, while capturing something true and important about humiliation, fail to get it completely right. I will give two examples that we can straightforwardly understand as cases of humiliation, but which are inadequately characterised by either of these definitions. Of course, it may be that the word 'humiliation' is philosophically imprecise, and applies to more than one type of emotional experience. However, if we can find a definition that adequately covers both of my examples, this will be a mark in its favour.

The first example is one of what I shall call 'revelatory' humiliation. In this case, we imagine a young person who has nervously brought a first boyfriend or girlfriend home to meet the family. The young person's mother, in discussion, deliberately brings up some normal but embarrassing incident from childhood, and the young person feels humiliated.

In the second example, which is a case of 'active' humiliation, a third party actively brings it about that a person is in a humiliating position. A real life example is the case of the Israeli soldiers manning a roadblock who forced a Palestinian man carrying a violin to play for them before he was allowed to cross[127]. Leaving aside the particular historical resonances of such an act, we can understand why the Palestinian man may have felt humiliated by this.

Nussbaum's and Taylor's accounts each have difficulty in accounting for these cases. First, Nussbaum's identification of humiliation with shame does not seem to fit. In each case, we could understand the person feeling humiliated, but thinking that there was nothing to be ashamed of, and feeling no shame. In the 'revelatory' example, it would be natural to think that everyone was immature once, and there is nothing shameful about having been so. In the 'active' example, one might think that there is no shame in being forced to do something by

---

[126] Taylor, 68.
[127] See, for instance, "Israel shocked by image of soldiers forcing violinist to play at roadblock", *The Guardian*, 29 November 2004.

people in a powerful position who are abusing their power. Of course, shame could be present in such cases, and could even be appropriate. Nonetheless, we can make sense of somebody making the type of statement set out above, yet still feeling, and recognising that they are feeling, humiliation. While Nussbaum might argue that this is a case of conceptual misunderstanding, I think we understand such cases readily enough, that they support my contention that we recognise a genuine distinction between shame and humiliation.

Second, Taylor's claim that humiliation always involves the humiliated person believing that the audience is seeing him as having pretended to a higher position than he merited does not seem to fit with these examples either. Surely neither the young person bringing the boyfriend or girlfriend to meet the family, or the Palestinian forced to play the violin need to believe that others think they had pretended to a more elevated status than they are now shown to occupy. As I said before, we were all immature once, and so the notion that the young person was pretending to the status of never having been so is absurd. In the case of the Palestinian, it seems that the status to which he has been reduced is that of someone who is unable to act autonomously, but surely no-one could think that it displayed presumptuousness on his part to have presented himself as an autonomous agent in the past.

Even though I do not think that either of these views can make sense of these two examples, we can take something from each of them to come up with a more robust conception of humiliation. From Taylor's view I take the idea that humiliation is somewhat like embarrassment, as it is the appearance to others that counts rather than the person's own assessment. This seems to be true of both my examples, as in each case we can imagine the person saying they feel they have nothing to be ashamed of, but feeling humiliated nonetheless. From Nussbaum, I take the idea that humiliation necessarily involves another person acting deliberately to make others see one as "lesser" in some way. In the revelatory example, it is vital that the mother acts as she does to deliberately humiliate the young person. If the damaging information had come out by an unintentional word from the mother, or the young person, or in some other accidental way, then humiliation, I suggest, would not have been an appropriate response (absent further information); embarrassment would be much more fitting.

Therefore I conclude that the paradigm situation at the heart of humiliation is the same as that for embarrassment, except that the lower image others make of one is cruelly and deliberately created or fostered by a third party. Again, there will clearly be links with shame. For

instance, one may feel ashamed by one's lack of ability to prevent one's own humiliation. However, we should not confound the two – they are distinct emotions.

One possible objection to this characterisation of humiliation is that we sometimes speak of humiliation in cases where there is no deliberately cruel action by others, or even no action of others of any kind. For instance, we can make sense of the humiliation of a friend who has been talking himself up as a runner for weeks, but then performs badly in a race. In a sense, his humiliation is due to the better performance of the other competitors, but they did not do anything unjustified and they certainly did not set out to humiliate him. Alternatively, we can imagine a case where someone with no agency by others at all; for instance, the friend might have been boasting about his time for the mile and then perform poorly against the clock, resulting in humiliation. How are we to account for these cases?

If I want to hold on to the conception of humiliation that I have offered, I have three broad options for responding to non-standard cases such as these, which are to argue that they are unintelligible, inappropriate but intelligible, or rational but non-standard. The claim that these cases are unintelligible as experiences of humiliation is unattractive; such cases are sufficiently familiar to us to invalidate any understanding of humiliation that completely fails to accommodate them. I will therefore disregard this option.

To claim that in these cases humiliation is inappropriate but intelligible would be to say that the situations cannot be seen as appropriately analogous to the paradigm scenario, because of the lack of deliberate agency of others, but we can understand how somebody could see it this way. In the examples I gave, the subject might be seeing the situation as one in which he has been deliberately belittled by others (such as the friends to whom he had boasted and who then called him on his claims). Under the model of the emotions I am offering, of course, this does not require that the subject believes this to be the case, only that he sees the situation in this light. We can certainly posit unconscious motivations for seeing the situation this way; while humiliation is an unpleasant emotion, it at least offers some consolation in comparison to shame, in that its focus on appearances means that the subject can suggest it is not reflective of the underlying reality. Equally, humiliation offers the subject someone other than himself to blame. We may therefore argue that humiliation in such cases, while unjustified, is both comprehensible and explicable.

Alternatively, we might conceptualise these cases as appropriate cases of humiliation by arguing that although the cases differ from the paradigm case in lacking the element of

deliberate agency of others, there is a legitimate sense in which they can be seen as appropriately analogous to the paradigm scenario. One way to make this argument is that the subject's view of the situation involves taking an external view of the boastful side of his personality, which he sees as deliberately and foolishly exposing his inadequacy; on this interpretation, the subject would feel humiliated by himself. Another possibility would be that the subject is personifying fate, which he sees as having humiliated him. Again, it is important to note that the subject does not need to believe these scenarios, but merely must see the situation as conforming to them.

There is, in the greater scheme of things, little to choose between these two options. Both of them are preferable to adjusting the paradigm scenario to attempt to encompass these cases. One of the benefits of the paradigm scenario view of emotions is that it allows us to maintain a clear view of what constitutes the central cases of emotions, while maintaining an open mind on whether more marginal cases apply. It will often be a matter of fine interpretation whether a particular non-standard case counts as unintelligible, inappropriate or appropriate qua a particular emotion, but the additional space the paradigm scenario view gives us to make this judgement, I contend, fits better with the way we use emotion terms in normal discourse.

One may legitimately ask whether I am setting up my account of the individual emotions in such a way as to make it impossible to refute – if anyone points to an apparent example of an emotion that is not analogous with the type of paradigm scenario I have placed at its centre, then what is to prevent me from simply describing it as a "marginal" case? My answer to this is that nothing is preventing me from doing so. We should expect there to be a penumbra of marginal cases on which reasonable people disagree, for two reasons. First, the definition of the emotions is, on my view, a folk psychological concept. While what the generality of people think differentiates the emotions should not be the last word, it probably should be the first. These views also shift over time, so we would expect reasonable people to disagree on what falls under them. Second, the notion of being "appropriately analogous" is inevitably going to be one of degree, with a realm of vagueness. Therefore there is nothing to stop me from claiming that a particular case is "marginal" other than an opponent saying "no it is not".

Is this problematic for my argument? Given that I am going to go on to argue that we need shame alongside guilt, due to the fact that the former can be appropriate in certain situations

where guilt cannot, might this vagueness in the emotions undermine my arugment? I do not think so; the fact that there are marginal cases in this penumbra of vagueness does not change the fact that there are central cases with which it would be hard to disagree that they are examples of the relevant emotion, just as there are ones that are clearly not such cases. So long as the cases I am relying on to make my argument fall clearly within the scope of shame and outside the scope of guilt, the existence of other, marginal cases should not be of great concern. Whilst it is folly to seek an excessive degree of precision in this area, this does not mean that anything goes.

So, to summarise, I have argued that emotion and humiliation differ from shame in being concerned with appearances rather than (and sometimes, in opposition to) the reality of the self. They differ from each other in terms of the deliberate agency of others that the subject sees in the situation. Of course, it may be argued that I have failed to distinguish embarrassment and humiliation from all kinds of shame. I have tried to show how they are distinct from mature shame, but have not mentioned the immature heteronomous variety of shame. According to the view I am offering, in small children embarrassment and humiliation, in their mature forms, are essentially impossible, due to a lack of a firm self-conception and a full view of how they are seen by others. In such cases, reactions that we might refer to in adults as shame, humiliation or embarrassment are really all of one kind: the experience that I have called immature shame. In each case the paradigm scenario is of somebody making a negative judgement of one. It is only as we mature, and the mature form of shame comes to the fore, that these three emotion-types become distinct.

Guilt

I now turn to the final emotion that I wish to contrast with shame: guilt. Unlike shame, guilt can name an objective status as well as an emotion: one is guilty if one breaks the law or some other injunction with a similar status. While being guilty is often an important part of the explanation of why someone feels guilty, the objective state and the emotion are not co-extensive; one can *be* guilty without *feeling* guilty and vice versa. Here I will concentrate on the *emotion* of guilt, which is sometimes known as guilt-feeling or remorse (e.g. as in Gaita).[128]

---

[128] Raimond Gaita, *Good and Evil: An Absolute Conception* (Abingdon: Routledge, 2004). I am aware, however, that some writers (e.g. Taylor, Greenspan) would dispute my identification of guilt and remorse. I will address this point in the next chapter.

Shame and guilt are, of course, quite similar in some important ways. They are both negative emotions that involve negative judgements being passed on ourselves. As a result, they are both sometimes invoked as 'moral emotions'. We do, however, distinguish between these two emotions in everyday life, and this can lead to differential recommendations as to the role the two emotions ought to play in moral life. In order to respond to these recommendations, it is important to distinguish guilt from shame. As may be expected from the foregoing, my strategy will be to consider the essential differences between paradigm scenarios at the centre of each of shame and of guilt.

Shame and guilt are often not very well distinguished in the popular mind. For example, Tangney and Dearing, experimental psychologists working in the area of shame and guilt, note that [Third party material excised].[129] However, where serious attempts have been made to distinguish between shame and guilt, we can distinguish a number of putative differentiating characteristics.

First, it is sometimes claimed that shame is a social emotion, whereas guilt is an individualistic one. There are two ways in which we may take this claim. On the one hand, it could mean that shame always requires an audience (either real or imagined) whereas guilt does not. Alternatively, it could be claiming that shame is essentially heteronomous, whereas guilt is essentially autonomous, in the sense that I described in the previous chapter (i.e. shame is brought about by the negative assessments of others, guilt by the negative assessment of oneself).

Second, it may be claimed that shame and guilt are elicited by different types of wrongdoing or failure. Again, this can plausibly be read two ways. The claim may be that different types of concrete behaviour elicit guilt and shame; for instance, it might be claimed that cruelty to animals elicits shame, whereas cruelty to human beings elicits guilt. Alternatively, it might be claimed that shame and guilt are elicited by different formal objects: shame is elicited by a failure to live up to an ideal, guilt by the transgression of a norm.

Third, shame may be said to differ from guilt in that the former relates to the whole self whereas the latter relates only to some deed of the self. There are numerous ways in which this claim might be cashed out.

---

[129] June Price Tangney and Rhonda L. Dearing, *Shame and Guilt* (New York: Guilford, 2002), 24.

I will assess these claims in turn, in order to see if, and if so in what sense, they are distinguishing factors between shame and guilt.

Distinguishing Shame and Guilt: Social vs. Individual

One way in which writers have attempted to mark out the distinction between shame and guilt is through the claim that shame is a social emotion, whereas guilt is an individual one. As Deonna and Terroni note, [Third party material excised] [130]

This type of claim finds expression in the work of the experimental psychologists Gehm and Scherer, who report that [Third party material excised][131] The idea here seems to be that actual public exposure of one's wrongdoing or failure is required in shame, whereas this is not the case in guilt.

A similar, though more nuanced claim, is made by the social anthropologist Helen Benedict:

[Third party material excised][132]

Again, the claim is that shame requires an audience, whereas guilt does not, but here, unlike in Gehm and Scherer's claim, it is conceded that the audience that is required for shame may be imaginary, rather than actual.

A similar claim goes beyond the requirement of an audience for shame to be felt, to claim that shame is essentially heteronomous, whereas guilt is essentially autonomous. As discussed in the previous chapter, this means that shame depends on the judgements of others about us, whereas guilt depends on our own judgements about ourselves. This point of view is expressed in Wallbott and Scherer, where they claim that [Third party material excised].[133] This goes beyond the claim about the requirement of an audience in shame and guilt, by suggesting that the negative judgements of the self that are central to cases of shame and guilt come from two different sources: from others in the former case, and oneself in the latter.

---

[130] Deonna and Terroni, 729.

[131] Theodor L. Gehm and Klaus R. Scherer, "Situation Evaluation/Emotion Differentiation" in Scherer (Ed.) *Facets of Emotion: Recent Research* (Hillsdale: Lawrence Erlbaum Associates, 1988), 74.

[132] Ruth Benedict, *The Chrysanthemum and the Sword* 2nd edn (New York: Mariner, 2005), 223.

[133] Harald G. Wallbott and Klaus R. Scherer, "Cultural Determinants in Experiencing Shame and Guilt" in *Self-Conscious Emotions: The Psychology of Shame, Guilt, Embarrassment and Pride* ed. by June Price Tangney and Kurt W. Fischer (New York: The Guilford Press, 1995), 465-487 (474).

As will be apparent from the previous chapter, I do not find these attempts to distinguish between shame and guilt attractive. These claims all mischaracterise shame; as I have argued, it does not require the presence of an audience, either physically or in imagination, nor is it an essentially heteronomous emotion.

As I argued in the previous chapter, while audiences, either physical or imagined, have a particular role to play in many central cases of shame, they are not essential to the experience of shame. In some cases, being seen by others in a certain way (e.g. being seen naked, or being seen as an object of sexual desire) may be what one is ashamed of, and so a physical audience is obviously required. Clearly, however, this is true of only a subset of the possible things of which one might be ashamed. Equally, the same could be said of guilt: one might feel guilty for having indulged one's exhibitionist tendencies, and obviously an audience is required for such a situation (it is difficult to be exhibitionist on one's own), but this cannot be taken to say that guilt necessarily requires the presence of an audience. Thus the requirement of a physical audience cannot be used to distinguish shame from guilt.

The idea of a more thoroughgoing requirement for the presence of an audience, either real or imagined (or even imaginatively constructed out of one's own values, as Williams suggests), is true only of hetereonomous types of shame. In autonomous shame, what counts is the subject's own assessment, with no need for an intervening observer. While in autonomous cases, imagining an audience may be a useful heuristic to help us gain the overall view of ourselves required to judge that we are failing to live up to the way we should be, I see no reason to deny the possibility of reaching such a judgement without relying on this heuristic, and so feel shame without imagining the reactions of an audience. As Deonna et. al. argue, [Third party material excised][134].

This argument is backed up by experimental findings. Tangney and Dearing conducted experiments in which they examined narrative accounts of real-life pride, shame and guilt experiences provided by subjects. They assessed whether others were aware of the respondent's emotion-eliciting behaviour. They found that:

> [Third party material excised][135]

The key point for my purposes is that when respondents described times when they felt ashamed of something, other people were no more likely to be aware of the cause of their

---

[134] Deonna, Rodogno and Teroni, 139.
[135] Tangney and Dearing, 15.

emotion than in cases of guilt. While there are clear limitations to the reliability of self-reported emotional narratives, this seems to be the only experimental route open to us in this area, and the support of these findings bolsters my own arguments in the case of a physical audience, at least; others do not need to 'observe' our actions for us to feel shame, any more than they do for us to feel guilt.

The leading candidate for a fundamental distinction between the type of paradigm scenario that is acceptable for shame as opposed to that for guilt, is heteronomy versus autonomy, finding its expression in the claim that one feels shame based on negative judgements of others about oneself, whereas one feels guilt based on one's own negative judgements about oneself. As Deonna and Terroni put it:

> [Third party material excised][136]

In the previous chapter, I argued that shame is not necessarily heteronomous; rather the standard, mature form of the emotion is autonomous. In immature moral agents shame is heteronomous, and this immature reaction can resurface from time to time, even in agents who otherwise react maturely. Apart from this, there is a mature heteronomous type of shame where the agent does not judge herself negatively, but, due to the negative judgements of others, suspects that she may deserve negative judgement, and feels shame on this basis.

This more nuanced picture of the autonomous/heteronomous possibilities in shame means that we cannot use this supposed distinction as a way of distinguishing between shame and guilt. In addition, however, I would question whether guilt can be described as genuinely autonomous. It may be the case that the immature heteronomous form of shame has no analogue in guilt, as to feel guilt we must be capable of recognising the supposed transgression we have made as being a transgression of a norm with genuine force (more on this later).[137] However, I see no reason to deny that the mature type of heteronomy is possible in guilt. One way this may work is if we are involved in an incident, for which we do not think we are responsible. If others were to criticise us for this (or we were to imagine them doing so) this may lead us to feel guilt. It is likely, however, that this guilt would lead us to focus more closely on the relevant aspects of the incident, and enable us to come to our own

---

[136] Deonna and Teroni, 729.

[137] Of course, it is still possible to feel guilt over 'transgressions' one no longer views as transgressions, such as a person with a strict Catholic upbringing, which he has now rejected, feeling guilty for eating steak on a Friday, but this type of phenomenon, referred to in the last chapter as 'recalcitrant emotions' in is common to all emotions, and is not 'radically heteronomous' in the same way as those cases now under discussion.

view about whether guilt is appropriate. One might conceptualise "survivor's guilt" as a form of mature heteronomous guilt – despite not being able to pick out any norm that they transgressed, survivors may experience guilt due to the idea that they should have done something differently, perhaps reinforced by thoughts of others asking "what did you do to survive?". Alternatively, it may be that survivor's guilt is best understood as a variety of shame/humiliation rather than guilt. My best stab at an answer to this extremely difficult question is that it is probably a complex emotional syndrome that encompasses all of these individual emotions, and probably more besides (anger, envy etc.).

I thus conclude that none of the options I have considered under the heading of social vs. individual are able to adequately distinguish shame from guilt.

Distinguishing Shame and Guilt: Eliciting Situations

A second way of attempting to distinguish between shame and guilt consists of saying that different types of situation are fitted to elicit the two emotions. This option can be divided into two different sub-options.

The first sub-option is to say that different types of concrete behaviour elicit different types of fitting emotional response. For instance, we might say that a negative judgement passed on oneself as a result of one having told a lie, say, will elicit guilt, whereas a judgement resulting from kicking a dog will elicit shame. For this type of distinction to be anything other than ad hoc, we would need to be given some story about why different types of behaviour fall on either side of the line.

Irrespective of what story we may be told in this regard, however, this strategy is unpromising, for the simple reason that certain types of situation can clearly elicit either of guilt or shame (or perhaps both at once). For instance, we can easily imagine somebody who fails to go to the aid of somebody being beaten up in the street feeling either guilty, for their failure to help the person in need, or shame for their lack of physical courage in this situation. Whether or not we think one emotion may be more appropriate than the other in this situation (a question to which I will return in the next chapter), it seems clear that we can make sense of this situation eliciting either emotion.

This intuition appears to be supported by empirical research. For example, Tangney and Dearing, in their research into self-reported shame and guilt narratives, found that:

> [Third party material excised][138]

These results seem to support my key contention here that shame and guilt can both be intelligible reactions to the same circumstances.

I do not wish to deny that there are certain circumstances in which one of shame or guilt is a more likely, or a more appropriate response. Nor do I wish to claim that in any situation where either shame or guilt is an intelligible response, both of them are equally likely or appropriate; I will return to these questions in the next chapter. However, it is sufficient for my purposes if we are convinced that there are at least some circumstances in which either emotion would be intelligible. This being the case, the two emotions cannot be distinguished in the way proposed.

An alternative way of making this distinction would be to say that the emotions are to be distinguished by the different formal objects that they take. The paradigm scenario at the centre of shame, as I argued in the last chapter, involves some situation being seen as representative of a failure to live up to some ideal of how one thinks one ought to be. This may be contrasted with guilt, which is said to involve transgressions of norms.

This type of distinction has a number of proponents in the literature. For instance, Wollheim picks it out as one of the important differences between the two emotions, writing that [Third party material excised][139] Wollheim writes of "criticism" invoking ideals and injunctions in the two cases because he regards both shame and guilt as essentially heteronomous. I have discussed this point at length in the previous chapter, and given my reasons for rejecting it, but the point we are interested in at this stage, relating to failure to live up to ideals, versus transgressing injunctions, still stands.

A similar point is made by Sandra Bartky. She quotes with approval Deigh's claim that [Third party material excised], before adding that:

> [Third party material excised][140]

---

[138] Tangney and Dearing, 17.
[139] Wollheim, 155.
[140] Sandra Lee Bartky, *Femininity and Domination: Studies in the Phenomenology of Oppression* (New York: Routledge, 1999), 87.

Deonna and Terroni summarise this position as the claim that [Third party material excised] before adding that [Third party material excised][141]

I am inclined to agree with Deonna and Terroni that this kind of approach is promising. While we can, of course, feel shame for transgressing a norm, this is because of what this behaviour says about us in relation to our values. As I will go on to argue, the crucial difference is not in what causes the emotion, but the way one sees oneself in relation to that cause.

We must, however, be careful how we interpret this distinction, so as to avoid falling into an overly simplistic view of each emotion. This danger is highlighted by Gaita, who gives the example of a man who murders his lover's husband so that they can live together, but finds himself wracked with guilt[142] and so unable to live the carefree life for which he committed the murder in the first place. Gaita writes:

> [Third party material excised][143]

What this points to is the danger of taking the claim about the formal object of guilt, and for that matter shame, as indicative of the conscious thoughts a person will have when experiencing each of these emotions. When a person feels guilt it will be because he has seen that he has transgressed some important norm or rule, but to insist that the thought that this is the case must be consciously before him, or even that it must be the thought that is most prominently before him, risks creating a parody of the thoughts of a person experiencing guilt. Rather, central thoughts include 'how could I have done this?', 'why did I do it?' and thoughts about the negative effects of the relevant actions on others. In short, we need to remember that the transgression of a norm is a formal constraint on what type of situation may be at the centre of guilt, not a description of the psychological focus of the person experiencing the emotion.

A similar point may be made for shame: it would be absurd to imagine the person experiencing shame thinking 'Damn, I really failed to live up to one of my ideals there". Instead thoughts are likely to focus on the particular situations in which this became apparent, what one should have done, etc. While the nature of the paradigm scenario guides the direction of one's conscious thoughts onto the relevant situation in the real world, and is thus

---

[141] Terroni and Deona, 734.
[142] Which, as I mentioned above, Gaita identifies with remorse.
[143] Gaita, 234-35.

able to address the attentive frame problem, it does not, in the normal run of things, constitute the content of those thoughts,

This way of differentiating guilt and shame in terms of the types of situation from which it may arise, does not fall foul of the same difficulties as the proposed distinction in terms of types of concrete eliciting circumstance described above. The reason for this is that a single situation may be seen as either a transgression of a moral principle or a failure to live up to an ideal (or both). As Bartky notes:

> [Third party material excised][144]

A further point on which it is important to get clear is that guilt does not, of course, require a genuine transgression of a norm; all that is required is that the person who is subject to the guilt sees their action as a transgression of a norm. This is just another way of expressing the thought I raised at the beginning of this chapter: one may feel guilty (emotion) without being guilty (state). One form this phenomenon might take is somebody seeing something as a norm which, we think, ought not to be seen as such. For instance, somebody may feel guilty for dancing on a Sunday. The fact that we may not regard there as being any genuine norm against such behaviour does not mean that we are faced here with a case of guilt that counters the view of guilt as being about norm transgression; it simply indicates that the subject in question, but not us, sees dancing on Sunday as being contrary to a genuine norm (the stronger claim, that they *believe* that dancing on Sunday is contrary to a genuine norm, may or may not be true).

An alternative way in which this type of scenario may play out is that somebody is concerned with a genuine norm, but sees themselves, mistakenly, as having transgressed it. For example, we could imagine somebody who sees themselves as having transgressed a norm against harming others by outperforming them in a test, and so harming their chances of getting a job they both want, and so feels guilty for this. We might think that they have nothing to feel guilty for, as their actions do not constitute a transgression against the relevant norm, but we can make sense of their feeling guilt, on the characterisation of guilt I have offered, if we regard them as inappropriately seeing their action as such a transgression. Ultimately, of course, one might think that these two types of situation are really the same, as the claim may be made that part of what it is to understand a norm is to understand what would constitute as transgressing that norm. I have some sympathy with this view, but we seem capable of

---

[144] Bartky, 88.

making a commonsense distinction between the two kinds of case, and not much hangs on whether they are genuinely distinct.

It seems then that we have found an attractive means of distinguishing shame from guilt: the paradigm situation at the heart of the former involves a failure to live up to ideals, whereas the latter's paradigm situation involves the transgression of norms. This distinction is compatible with my description of shame from the previous chapter, tracks common-sense intuitions, and does not seem open to counterarguments, when interpreted in the correct sense.

Distinguishing Shame and Guilt: Role of the Self

Before we pronounce this distinction the winner, however, there is one possibility that remains to be considered: that shame and guilt differ in respect to the role of the self in the two emotions. Specifically, the claim is that shame involves a judgement on the whole self, whereas guilt involves a judgement only on part of the self: the self is judged only insofar as it is the doer of certain actions.

Perhaps the most prominent proponent of this distinction is the psychologist Helen Block Lewis. She writes:

> [Third party material excised][145]

This thought is echoed by a number of philosophers who have made a similar distinction. For instance, Wollheim: [Third party material excised][146] In a similar vein, Williams writes that [Third party material excised][147], before going on to say that guilt looks primarily in the former direction, shame in the latter.

Taylor also makes the point about the different attitude towards the self involved in shame and guilt:

> [Third party material excised][148]

In guilt, she suggests, we are somewhat self-alienated from our actions; while taking responsibility for them, we do not view them as a genuine reflection of the person that we are.

---

[145] Lewis, 30.
[146] Wollheim, 155.
[147] Williams, *Shame and* Necessity, 92 (emphasis in the original).
[148] Taylor, 92.

In shame, on the other hand, our attention is on our actions and behaviour precisely insofar as they *do* illustrate the type of person that we are or have become, and it is this latter idea that is the focus of shame.

There also appears to be some empirical support for this distinction. For instance, Tangney and Dearing ran a series of experiments where they asked people to engage in counterfactual thinking about cases of shame and guilt they had experienced, and imagine how the situation would have had to be different to secure a different outcome (i.e. one in which they felt no shame or guilt).

[Third party material excised][149]

The idea that 'undoing' aspects of the self (e.g. 'If only I was not so selfish') is more commonly thought to be a way of avoiding shame experiences, whereas 'undoing' aspects of behaviour (e.g. 'If only I had not eaten her cake') is more commonly thought to be a way of avoiding guilt experiences, seems to speak in favour of the view under consideration.

This distinction, with its talk of shame as being about the "whole self" may be thought to be in conflict with the distinction between localised and globalised shame that I discussed in the last chapter. If this amounts to a claim that shame is always globalised, and in the last chapter I argued that shame is not always globalised, then I will be unable to accept this as a distinguishing factor between shame and guilt.

We have reason to believe, however, that the claim that shame is about the whole self as made by Lewis and others is not equivalent to the claim that shame is globalised, in the sense I set out. This claim is meant in the same way that, I argued[150], Wollheim and Nussbaum mean it; one's (whole) self is necessarily the object of shame (when one is ashamed, one is always 'ashamed of oneself') but it is the level of generality of the content of the relevant view of the self that determines whether an episode of shame is localised or globalised (e.g. whether one feels ashamed of oneself for being prone to greediness or for being such a terrible person). Thus this account of the distinction between guilt and shame is not in conflict with my earlier description of shame.

Following this through to the case of guilt, the argument would be that the object of guilt is always an action that the subject has either done or failed to do. This is not to say, however,

---

[149] Tangney and Dearing, 23.
[150] C.f. Chapter 3, above.

that guilt is not at all about the self; guilt, as I have said, is still an emotion of self-assessment. As Deonna and Terroni remark:

> [Third party material excised][151]

This being the case, how are we to understand the distinction that Lewis and others are pushing towards? Deonna and Terroni appear to be on the right track when they write that:

> [Third party material excised][152]

This allows us to mark out a way in which shame is especially concerned with the self, while maintaining the essential connection between guilt and the self, as well. This view distinguishes guilt from shame by arguing that, in my terms, the paradigm scenarios at the heart of each emotion will involve different things being judged negatively. In shame, the self is judged negatively (although this may be in respect of some behaviour or even some single action) whereas in guilt, an action is judged negatively (although this must be some action in which one is especially implicated in some way).

Understanding the distinction in this way allows it to avoid falling foul of the fact that either guilt or shame can be felt over a single situation. This is because a single action can be seen both as a transgression against a norm, and as a failure to live up to an ideal: I commit a bad act when I steal her piece of cake, and I also display my own dishonesty and unkindness. By making the distinction depend on where the evaluative focus of the emotion lies, we make room for this possibility.

It seems, then, that distinguishing shame and guilt by reference to the role of the self is an attractive route to take: it accords with our intuitions about the two emotions, it appears to have some empirical support, and it can deal with relevant objections.

Two Sides of a Coin?

This seems to leave me with something of a problem; in seeking for a distinction between shame and guilt, I have found two that seem promising – which one should I adopt? Should I say that shame and guilt are distinguished by the fact that one involves failing to live up to

---

[151] Deonna and Terroni, 730.
[152] Deonna and Terroni, 732.

ideals and one transgressing norms, or that one has an evaluative focus on the self and one on actions?

On closer inspection, however, this turns out to be a false dilemma, as the two distinctions are really two sides of the same coin. As Deonna and Terroni note, the two distinctions [Third party material excised].[153] Combining the two, we get a picture of guilt as involving a negative judgement on an action that is perceived as a transgression of a norm; this is to be expected, as only actions can count as a transgression of a norm. Similarly, we come to a view of shame as a negative judgement about the self for failing to live up to an ideal; again, this is as it should be, as only the self can fail to live up to an ideal. An action cannot count as failing to live up to an ideal – it can only make it the case that the author of that action has themselves failed to do so. This tight fit means that the two distinctions cannot come apart, so I am able to argue that they are both aspects of a single distinction between shame and guilt.

Adopting this distinction allows us to make sense of a number of commonly held beliefs about shame and guilt.

First, consider the widely held idea that guilt is restricted to the moral sphere, whereas shame ranges over both the moral and the nonmoral[154]. The distinction I favour can make sense of this, as, it may be thought, norms, the transgression of which are at the centre of guilt, deal exclusively with the realm of the moral. Ideals, on the other hand, which are the central concern of shame, can range over both the moral and the nonmoral. One of my ideals may be to be a considerate person, another to be a conscientious student; these two ideals seem to fall on different sides of the moral/nonmoral divide, as usually conceived.

Second, it is often thought that we can feel vicarious shame but not vicarious guilt; i.e. we can feel ashamed for something somebody else has done, but we can only feel guilty for our own actions. For instance, Taylor writes that:

> [Third party material excised][155]

My favoured distinction between guilt and shame allows us to make sense of this, too. As I argued above, in guilt the self limits the type of action over which one may feel guilty: I will only feel guilty if I believe *I* have transgressed some norm. In the case of shame, however, the

---

[153] Deonna and Terroni, 736.
[154] For instance, see Lewis: "Another way in which the multiple stimuli characteristic of shame have been described is to say that there is both moral and nonmoral shame. There is only one kind of (moral) guilt." (84).
[155] Taylor, 91.

range of possible ideals is sufficiently wide that the actions of another can be interpreted as my failing (in a passive sense) to live up to the ideals I have set myself – hence if one of my ideals is to be the father of a well-behaved child, the bad behaviour of my child may make it the case that I have failed to live up to this, and so elicit shame, without any identifiable action on my part being responsible for this.

Finally, we may consider the differing action tendencies that are often attributed to people experiencing shame and guilt. For instance, it is sometimes remarked that guilt often leads one to seek to make specific reparation for the wrong done, whereas this is less often the case in shame[156]. This, however is as we should expect, given the differing central objects of evaluation in the two emotions. As Taylor notes:

> [Third party material excised][157]

While I would dispute Taylor's claim that in shame "no steps suggest themselves" (as I will explore in the next chapter), she is quite correct to say that a reparative option is more obvious in the case of guilt, simply because it focuses on a particular action, as opposed to the self, which takes a more practically diffuse object: the status of the self.

Guilt and the Voluntary

The distinction I have drawn between guilt and shame relies on guilt being essentially tied to voluntary action; guilt is about actions one takes that violate norms. This is contrasted with shame, which is focused on the status of the self. In particular, one can feel ashamed because of something one has deliberately done, something one has non-intentionally or unknowingly done, or some fact about oneself totally unrelated to one's actions. There seem, however, to be two types of case where guilt appears to violate one of these distinctions. First we have the case of guilt for original sin, where it seems as if the subject experiences guilt for something someone else did, or for some non-action-based feature of oneself. Second, there is the example of "agent regret" where one feels guilt (or some emotion that is phenomenologically extremely close to it) for an action that is non-culpable because it has bad effects due to unforeseeably bad luck. Unless I am able to somehow account for these cases, they will stand to undermine the distinction I am aiming to draw between shame and guilt.

---

[156] See, for instance, Williams, *Shame and Necessity*.
[157] Taylor, 90.

Turning first to original sin, the doctrine seems to consist of two elements. Alistair McFayden sets out these two elements as follows:

> [Third party material excised][158]

As McFayden goes on to note, guilt for original sin, on the Augustinian account, can in no way be said to result from any action we take ourselves: [Third party material excised].[159]

If people feel guilty over original sin, then, this poses at least a prima facie problem for my characterisation of guilt, as it would appear to be either a case of vicarious guilt, felt over the sin of Adam, or a case of feeling guilty because of some fact about oneself (most likely one's relationship to Adam, who brought sin into the world). Given that I argued that the key dividing line between guilt and shame lies in the former's focus on personal actions, guilt over original sin seems difficult to accommodate in this schema.

One option would be to argue that while we do all inherit the guilt of original sin in the sense of being personally responsible for it, we do not, and should not, feel guilt for this. This line is at least suggested, if not wholeheartedly taken, by the theologian Reinhold Niebuhr, who claims that:

> [Third party material excised][160]

This line seems unappealing for my purposes, however, for two reasons. First, far from denying the possibility of feeling guilt for original sin, Niebuhr's position takes as read that such guilt *can* be felt (as it was by early Protestants and evangelicals), merely arguing that it need not be felt; this does not therefore address my worry about whether such guilt can be accommodated by my picture. Second, it is hard to see how, while maintaining that we are each (state) guilty for original sin we cannot be (emotion) guilty for it: the tie between state and (potentially) emotion is just too tight in our everyday folk psychological concept of guilt that to break it would do a level of violence to our concept of guilt that I am not prepared to countenance. Emotion concepts are, first and foremost, folk psychological concepts, so any account of them that strays so far from our core understanding is unlikely to be attractive. If we are to hold on to the picture of guilt that I am proposing, then, we will have to either argue that the doctrine of original sin is best conceived of so that our responsibility for original sin is

---

[158] Alistair McFayden, *Bound to Sin: Abuse, Holocaust and the Christian Doctrine of Sin* (Cambridge [UK]: Cambridge University Press, 2000), 16.

[159] McFayden, 16.

[160] H.Richard Niebuhr, "Man the Sinner", *The Journal of Religion* 15 (1935), 272-80 (278).

direct enough to make feeling guilt about it intelligible, or deny that we can appropriately feel guilty for original sin and find some way of explaining away supposed cases of this emotional phenomenon.

It is worth noting that the reality or otherwise of original sin has is not the central issue here. The key task for me in defending my view of guilt is to explain how, given that some people believe in the doctrine of original sin, we can accommodate the claims of some of them to feel guilt for original sin within the picture of guilt I am offering.

The option of positing a notion of original sin that does imply direct responsibility (and so renders feeling guilt in response fairly unproblematic) has been a popular one amongst Christian thinkers. As McFayden points out, the traditional doctrine of original sin is counter to the central moral tenet that we can only be morally responsible for our own acts: [Third party material excised][161]. Given the difficulty of squaring such a picture with our ordinary moral outlook, the drive to find a way of "taming" original sin is certainly understandable.

One way of doing this might be to regard the traditional notion of original sin as a metaphor, expressing only the inevitable tendency to sin that is present in human beings. On this account, the fall of Adam would represent the catastrophic shortcomings of our nature that make inevitable our committing of sins through our personal acts. Alternatively, we could argue that we are partially responsible for our sinful nature in a direct way, because the committing of sinful acts builds and reinforces our sinful nature. In both of these cases, we would only be morally responsible for our own sinful acts, but these would be against a backdrop of generalised human sinfulness, rendered vivid by the story of the fall[162].

In adopting either of these interpretations, we would be removing the sense of direct responsibility for original sin, reducing it to responsibility for our individual sins, or our sinfulness insofar as it is caused or sustained by our sinful acts. On this interpretation, the pressure to account for how people can feel guilt for original sin is removed – the answer that they cannot and do not is sustainable. We can supplement this position, however, by detailing the range of emotions that people could feel in this area. First, they could feel guilt either for their personal sinful acts, or for creating/sustaining their sinful character through their own sinful actions – there is a metaphorical or causal link to original sin here, but original sin is not the object of the emotion. Second, they could be feeling shame, as opposed to guilt, for

---

[161] McFayden, 22.
[162] These and other strategies along similar interpretative lines are set out in McFayden, *op. cit.*

original sin, either in a vicarious way through feeling shame for what Adam, as a fellow human being, did, or more directly, through feeling ashamed for their sinful nature (whether self-caused or not). Such a scenario would fit with the distinction between guilt and shame that I am proposing. Finally, they could be feeling collective guilt for the sins that we human beings have committed, including the original sin committed in Eden. I will come on to the experience of collective guilt in the next chapter, but suffice to say here that I do not believe it requires guilt to be felt by an individual – again, original sin is the object, but there is no feeling of individual guilt about it.

If we are able to re-frame the phenomenon of collective guilt in any of these ways which removes the attribution of personal responsibility for the act of another person, then we remove the problem of explaining how individuals can feel guilt for original sin. While this might be an attractive solution in itself, however, it does not seem sustainable to say that this is really what is going on with all cases where people feel emotions related to original sin. Not least among the problems with positing this view of original sin is that it seems contrary to the official doctrine of the Catholic church, as expressed in the Catechism:

> [Third party material excised][163]

The type of tampering that the Catechism warns against would include the attempted means of conceptualising original sin so as to remove personal responsibility for it. In order to maintain the universal need for salvation through Christ, the church must maintain that we are personally responsible for original sin, even in advance of any sinful acts we make.

For those who subscribe to the official Catholic view of original sin, then, individuals are personally responsible for original sin, and so feeling guilty for original sin should be both possible and appropriate. How are we to square this with my notion of guilt as an emotion we feel in response to our own, personal transgressions? It seems clear that one side or the other will have to give – either I will have to argue that the Catholic notion of original sin provides an inappropriate focus for guilt, or revise my notion of what can constitute an appropriate case of guilt.

My preferred option would, predictably, be the former. I agree with Paul Ricoeur's contention that, the traditional notion of Original Sin [Third party material excised][164]. While it is

---

[163] *Catechism of the Catholic Church 389,* www.vatican.va.
[164] Paul Ricoeur, "'Original Sin': A Study in Meaning" in *The Conflict of Interpretations Essays in Hermeneutics* 3rd edn (London: Continuum, 2004) , 266.

undoubtedly an extreme step to dismiss a concept as incoherent, we can in this case point to the intellectual pressures that may be driving people to operate with such a concept. Augustine, and those modern Catholics who follow him on the meaning of original sin, seem caught between their belief that everything is the creation of a benevolent God (as opposed to the Manichean doctrine that he is not responsible for the evil in the world) on the one hand, and their belief that all human beings need to be saved through baptism, regardless of whether they have wilfully committed any sinful acts themselves, on the other. As a result they find themselves pressured to adopt a notion of original sin whereby the sinful nature of all people is as a result of a person's action (and so God is not responsible), but other people are each directly responsible for that sinfulness through their biological inheritance.

In the absence of an antecedent belief in any of these propositions, we have no reason to accept the notion of original sin as coherent, let alone one that refers to a real phenomenon. Once we have adopted the position that the notion of original sin is incoherent, we can argue that any episode of guilt that has such a notion at its centre must be either based on a paradigm scenario, or an interpretation of a situation as analogous to an acceptable paradigm scenario, that is radically inappropriate. While we could therefore argue that this makes guilt for original sin inappropriate, I would go further and argue that it is so off-kilter that it is hard to conceptualise it as a case of guilt at all.

Thus I would argue that, unless we are independently committed to premises that lead us to adopt the "juridical-biological" notion of original sin, we should resist any pressure to revise our notion of guilt on the basis of its supposed inability to accommodate cases of guilt over original sin. Instead, we should argue that original sin is either (a) to be conceived of in such a way that it is not an intelligible object of individual guilt or (b) to be dismissed as an incoherent concept.

A second type of case that could threaten my account of guilt is that of "agent regret". Like original sin, this appears to be a case where we have an emotion like guilt felt in the absence of deliberate transgression. Unlike original sin, however, agent regret focuses on a particular action or omission of the subject; the distinction is that while the subject is causally responsible for this "act", they are not morally responsible for it (on our usual interpretation of moral responsibility) as they did not freely choose to do it as an act of will. If agent regret is to be understood as a variety of guilt, then if it is a real phenomenon it may cast doubt on

my notion of guilt as being based around transgressions deliberately committed by individuals.

In his classic formulation of moral luck, Williams uses the example of a lorry driver who hits a child who has run out into the road. For illustrative purposes, I would like to distinguish four different versions of this story. In the first version, the driver is *strongly culpable and unlucky*, say because he is drunk behind the wheel, a child runs out in front of him, and he kills the child. In the second version, the driver is *weakly culpable and unlucky*; he has failed to check his brakes as recently as he should have, a child runs out in front of him, and he kills the child. The third driver is *weakly culpable and lucky*; he has failed to check his brakes as recently as he should have, but no child runs out in front of him and no-one gets hurt. The fourth driver is *non-culpable and unlucky*; he has done everything right, neglected nothing, but a child runs out in front of him, and he kills the child.

On my notion of guilt as a reaction to one's transgression of a moral norm, the level of guilt-feeling that would be appropriate in each of the four drivers seems clear: the first driver should feel a significant amount of guilt, the second and third should feel a lesser amount than the first (because their omissions were less risky) but equal to each other (because they transgressed by the same omission), and the fourth driver should feel no guilt (because he did not transgress by any deliberate action). Yet this does not seem plausible; we would expect the second driver to feel worse about his actions than the third, and the fourth driver to feel somewhat bad about his actions too. Rather than having their negative self-directed feelings limited strictly to the limits of their voluntary action, we would expect the results of their actions to influence these reactions. Moreover, such reactions would seem appropriate. Williams argues that [Third party material excised][165]. This is surely correct; someone who could shake off running over and killing a child (even when completely non-culpable) as easily as hearing that a child has been run over by someone else would be a strange and concerning individual.

If this self-directed negative feeling were to be conceptualised as a case of guilt, then this would pose at lease a prima facie problem for my notion of guilt. On my understanding, guilt is based around voluntary action that transgresses moral norms, but here we would be claiming that it is commonplace and, moreover, appropriate, for people's feelings of guilt to be based on factors outside their own control (in this case, whether or not a child happens to

---

[165] Bernard Williams, "Moral Luck", *Proceedings of the Aristotelian Society* Supp. 50 (1976), 115-35 (125).

run out into their path). Given that I do not wish to deny the appropriateness of such feelings, I must either account for them within my current conception of guilt, or revise that conception to accommodate them.

One tack would be to follow Williams in arguing that where there is negative self-directed feeling in these cases above and beyond guilt over one's transgressions, this is not additional, 'outcome-dependent' guilt, but rather a different emotion, namely "agent regret". Williams describes agent regret as:

> [Third party material excised][166]

We are therefore left with a typology of three emotions: guilt, which is felt over voluntary transgressions we have committed, regret, which is the reaction to situations that we wish could have turned out differently, and agent regret, which is felt in response to situations where we think we could have acted differently so as to result in the situation turning out differently.

This typology opens the door to what Wolf calls a "rationalist" position:

> [Third party material excised][167]

Wolf's nameless virtue is one that involves taking responsibility for the effects of our actions, even when they are unforeseen and unforeseeable (one might ask whether "magnanimity" could be an appropriate name for it). If we were to adopt this position, my notion of guilt would be completely untouched by the phenomenon of agent regret. Where one has committed a transgression against a moral norm, even when this has no bad effects (like the culpable but lucky driver) guilt is in order. Where there is no transgression but our action has bad effects (like the non-culpable but unlucky driver), guilt is not in order, but agent regret is. Finally, where there is a transgression but, through bad luck, this contributes to much worse consequences than could have reasonably been foreseen (like the weakly culpable but unlucky driver) then guilt is in order for the transgression, but only agent regret is in order for that portion of the bad consequences that could not have reasonably been foreseen to follow from the transgression.

---

[166] Williams, *Moral Luck*, 117
[167] Susan Wolf, "The Moral of Moral Luck" in *Setting the Moral Compass: Essays by Women Philosophers* ed. by Cheshire Calhoun (Oxford: Oxford University Press, 2004), 113-127 (124).

As Wolf correctly argues, however, this typology seems much too neat. Instead, she suggests a second, more nuanced position. On this view, like the rationalist perspective just discussed, in cases of non-culpability, agent regret can be an appropriate response, and in cases of culpable wrongdoing that, luckily, had no bad effects, guilt can be appropriate. Where the second view differs, however, is in its claim that [Third party material excised].[168] In other words, in such cases we cannot disentangle guilt and agent regret; the bad effects of the action suffuse the guilty acknowledgment of the wrong we have committed, deepening and colouring that emotional state. This is surely a more compelling picture – the notion that someone could plausibly feel guilt for a particular wrongdoing, but only agent regret for bad results coming from unforeseeable bad luck, all in relation to a single incident, is fanciful.

With this typology of emotional responses in hand, we can give an account of how my proposed notion of guilt fares against the examples in question. In the case of the non-culpable but unlucky driver, we would say that agent regret is appropriate, but guilt is not. That someone might feel guilt in this situation is, however, intelligible. First, in such cases we might expect people to be questioning whether they had, in fact, transgressed a moral norm in their prior behaviour; the fact that they are experiencing agent regret will mean that their attention is focused on the events leading up to that, and the suspicion of possible culpability may be enough to lead them to see the situation as analogous to the paradigm scenario for guilt. Second, even if the agent is of the firm conclusion that she did nothing wrong, the situation is close enough to one where a moral norm has been transgressed (the agent has caused harm, and feels bad for it) for us to understand how she might view the situation in those terms anyway, and so feel guilt as a recalcitrant emotion. It is an advance of the paradigm scenario view over the traditional cognitivist position that it can easily make sense of such a position. In fact, I would suggest that in situations like that of the non-culpable but unlucky driver, we might expect people to oscillate between (appropriate) agent regret and (inappropriate) guilt reactions.

Turning to the contrast between the levels of guilt between the lucky and unlucky versions of the weakly culpable driver, it seems clear that, for the reasons stated above, we can see how more intense guilt-reactions in the latter case are intelligible (and, indeed, to be expected). I would go further than this, however, and say that this disparity can be appropriate. Given that agent regret can be an appropriate reaction to harm caused by our actions, and that in cases of

---

[168] Wolf, 125.

limited culpability this agent regret can cause a blended emotional reaction with guilt, I would argue that it can be appropriate in such cases to feel a greater degree of guilt, a result that reflects what Margaret Urban Walker calls our "impure agency"[169]. In a sense, however, the question of appropriateness is less pressing for present purposes – it is enough to preserve my notion of guilt that I show how this common type of reaction is at least intelligible under it.

Thus I conclude that neither the phenomena of guilt for original sin nor agent regret renders my notion of guilt unsustainable. In the first case, I argued that the phenomenon must either be re-described so that it is not a case of guilt being felt for the transgression of some other person, or dismissed as unintelligible as a case of guilt. In the second case, I concluded that agent regret, as a distinct emotion from guilt, can be an appropriate response to non-culpably caused harm. Guilt, by contrast, can be intelligible, though inappropriate in such cases, and a greater degree of guilt may also be an appropriate response to an identical transgression that unluckily causes greater harm. Neither of these cases, therefore, present me with a reason to revise my characterisation of guilt.

Now that I am satisfied with the distinction I have drawn between shame and guilt, I can go on to consider their respective roles as moral emotions.

---

[169] Margaret Urban Walker, "Moral Luck and the Virtues of Impure Agency" in *Moral Luck* ed by Daniel Statman (Albany: State University of New York Press, 1993), 235-50 (243).

CHAPTER FIVE: SHAME AS A MORAL EMOTION

Having distinguished shame from other, related emotions, including guilt, I am now able to address the question I posed at the outset: is there a necessary role for shame in moral life? I have already argued that there is a necessary role for the emotions, in general; I now need to argue that shame ought to be numbered among the group of emotions that are necessary. I will approach this task by considering, and rejecting, the most prominent arguments available, from within a view that sees an important role for the emotions more generally in moral life, for denying shame such a role.

The type of argument that I am aiming to refute is based on the claim that guilt can perform the functions we require of a moral emotion better than either shame alone, or a combination of guilt and shame. In its strongest form, this argument comprises a two-pronged attack. First, there is the argument that appropriate guilt alone is able to pick out all and only cases of moral disvalue, whereas even appropriately-directed shame picks out cases where no moral disvalue is in evidence. Second, there is the argument that shame is a dangerous emotion, with bad effects, whereas guilt is comparatively benign. Taken together, these arguments paint shame as superfluous in moral life, and potentially dangerous; the obvious conclusion is then that we should, ideally, attempt to expunge shame from our moral lives, in favour of guilt. I will attempt to avoid this conclusion by arguing that both strands of the case against shame are unsustainable.

Shame as an Inferior Moral Emotion to Guilt (1): Shame is Superfluous

The first strand of this argument against shame claims that some other emotion (or combination of emotions) - most plausibly guilt - is sufficient to fulfil the entirety of the role we need the moral emotions to play in our moral lives, from a first-person view, at least. The particular form of this argument that I am going to consider is that put forward by Alan Gibbard in his book *Wise Choices, Apt Feelings*.

As I discussed in Chapter 1, Gibbard can be described as a neo-sentimentalist, in that he sees an essential role for the emotions in picking out instances of moral value. More specifically, his view is a norm-expressivist one, which says that the way to tell whether an action is wrong is to consider whether we accept norms saying that the doer of such deeds ought to feel guilt, and whether observers ought to feel anger, over the doing of them:

[Third party material excised][170]

It is important to note that guilt and anger exhaust the range of moral emotion that is required on Gibbard's picture; shame does not feature. Why does Gibbard view shame as surplus to requirements in this way?

The thrust of his argument is that, even before the imposition of societal-based norms – even before guilt is 'civilised' by societal pressure - that emotion is tied to the voluntary in a way that shame is not. Gibbard goes on to argue that this means [Third party material excised][171] As a result, he continues, [Third party material excised][172] Guilt is ready-made to be an excellent moral emotion, he suggests, as it attaches to all and only voluntary, morally disvalued acts.

Shame, on the other hand, is not so restricted:

[Third party material excised][173]

As my earlier exploration of shame affirms, one can clearly feel ashamed over one's parents, one's physical appearance, one's lack of natural athletic ability and other factors that are largely or wholly outside one's voluntary control. This means that, in its current state, shame is far less suited to be a moral emotion than guilt. When we feel appropriate guilt, in the central case we know (or at least believe) that the action on which the emotion is focused is morally wrong; we can have no such certainty with even appropriate shame, as it can be felt over either a moral failing or a non-moral shortcoming.

Of course, one option would be to 'moralise' shame by adopting norms that say that shame is only appropriate if it is being felt over some voluntary action or pattern of behaviour. Gibbard argues, however, that this is not a feasible strategy:

[Third party material excised][174]

Given that we already have guilt, the norms for which already sanction feeling it only over voluntary actions, it would seem totally pointless to even attempt such a difficult (if not impossible) task in the case of shame.

---

[170] Gibbard, 45.
[171] Gibbard, 297.
[172] Gibbard, 297.
[173] Gibbard, 297.
[174] Gibbard, 297.

Shame as an Inferior Moral Emotion to Guilt (2): Shame has Negative Effects

The second strand in the case against shame as a moral emotion suggests that shame is problematic as a moral emotion in ways that guilt is not. On its own, of course, this may not be enough to rule shame out of contention as a moral emotion; we might want to accept the supposed downsides of shame as a price worth paying for the deeper moral insight it affords us. If, however, we combine this with the first strand, then this makes it seem that shame is more trouble than it is worth. We would be accepting the downsides of shame for no corresponding upside that we could not get just as well (or, as some would claim, better) from guilt. The obvious conclusion from this is that we ought to take steps to eradicate shame from our moral lives, focusing solely on appropriately-felt guilt.

Turning to the second stand, we see arguments, often in the psychology literature, suggesting that shame can have strongly negative effects, both in terms of the individual's well-being and their ability to make good moral judgements. In the former category, we may place claims that shame can lead to depression and other-focused anger; of course, a secondary effect of these conditions could be to negatively affect one's capacity to make good moral judgements, so perhaps we would do better to view them as having a foot in both categories. In the latter category, we can locate claims that shame reduces one's ability to empathise with others, so reducing one's ability to have appropriate moral reactions. Guilt, by contrast, is said to carry none of these costs. I will set out each of these supposed costs of experiencing shame, as opposed to guilt, in turn.

Shame and Depression

Consider first the claim that shame can lead to depression. Perhaps the most famous instance of this argument is in Helen Block Lewis's *Shame and Guilt in Neurosis*. Lewis writes that [Third party material excised][175]

This claim is taken up, and apparently confirmed, by Tangney and her collaborators.[176] Using subjects' reports of how they would feel in different hypothetical situations, they divide the subjects into guilt-prone and shame-prone categories. Then, using more interviews, they assess the prevalence of certain psychological symptoms. They found that:

---

[175] Lewis, 89.
[176] June Price Tangney, Patricia Wagner and Richard Gramzow, "Proneness to Shame, Proneness to Guilt, and Psychopathology", *Journal of Abnormal Psychology* 101 (1992), 469-78.

[Third party material excised][177]

These 'symptom clusters' included symptoms typical of people suffering from depression.

By contrast, the authors claim that proneness to guilt is not associated with psychological disorders in the same way:

[Third party material excised][178]

In other words, while guilt-prone people were found to be somewhat more susceptible to psychopathological syndromes (notably depression) than the average person, this could be attributed solely to the fact that they were also more prone to shame than the average person. Once they discounted the effects of the propensity to experience shame, the propensity to experience guilt was found to have no connection to a heightened prevalence of these symptoms.

While the authors recognise that their study, being correlational, cannot prove a causal link between shame-proneness and propensity to suffer these psychological symptoms, they do offer some speculations as to why we may expect such a causal connection to exist.

One such explanation is that [Third party material excised][179] Elsewhere in the literature, we are told, it has been suggested that stable and global attributions of negative characteristics to the self have been identified as key causes of hopelessness, and the depression that hopelessness can induce. Tangney et. al. argue that as such self-attributions are typical of shame, it is to be expected that a tendency to experience shame should be linked to increased incidence of depression.

The authors also argue that this linkage between shame and depression is liable to be self-strengthening, through a type of mental feedback loop. On this account, feelings of shame cause the subject to feel depressed, with the subject becoming ashamed of his depression ('What have *I*, with my house, job, loving family, got to be depressed about') leading to a strengthening of the existing feelings of depression and so on.

Finally, they argue that [Third party material excised][180] The suggestion here is that certain 'family processes' (factors in the style of family interaction experienced in a subject's early life) are likely to result in subjects who are prone both to experiencing shame, and to suffering depression.

---

[177] Tangney, Wagner and Gramzow, 475.
[178] Tangney, Wagner and Gramzow, 472.
[179] Tangney, Wagner and Gramzow, 476.
[180] Tangney, Wagner and Gramzow, 476.

If shame and depression are inextricably linked in this way, this would count against adopting shame as a moral emotion in two separate ways. First, and most obviously, depression is damaging to the person who is suffering from it. At the mild end it can prevent us from taking and enjoying valuable opportunities that are presented to us, while in the most serious cases it can result in significant damage to one's well-being, and even suicide. As a result, a linkage between shame and depression would mean there were serious costs associated with feeling shame, which would need to be weighed against any efficacy as a moral emotion it is thought to possess.

Second, a link between shame and depression would actively serve to undermine shame's role as a moral emotion. One of the reasons I invoked in Chapter 1 to argue that we needed emotions in our moral lives was to solve the motivation problem: it is painful to feel negative moral emotions, so we are therefore led to take steps to avoid them, or to alleviate them if we fail to do so. In depression, however, one characteristically loses motivation to act, seeing any attempt at positive action as pointless. This difficulty is noted by Isenberg:

> [Third party material excised][181]

If it is true that shame is necessarily linked to depression, we would still have a motivation to avoid doing acts that would provoke it, but when we did do acts that led to shame, and this resulted in depression, this would strongly counteract any motivation we would have to put things right. Thus shame would partially undermine its own ability to do the work that we need a moral emotion to do.

Shame and anger

Moving now to the case of anger, Mollon gives an account of one case which is indicative of the general view connecting shame with violent anger:

> [Third party material excised][182]

The usual explanations given in the literature[183] for reactions to shame of this sort, focus on the idea that shame is an intensely unpleasant emotion, and so this can sometimes lead to a desire to externalise the source of the emotion by blaming somebody else. By transmuting shame into anger,

---

[181] Arnold Isenberg, "Natural Pride and Natural Shame" in *Philosophy and Phenomenological Research* 10 (1949), 1-24 (19).
[182] Phil Mollon, *Shame and Jealousy: The Hidden Turmoils* (London: Karnak, 2002), 3.
[183] See, for instance, Tangney and Dearing, *op. cit.*

and attending to supposed failings of others, rather than his or her own, the subject can avoid (some of) the painful experience associated with the former emotion.

Guilt, on the other hand, may be thought not to suffer from similar links to anger. With the focus of guilt being on one's actions, rather than one's self, it is seen as being a less damaging emotion to one's overall self-regard than shame, and so not motivating the externalisation reaction to the same degree. Also, it is argued, by being focused on a transgression, for which one might make amends, guilt presents an obvious route for its own relief. Shame, on the other hand, being about the status of oneself, does not offer such simple answers; as a result, externalisation might be the only way (at least in the short term) to relieve oneself of its painful feeling.

Tangney and her collaborators, once again, claim that empirical evidence supports the hypothesis of a relationship between shame and anger, and the non-existence of a similar relationship between guilt and anger. Using the same techniques as in the study into guilt, shame and depression, they found that:

> [Third party material excised][184]

While, obviously, both guilt and shame-prone individuals were liable to experience anger in some situations, the former were found to handle this in a more constructive manner than the latter. Whereas the shame-prone often resorted to active aggression towards another person, the guilt-prone tended to adopt other strategies, such as non-aggressive discussion, in interacting with the target of their anger.[185]

Once again, it is clear that this kind of link between shame and the most damaging forms of anger, in the absence of a similar link between guilt and anger, would count against any claim we wish to make for shame as a moral emotion. Anger can have negative effects both on others, if it causes the subject to lash out, and on the subject herself, ranging from damage to relationships to possible legal repercussions.

A link to anger would also undermine shame's ability to perform the functions required of a moral emotion. On the motivational side, it would create a host of alternative desires, most notably for revenge, which would weigh against any positive desires to put things right arising from the shame reaction. In addition, if our shame reaction is transmuted into one of anger, then we will see the painful feelings as brought on by the actions of others, rather than ourselves. As a result, the task of

---

[184] Tangney and Dearing, 97.
[185] Tangney and Dearing, 100.

shaping the self-monitoring mechanisms of the subject, so he is able to avoid similar situations in future, will not be achieved, as the situation has been understood as one that is not due to the subject's behaviour or characteristics.

As I have argued, we also need moral emotions to play an epistemological role; as well as marking some piece of behaviour as being of moral concern, they focus our attention on this, leading us to contemplate why and in what way it is of moral concern, so allowing us to overcome the attentive and conceptual frame problems in particular instances. This in turn allows us to come to a deeper understanding of different morally relevant factors, shaping our future moral responses. If shame is transmuted into anger, however, we lose this epistemological aspect of the emotion; the anger takes away our attention from the morally salient factors, and instead focuses attention on its own object.

Shame and Withdrawal

A further negative reaction that is often associated with shame is withdrawal from others. Williams notes this tendency:

> [Third party material excised][186]

This possible reaction is also noted by Tangney and Dearing, who claim that [Third party material excised].[187]

Again, we can understand why this might be the case. Shame is an unpleasant emotion, so the urge to escape it is strong. If the shame is purely heteronomous, then one way of doing this is to escape from the judgements of others by withdrawing from view. This is not such an easy option when the shame is autonomous – we cannot withdraw from our own gaze (a point to which I will return in due course).[188] However, it is true that the negative judgements of others can strengthen even autonomous shame, by continually drawing the subject's attention back to the putative failings that are at the centre of the emotional reaction. Even in these cases, then, it may be possible to somewhat relieve feelings of shame by withdrawing from others.

Guilt, on the other hand, is supposed not to encourage such behaviour, for two reasons. First, as it is focused on a particular action one has committed, it is unlikely to be eased by escaping from public

---

[186] Williams, *Shame and* Necessity, 89.
[187] Tangney and Dearing, 92.
[188] This is true physically at least. One might attempt to withdraw from one's own gaze by distracting activity, e.g. alcohol or drug abuse.

view. The central judgement in guilt is the subject's own judgement that he or she has transgressed some norm, so withdrawing from the sight of others will not remove the basis for the emotion. Second, because of guilt's focus on an identifiable transgression, there tend to be more easily identifiable options for easing the guilt feeling, such as apologising or trying to right the wrong one has done. In shame, it is claimed, there is often no such easy option for putting things right. Thus in guilt the negative reactions identified as a means of easing shame feelings are not required, due to this alternative outlet.

The damage that a tendency for shame to lead to withdrawal from others would do to shame's prospects as a moral emotion is obvious. In leading us to turn away from others, shame makes it more difficult for us to focus on the morally salient characteristics of a situation and to be motivated to make amends to those we have wronged. If shame has this effect, then, but guilt does not, so much the worse for shame as a moral emotion – or so one might argue.

Shame and Empathy

Finally, a charge that is sometimes made against shame is that it reduces one's ability to empathise with others. Anecdotal evidence such as the case of Jim, reported by Tangney and Dearing[189], seem to support this. Asked to describe an occasion when he felt shame, Jim recalls an incident where some work colleagues of his were telling jokes, which swiftly began to take a racist tone. Despite feeling uncomfortable about the jokes, Jim did nothing to express his distaste, and in fact came to somewhat enjoy them. On noticing, however, that a black colleague, whom he liked, was sat behind him and must have heard everything they were saying, Jim immediately felt ashamed.

Tangney and Dearing remark on the type of language that Jim uses in recounting this experience, calling his reaction "solidly self-focused". By contrast, they argue, Jim does not mention the effects of his actions on his friend, completely ignoring the way she must have been feeling. This, they suggest, is symptomatic of the way shame turns the subject's attention inwards, so blocking identification with others; [Third party material excised][190]

Tangney and Dearing argue that the negative impact of shame on empathic responses is a result of the general tendency of shame to draw the subject's attention inward:

---

[189] Tangney and Dearing, 83-84.
[190] Tangney and Dearing, 84.

[Third party material excised][191]

Guilt, by contrast, is not thought by them to have such negative effects on empathy; on the contrary, they argue that guilt can actually enhance one's ability to have empathic reactions. They give three reasons for the asymmetric relationships of guilt and shame to empathy.

First, guilt focuses on a specific wrong act that one has done, and so is more likely to focus one's attention on the harm that one has done to others, so increasing the likelihood of an empathic response. Second, guilt is said to carry with it a sense of personal responsibility, meaning that it is more difficult for people to externalise the blame to which their own feelings of guilt are related, which could serve to cut off any incipient empathic reaction. Finally, phenomenal studies have consistently linked guilt with reparative actions, which may themselves be motivated, in part, by empathic reactions.[192]

Once again, Tangney (this time with George Mason) attempts to supplement the theoretical argument that shame is more prone to lead to this type of bad outcome than guilt with empirical results. She writes that:

[Third party material excised][193]

The studies in question are organised along similar lines to those already cited, with subjects being divided into shame- and guilt-prone groups on the basis of qualitative assessments, and then assessed for empathic responsiveness in different scenarios, on the basis of how they respond to questions about their actions in different circumstances.

Clearly, if we were to accept that shame in general inhibits empathic responses, but guilt has no such effects, this would be damaging for an argument that shame ought to be regarded as a moral emotion alongside guilt. There are two main ways in which this damage would be done.

First, far from being a commendable moral response, feeling shame rather than guilt would be indicative of a vicious, or at least significantly less than virtuous, moral disposition. There is something hideous about the perpetrator of a serious wrong constantly considering only what his behaviour means about him as a person. John Doris (whose situationist critique of shame I shall consider in the next chapter) notes this point:

---

[191] Tangney and Dearing, 83.
[192] Tangney and Dearing, 86.
[193] June Price Tangney and George Mason, "Moral Affect", *Journal of Personality and Social Psychology* 61 (1991), 598-607 (604).

[Third party material excised][194]

While it is true that a murderer who feels genuinely ashamed of his actions displays a better moral character than one who feels no negative emotion at all in response to his behaviour, we may be justified in thinking that he should be feeling less concerned with what his actions mean for *him* and more concerned with how they have affected his victim, and his victim's loved ones. In short, we may be justified in saying that he should be feeling less shame and more guilt. Similarly for Jim in the earlier example; he ought to be less worried about how his actions reflect on him, and more about how they have hurt his friend.

Second, if shame causes an undue focus of attention on oneself, as against others, this would directly prevent shame from performing one of the vital functions of a moral emotion. As I argued in Chapter 1, one of the key roles of the moral emotions is to focus our attention where it ought to be focused, so helping us to address the attentive and conceptual frame problems of moral judgement. If shame necessarily focuses it in the *wrong place*, then shame would, by its very nature, undermine its own status as a moral emotion.

We now have a number of allegations relating to negative tendencies associated with shame, which are said not to be associated with guilt; in the words of Tangney and Dearing, shame is [Third party material excised][195]. Rather than view these as distinct criticisms of shame that all, independently, make a case against shame and in favour of guilt as a moral emotion, we do better to conceptualise these as aspects of a single case arguing for the potential dangerousness of shame, and the relative safety of guilt.

On this picture, shame is particularly prone to both induce and facilitate displacement activity, where the blame for the negative feeling is externalised by the subject. This leads to the possibility of damaging responses such as outward-focused anger and withdrawal from the society of others, both of which make constructive, remedial action less likely. With its focus on a degraded self, as opposed to particular bad actions of the self, and thus offering no easy answers on how to put things right, shame can lead to feelings of hopelessness and depression. As well as being bad in themselves, both for the subject and for others, these angry, withdrawn and depressive reactions take the attention of the shamed person away from where it should be – on the person who is

---

[194] John M. Doris, *Lack of Character: Personality and Moral Behaviour* (Cambridge: Cambridge University Press, 2002), 156.
[195] Tangney and Dearing, 3.

suffering on account of his behaviour. This means that empathic reactions are dulled, so adding a negative moral effect to these damaging tendencies of shame.

Guilt, by contrast, is said to suffer from none of these failings. By focusing on the restricted case of a particular bad action, or set of actions, guilt does not provide the same kind of motivation and opportunity for externalisation of blame, and provides more obvious opportunities for direct relief through apology and reparation; it therefore does not pose the same kind of risks in terms of anger, withdrawal and depression. At the same time, it focuses the attention of the subject on the morally salient aspects of the situation – the person who has been harmed or wronged – and so performs its role as a moral emotion better than shame.

We are therefore presented with a powerful second strand in the case against shame as a moral emotion *vis-à-vis* guilt: unlike guilt, shame is prone to induce a whole host of negative reactions that are bad in themselves and/or undermining of shame's own role as a moral emotion.

Added to the first strand, this completes the case against shame, vis-à-vis guilt, as a moral emotion. If we are persuaded by this case, then we will see shame as an emotion that brings with it scope for numerous kinds of negative reaction, with potentially dangerous consequences for the subject and for others, and which undermines the moral role that it is supposed to be playing; guilt, on the other hand, is believed to have none of these flaws. We will also accept that shame is superfluous as a moral emotion; guilt can capture all the cases of moral disvalue, whereas shame ranges beyond these boundaries, being felt (appropriately) in cases of purely non-moral disvalue.

Where we have something that is (i) surplus to requirements and (ii) potentially dangerous and undermining of our important purposes there can be only one prescription: get rid of it, to the extent that we are able. In the case under consideration, this means that we ought to expel shame from our moral reasoning, leaving the field of first-personal moral emotions open to guilt alone.[196]

In Defence of Shame (1): Shame is a Necessary Supplement to Guilt

I have set out what I believe to be the strongest case for excluding shame as a moral emotion and proceeding solely with guilt in this role. I will now explain how I think this case can be resisted, and why we ought to grant shame its rightful place as an important moral emotion alongside guilt. I will address the two strands of the case against shame in order. First, I argue that guilt alone is not

---

[196] I insert the modifier "first-personal" in recognition of Gibbard's claim that anger is a third-personal moral emotion, which should stand alongside guilt. As Gibbard sees the two as perfect analogues however (meaning that appropriate anger will be felt at all and only those actions of others for which it would be appropriate for the agent to feel guilt) I will ignore this nicety from now on.

sufficient to capture all cases of moral disvalue, and shame is well-placed to fill some of these gaps. Second, I put the case that shame is not as harmful as it is made out to be (and perhaps guilt not as harmless), so by admitting shame to the pantheon of moral emotions we do not have as much to fear as critics of the emotion would have us believe.

Turning first to Gibbard's argument for the superfluity of shame as a moral emotion, one way to defend against this would be to suggest that guilt is just as bad, arguing that guilt (or some emotion very like guilt) is not confined to the voluntary, but sometimes ranges over the non-voluntary as well. This type of case is put by Bernard Williams in *Shame and Necessity*. If a sustainable argument can be produced to this effect, this could undercut the supposed distinction Gibbard draws between shame and guilt, and put the two emotions back on all fours.

Against this, however, we could expect Gibbard to argue that we ought to (and, in fact, that we do) adopt norms saying that guilt is only appropriate when it is felt over the voluntary. Williams, of course, could reply that it would be both difficult and arbitrary to restrict the appropriateness of guilt in such a way; why should we accept that such supposedly natural and intelligible cases of guilt are in fact 'irrational' or 'inappropriate'?

I am unsure that there is any killer move that can be made in this argument to make it implausible to come down on either side. Furthermore, I do not need to show that guilt can appropriately be felt over non-voluntary actions to make my case defending shame from Gibbard's charge of superfluity. For these reasons, and without making any claim about which side is in the right, I will put this argument to one side.

Instead, my counterargument against Gibbard will focus on the argument that guilt alone is not enough. There are some cases where the following three conditions hold: (i) it is not appropriate to feel guilt over some behaviour, (ii) the behaviour is morally disvalued and (iii) it *is* appropriate to feel shame over that behaviour. If there are such cases, this will indicate that guilt needs to be supplemented by shame as a moral emotion in order to capture them.

One way to achieve this would be to argue that Gibbard has made a fundamental error in attempting to restrict the sphere of moral value to actions, whereas in reality it also covers aspects of motivation and character. As guilt is an emotion related to actions, with shame being related to character traits, this would immediately suggest a distinctive role for shame in picking out certain varieties of moral value. While I am personally attracted to such an approach, I am aware that a significant proportion of potential readers would be likely to have qualms about it, given the deep

waters of questions of the extent to which we can be held morally responsible for the content of our characters. Instead, I propose to adopt a more consensus-driven approach, by picking out two areas where I would argue moral value goes beyond personal actions that transgress norms, and so beyond the purview of guilt, while still remaining close to the general sphere of action, and so the more consensually-accepted space for moral assessment.

Shame and the Supererogatory

One area where shame seems capable of marking moral value, where guilt cannot, is the supererogatory. The basic idea behind characterising an act as supererogatory is that the act somehow goes beyond the call of duty, so the agent does more than is required. More precisely, we may follow Gregory Mellema[197] in defining a supererogatory act as one which fulfils the following three conditions:

1. Performance of the act directly fulfils no moral obligation or duty.

2. Performance of the act is morally meritorious or praiseworthy.

3. Non-performance of the act is not morally demeritorious or blameworthy.

This definition captures the commonsense notion that supererogation is about going beyond the call of duty, and so doing something that is deserving of praise, but for failing to do which one would not be deserving of blame. As such, if supererogatory actions exist, they clearly fall within the realm of moral value. This presents the neo-sentimentalist with a question: if we accept that supererogatory actions exist, what emotion is able to pick out the type of moral value they instantiate?

Several philosophers have suggested that shame, rather than guilt, is the emotion that is appropriate to picking out the type of moral value that we find in the realm of the supererogatory. For instance, Rawls writes that:

> [Third party material excised][198]

Rawls does not expand here on what he means by "the higher forms of moral excellence", but it is clear that he views shame as a more appropriate emotion to guilt in the realm of the supererogatory.

---

[197] Gregory Mellema, *Beyond the Call of Duty: Supererogation, Obligation and Offence* (Albany: State University of New York Press, 1991), 3.
[198] Rawls, 424.

Our intuitions about particular cases, I would argue, tend to cohere with Rawls' contention. One example of this can be seen in the way we can make sense of the speech that Shakespeare gives Henry V on the battlefield at Agincourt:

> And gentlemen in England now-a-bed
>
> Shall think themselves accurs'd they were not here,
>
> And hold their manhoods cheap whiles any speaks
>
> That fought with us upon Saint Crispin's day.[199]

Henry is talking about the reactions of men who failed to *volunteer* in his army. The language he uses in foreseeing these reactions – they will "think themselves accurs'd" and "hold their manhoods cheap" – is accordingly much more fitting with shame, the emotion that is concerned with the status of oneself, rather than guilt, which is concerned with the status of one's actions.

To take another example, consider a person who passes a burning building, with a crowd of people standing outside. One of the onlookers tells her that there is a person trapped inside. With the fire brigade nowhere in sight she decides that she should run into the building to try to rescue whoever is trapped inside. Looking at the raging flames, however, she has a loss of nerve and cannot bring herself to enter the building. None of the other onlookers attempt a rescue either, and the person inside is killed. This seems like a textbook example of failure to perform a supererogatory action; if you are not convinced, however, you are welcome to substitute an example of your own and use that for the following analysis.

Now imagine that the woman who did not attempt the rescue is talking about the event just described, some weeks later. If the woman said that she felt guilty about her behaviour, I think we would rightly feel that her emotion was inappropriate. We might tell her nobody could have expected her to put her own life at risk, and that nobody can blame her; in short, that she has nothing to feel guilty about. If, however, she said that she felt ashamed of failing to attempt a rescue, our response might be quite different. It is true that we might try to put her feelings of shame in perspective, by pointing out the great danger that would have been involved in any attempted rescue, along with the fact that none of the other onlookers tried to enter the building. I do not think, however, that we would find her shame to be fundamentally misplaced in this situation, in the way that we would find guilt to be.

---

[199] William Shakespeare, "The Life of Henry the Fifth" in *Complete Works* ed. by J. Bate and E. Rasmussen (Basingstoke: Macmillan, 2007), 1033-1097, iv. 3, 1077.

One of the characteristic thoughts of the person who is experiencing guilt is "What have I done?" If the woman in our example expresses this thought, however, we would clearly be justified in saying that she has done nothing wrong, and that she is wrong to feel guilty as, by hypothesis, one does not do wrong in failing to perform a supererogatory action. One of the characteristic thoughts of shame, on the other hand, is "What have I become?" If the woman in our example were to express *this* thought, we might not be so quick to dismiss her concerns. While she has clearly not done anything wrong, it is less clear that she has not failed to live up to intelligible standards of how she ought to be. If she had previously considered herself as someone who would go to great lengths to help others in distress, even at serious risk to her own wellbeing, her failure to attempt a rescue may be enough to provoke shame in her.

One may ask, however, how a failure to perform a supererogatory action can make sense as an object of shame when, by hypothesis, such a failure is not morally demeritorious. An initially tempting reply may be that shame, in contrast to guilt, can range appropriately across the moral and the non-moral. While this is true, however, it would not do to categorise such cases as examples of shame about non-moral issues, as this would defeat the argument that supererogatory actions foregone can represent cases where shame can detect a source of moral value that guilt cannot.

Instead, I would argue that in some cases one can believe that one should be living up to a certain ideal, which goes beyond what one regards as required to avoid transgressing a moral norm. Thus there are areas of moral life where a moral value is in play, but shame is appropriate and guilt is not. While I do not intend to argue for the existence of a realm of supererogatory action, if we exist that such a realm exists, then it seems clear that one may regard oneself as being the type of person who sometimes goes above and beyond what is morally required, and feel shame when one fails to do so.

Of course, not every failure to perform a supererogatory action will elicit shame from every agent; whether it does or not will depend on how high we set our own moral standards. It is the mark of an exceptional moral hero, not an ordinary person, to hold himself to performing supererogatory acts whenever the need arises. Indeed, we might think that holding such high standards is symptomatic of an unattractive kind of moral narcissism. If, however, we accept that people legitimately set standards for themselves which go beyond doing their moral duty but are still of moral concern, then this leaves open the possibility for occasions where guilt is inappropriate but shame is appropriate, in relation to a matter of moral relevance. The full range of the supererogatory will not be picked out by what ordinary moral agents like ourselves *would* feel shame over failing to

perform, but what we can make sense of somebody appropriately feeling shame over failure to perform.

This idea is present in the work of Gregory Trianowsky, who notes that when we fail to do a supererogatory act, we often offer excuses – think of the person who asked to give money to charity says "Sorry, I don't have any change" or "I gave earlier".[200] It somehow seems wrong to simply say "I don't want to give you any money" in these circumstances. This seems strange, however, as excuses are usually employed to deflect negative judgements about us by showing how something that would normally be a case of wrongdoing is, because of some feature of the wider circumstances, not a case of wrongdoing at all. Failure to do a supererogatory act, however, is, by hypothesis not a case of wrongdoing. So why do we feel the need to offer excuses in this situation?

The answer Trianowsky gives is that while no negative *deontic* judgements (i.e. judgements about the moral value of the act) are in order in such cases, it is possible that some negative *aretaic* judgements (i.e. judgements about the moral character of the agent) could be applicable. These could be negative judgements about some standing trait of the agent or negative judgements about a person's occurrent motivational structure. It may be the case, for example, that the woman in our example failed to attempt a rescue due to cowardice, or because at that time she was motivated in a cowardly manner.

As Trianowski points out, this illustrates a clear link between the supererogatory and virtue. The perfectly virtuous person will always perform a supererogatory action, when it is available to them (presuming we have in place a distinction between the supererogatory and the foolhardy[201]). When we fail to perform a supererogatory action, therefore, we show ourselves to be less than fully virtuous. As Burnyeat has pointed out, shame has a close relationship with failures to be fully virtuous – he refers to shame as "[Third party material excised].[202] When someone believes, or even suspects, that their failure to perform a supererogatory act is due to an unacceptable lack of virtue, shame may be a perfectly intelligible response.

This phenomenon is well-displayed in the case of imperfect duties, where doing something on any particular occasion is supererogatory, but one has a moral duty to do that thing on at least some occasion(s). While it may be argued that an act cannot be supererogatory and fulfil a duty, by

---

[200] Gregory Trianowsky, "Supererogation, Wrongdoing, and Vice: on the Autonomy of the Ethics of Virtue", *Journal of Philosophy* 83 (1986), 26-40.
[201] See Mellema, *op cit* for more on this distinction.
[202] Myles Burnyeat, "Aristotle on Learning to be Good" in *Essays on Aristotle's Ethics* ed. by A. Rorty (London: University of California Press, 1980), 69-92 (78).

definition, I would suggest that in the case of imperfect duties, if we specify the action as "doing X on occasion Y" then we have no duty to do *this* and doing it can be classed as supererogatory. An example of this might be giving money to charity; we might think that it is supererogatory to give money to charity on any given occasion, but it is a moral duty to do so on some occasion. If we conceive of an imperfect duty as a duty to perform a disjunctive action (e.g. giving money to charity on one of a list of possible occasions) then failure to perform this disjunctive action could appropriately elicit guilt. It seems, however, that there are some occasions when shame is an appropriate response to failure to give to charity, but guilt is not.

We can explain this type of situation by invoking Trianowski's distinction between deontic and aretaic criticism. The reason shame, but not guilt, may be appropriate over failure to fulfil to perform a supererogatory act after a certain number of opportunities is that while not enough opportunities have presented themselves for the non-performance to count as a failure to fulfil a moral duty, enough have passed that the agent may start to believe or suspect that his failure to perform the act is due to some character defect or vice on his part. Even failing to do a supererogatory act on the first occasion an act of that type presents itself can render shame an appropriate response if the agent believes that his inaction is due to a character flaw.

Thus the supererogatory seems to be an area of moral concern where guilt is not appropriate, but shame might be. If we believe that cases of supererogatory actions exist, then, we are committed to the claim that there are cases where moral value cannot be detected by appropriately-felt guilt, but can be detected by appropriately-felt shame.

Shame and Collective Guilt

A second area where we might think that shame is able to pick out moral value where guilt cannot is as the appropriate individual-level reaction to the moral failings of collectives of which we are members, but where we do not bear individual responsibility.

Martha Nussbaum picks out an example from Barbara Ehrenreich's study of low-paid work in America, *Nickel and Dimed*:

> [Third party material excised][203]

---

[203] Nussbaum, *Hiding*, 211 .

As Nussbaum points out, the situation of the working poor in America is not a result of isolated acts or omissions by individuals, but rather a failure of the will to act by society as a whole.

A similar point is made by Rai Gaita in relation to the attitude required of contemporary Australians to the historical wrongs committed by their country against the Aborigines: [Third party material excised][204] While guilt may be an appropriate response to individuals who played a part in the relevant deeds, or who failed to do enough to stop it, [Third party material excised], for whom shame is the correct response.[205]

Why do these philosophers claim that shame, rather than guilt, is the correct response to moral failings of groups of which we are part? The root of such claims lies in the idea that the object of guilt is always some action or omission of one's own. The object of shame, on the other hand, is oneself, seen as substandard in some way, and the reasons for which the self can be seen as substandard include relational properties; thus the actions of others to whom we are related in some way can form the basis of rational shame on our part.

Consider the example of somebody who has just discovered that a close relative is guilty of some terrible crime. If that person said that they felt guilty about it, an intelligible response would be to ask them what they thought they had done, or failed to do, that contributed to the relative's coming to commit the crime. If they were unable to point to any such act or omission we would be justified in describing their 'guilt' as irrational or even unintelligible. If the person said that they felt shame, however, we would not be justified in reaching the same conclusion; the person has recently discovered that they are the relative of a heinous criminal, they regard this as a negative relational property that applies to them, and they feel shame in consequence.

While one might argue that it is irrational to feel shame over actions for which one is not at all personally responsible, I would simply say that we quite readily understand such cases of shame, suggesting that they are accurate reflections of a widely-held paradigm scenario for the emotion, and so appropriate cases of shame. While it is of course right to say that the shame in these cases does not imply any moral failing on the part of the non-responsible individuals, the argument that their shame is irrational or inappropriate would require an argument that our paradigm scenario for shame is not well-fitted with our other emotional paradigm scenarios, and so needs adjustment.

---

[204] Raimond Gaita, *A Common Humanity: Thinking About Love and Truth and Justice*, 2nd edn. (London: Routledge, 2000), 92.
[205] See also Larry May, *Sharing Responsibility* (London: University of Chicago Press, 1992) for an insightful discussion of these issues.

Given the apparent strength of paradigm scenarios that do understand such cases of shame, such an argument would need to be a very strong one to convince.

When we consider collective moral failings then, it seems that an individual can only feel personal guilt when they have personally transgressed through action or omission, which has contributed to the wider failing of the group. This can include failing to protest sufficiently against the actions of the group (what one is morally required to do in this regard is likely to be more strenuous in relation to bad acts committed by a group of which one is a member than bad acts committed by those to whom one is unconnected). If you have discharged your moral duty to try to prevent the bad action of your group, or if you knew nothing of the bad action until it was too late, or if you were not yet born when the bad action takes place, then guilt will not be appropriate. Shame, on the other hand may be appropriate as a response to one's own negative characteristic of being a member of a group that has done a particular morally bad act, or set of acts.

This view is supported by Peter Forrest, who argues that:

> [Third party material excised][206]

While he is talking specifically about the wrong acts of past generations, there is no reason why his claim should not equally apply to contemporary wrong acts of collectives that we are members of, where we have not personally committed any wrong act or omission.

This schema is disputed, however, by Margaret Gilbert[207], who, cutting shame out of the picture altogether, argues that there are three different ways in which an individual can appropriately feel guilt in the context of a collective action. The first, which I have already discussed, is personal guilt felt over one's own acts and omissions relating to the bad action of the collective. This poses no problem to my argument.

The second type is what Gilbert calls "membership guilt". She quotes from Karl Jaspers' account of his self-directed negative emotions relating to German collective war guilt, but where Jaspers thought that such emotional reactions were irrational, Gilbert argues they can be perfectly rational. It is important to note that the reactions under discussion are not focused on personal actions and omissions, such as the failure to disassociate oneself from the actions of the group, or even leave it where possible:

---

[206] Peter Forrest, "Collective Guilt; Individual Shame", *Midwest Studies in Philosophy* 30 (2006), 145-153 (150).
[207] Margaret Gilbert, "Collective Guilt and Collective Guilt Feelings" in *The Journal of Ethics* 6 (2002), 115-143.

[Third party material excised][208]

While I do not doubt that the phenomenon that Gilbert is discussing is real enough, I am unable to understand why she wishes to call it guilt, rather than shame. The case seems analogous to the one discussed earlier of the person whose relative commits a serious crime. Regarding that example, I argued that shame is an appropriate response, but guilt is not; the same arguments seem to apply here, but Gilbert appears to overlook this. This omission is made even more inexplicable by the fact that she quotes Herbert Morris in support of her claim that membership guilt is possible:

[Third party material excised][209]

Noting in passing that Morris is talking about shame rather than guilt, Gilbert fails to remark on what this difference might signify, or why we should accept her claim that this type of emotional reaction is a case of guilt rather than shame. Absent any argument to the contrary, we have no reason to assume that the feelings under discussion are anything other than shame.

Where Gilbert is silent on this matter, however, another philosopher, Anita Konzelmann Ziv, offers an argument as to why guilt, rather than shame, is the correct individual-level response to wrongdoing by a collective of which one is a member, even when one is not personally implicated in the action(s) in question[210]:

[Third party material excised][211]

Guilt is fitted to play this role, she argues, because it can be experienced vicariously:

[Third party material excised][212]

I find both aspects of this argument unappealing. First, I do not agree that guilt can be felt vicariously. Second, I also disagree with Ziv's characterisation of shame, which renders the emotion unsuited as a reaction of a group member to wrong acts perpetrated by the group, in which she is not personally implicated.

It is widely believed that guilt can only be felt over actions for which one is personally responsible; indeed, this very characteristic has been used as the basis for an argument for guilt's superiority to

---

[208] Gilbert, 136.
[209] Gilbert, 134.
[210] In fact, she says that guilt or regret may be the correct response, but for the purposes of the current discussion, the relevant issue is her rejection of shame as a candidate for this role.
[211] Anita Konzellman Ziv, "Collective Guilt Feeling Revisited", *Dialectica 61:3*, 2007, 467-493 (490).
[212] Ziv, 472.

shame as a moral emotion, as we have seen earlier in this chapter. The supposed cases of vicarious guilt that Ziv suggests we can understand are susceptible to two competing interpretations.

First, they could actually be cases of vicarious shame – given that, as we have seen, shame is about the state of the self, which is partly constituted by our relational characteristics, vicarious shame, unlike vicarious guilt, is fairly straightforward to understand. Given the similarity of some of the affectivity of guilt and shame, and the psychological pressures that can lead us to falsely reframe episodes of shame as episodes of guilt, it is understandable that what are really cases of shame may be mistaken for guilt.

Second, these cases may actually be cases of straightforward, personal guilt (the only kind I think there is), rather than vicarious guilt. In cases where a child has committed a vicious crime, it is natural that the parent will think, or at least suspect, that there may be something that they did, or did not do, that contributed to their child coming to act in this way. Thus we would expect a parent in this case to feel guilt, not vicariously for their child's deed, but for their own, perceived or suspected, acts or omissions.

I also reject the claim that shame is not fitted to play this role. In claiming that shame is a character-directed emotion, Konzellman Ziv makes the same mistake I attributed to Rawls in Chapter 3, suggesting that shame is only ever directed at our intrinsic personal characteristics. As I argued, with Deigh, however, shame can also be about status, which can naturally include one's status as a national of a particular country, or a philosopher, or as a member of any group which is sufficiently important to one's self-conception. This being the case, anything which degrades the status of the group can cause shame in us, qua member of the group, regardless of whether we have any direct personal responsibility for the degrading incident. Ziv's mistake arises from adopting an excessively individualistic, even asocial, notion of the self which is the object of shame.

For these reasons, I reject Gilbert's assertion, and Ziv's argument, that guilt rather than shame is the appropriate emotion to feel in response to wrong acts by a group of which one is a member, when one is not personally culpable for those wrong acts.

The third way in which Gilbert argues that guilt can be felt in response to a collective act is by members of the collective constituting a plural subject of guilt, where the object of the guilt is the wrong act of the collective. She explains how such a state might arise as follows:

[Third party material excised][213]

What constitutes individual members of the group being jointly committed to feeling guilt as a body? In short, Gilbert says that this involves the individuals acting and speaking only in ways that are consistent with the judgement that the collective is guilty. For instance, such individuals will refrain from proposing similar actions for the group in future, will remonstrate with members of the group who deny that the group is guilty and will seek opportunities for the group to make appropriate reparations to the victim(s) of its acts.

Gilbert's plural subject conception of collective guilt invites the question, however, of who or what is *feeling* the guilt in question. As Christopher Kutz puts it:

[Third party material excised][214]

A collective cannot have any affective response beyond that of its individual members, and so long as they are not personally morally responsible for any wrongdoing, it is unclear why any member of the group should feel guilty. Gilbert's response is that this is not a problem; there is no particular phenomenology that is necessarily associated with personal guilt, so the fact that there is no phenomenology associated with collective guilt does not mean that it is not really guilt.

This approach, however, seems to leave an emptiness at the heart of instances of collective guilt. Specifically, it reduces feeling guilt to the mere recognition that one is in a state of guilt (i.e. recognition that some morally wrong action has been done); in other words the cognitive aspect of the emotion is present, but the feeling or affective aspect is conspicuously absent. Gilbert is incorrect to suggest that there is no necessary phenomenology that goes with guilt. She is led to make this mistake by employing an unduly narrow conception of phenomenology. Her comments suggest that she is thinking solely of physical feelings, and it is true that these may differ from case to case. When one person feels guilt, this may manifest itself as a nauseated feeling, another may experience it as a feeling of a heavy weight on the shoulders, while a third may experience guilt without any physical manifestation at all.

As discussed in previous chapters, these physical feelings do not, however, exhaust the feeling-aspect of guilt; we also have psychological feelings and effects. One defining characteristic of the emotions is that they affect our perception of moral saliencies, and so affect the direction of our attention and the way we conceive of what we attend to. These psychological effects are part of the

---

[213] Gilbert, 140.

[214] Christopher Kutz, *Complicity: Ethics and Law for a Collective Age* (Cambridge: Cambridge University Press, 2000), 196.

reason why the moral emotions are important; they are not merely passive detectors of moral wrongs, they focus our attention on our moral transgressions and failings, helping to clarify and deepen our sense of our moral values. While someone could recognise that they are in a state of guilt (i.e. that they have committed some moral transgression), unless this leads them to see certain things as morally important (for example, the suffering of their victim, or the need to avoid doing such actions in the future) then they are not experiencing the emotion of guilt.

In addition to this role in focusing the attention on salient features of the particular situation that elicits it, emotion also affects how we approach similar situations in the future. A useful concept from the psychological literature is the notion of a behaviour inhibition system, based on previous self-directed negative affect. This system is activated when we come across cues that have been associated with aversive reactions in the past. Once activated, it tags a possible course of action as 'faulty' directing our attention towards it, and so taking us off 'automatic pilot', leading us to think more carefully about whether our proposed course of action is the thing to do.[215] As Gilbert's account does not explain how it is possible for collective guilt to have any affective content, it is not clear how it could be expected to play this forward-looking role in the regulation of conduct.

While Gilbert gives an account of how a collective can recognise that it is guilty of some transgression, this does not amount to an account of how it is possible for a collective to feel guilt. While one option would be to say that collectives cannot feel guilt, this would be a problem for somebody who wants to adopt a neo-sentimentalist outlook. The neo-sentimentalist claims that something is morally disvalued if and only if it is appropriate to experience a particular moral emotion in response to it. If we believe that collectives can act, and act wrongly, then the neo-sentimentalist would need moral guilt to be appropriate to cash this out. But if collective moral guilt is never possible, then *a fortiori* it can never be appropriate. Of course, one option would be to deny that there is such a thing as irreducibly collective action, but this seems to go against some important intuitions.

One possible alternative is offered by Forrest, who argues that the feeling component of collective guilt is constituted by the feeling component of the shame felt by the individual members of the group[216]. As Forrest argues, there seems to be no reason why a single feeling component cannot be part of two different emotions. When we feel both shame and guilt on an individual level, it seems implausible to say that we have to alternate between, or combine, two separate affective states.

---

[215] Such accounts are discussed in Margo Monteith, "Self-Regulation of Prejudiced Responses: Implications for Progress in Prejudice-Reduction Efforts", *Journal of Personality and Social Psychology* 1993 (65), 469-485.
[216] Forrest, *op. cit.*

Rather, we seem to have one type of blended affective state, which forms part of both the shame and guilt that are defined by our self-directed negative assessments (of ourselves and our actions, respectively).

When we add Forrest's conception of the feeling element of collective guilt to Gilbert's conception of the judgement element – the paradigm scenario in my terms - (i.e. the individual's joint commitment to the judgement that the collective is guilty), this gives us a plausible picture of how a collective could genuinely experience guilt. The individuals' shame feelings will refocus their perception of moral saliencies, and as a result the collective's perception of moral saliencies (now of a summative kind) will also be affected. Thus the collective can be said to be experiencing guilt, rather than merely recognising that it is guilty. For example, my shame as a Briton over the wrongs committed by my country in the name of Empire, might lead me to regard certain issues around immigration and Britain's place in the world differently than would otherwise have been the case.

Where individuals are members of collectives that have committed morally wrong acts, but bear no personal moral responsibility for them, shame is the appropriate response on the part of this individual. This individual shame response on the part of members is partially constitutive of collective guilt. Thus shame is vital to marking the moral disvalue of wrongs committed by groups of the 'membership' kind. Indeed, one may go further, and argue that a tendency to feel shame at the wrong acts of a collective (and pride at its good acts) is partly constitutive of being a full member of a collective.

I have argued that there are two areas where appropriate shame is able to illuminate moral value but guilt is not: the supererogatory; and moral wrongs committed by groups. These are two cases where the field of moral concern extends beyond the moral transgressions of individuals; as individual guilt is the appropriate reaction just to the moral transgressions of individuals it is unsurprising that it is incapable of picking out these species of moral value. Shame, however, *is* capable of picking them out. As such, any neo-sentimentalist viewpoint that leaves shame out of the picture risks leaving us blind to these sources of moral value, giving us an etiolated picture of the moral landscape.

Gibbard notes that his proposal is not the only way of picking out morality, which consists of a cluster of features, but that [Third party material excised] and he [Third party material excised][217]. My response is that this is a sharpening too far. In restricting the range of the moral emotions to

---

[217] Gibbard, 293.

guilt, Gibbard fails to allow for important sources of moral value, which are capturable by shame, but not guilt.

One issue that Gibbard raises, however, still remains for me to address – the idea that shame is problematic as a moral emotion as it can be felt over a wide range of personal deficiencies, reaching from moral failings for which we would rightly hold ourselves responsible (such as our callousness in the face of people in need) to perceived personal weaknesses for which we are in no way responsible (such as our socio-cultural background, or a congenital disability). Gibbard seems to think that by encouraging the emotion of shame in the cases where shame recognises some moral value or disvalue, we will inevitably encourage it in those cases where it damagingly affects us for things that are morally irrelevant, and which we cannot change.

My initial reaction is that if this is the case, it is a price worth paying. I have argued that shame is able to capture important aspects of moral value that are missed by guilt alone, so if having shame at all means having these unfortunate cases of shame, then so be it. However, I think the choice is not really as stark as this suggests. While Gibbard is right to say that it would probably be impossible to restrict the appropriate feeling of shame to things for which we are personally culpable (and, given the importance of shame in picking out moral value beyond the bounds of personal culpability, nor should we attempt to), this is not the only way in which we can attempt to reform our experiences of shame.

The idea that proposing that shame should be taken seriously when it picks out moral value means that we will therefore be forced to take it seriously in all cases where it occurs, seems far too broad brush. Surely we can push at questions of whether particular instances of shame are appropriate, and what the appropriate response to different cases of shame are, on a case-by-case basis. Just because some cases of shame are futile and damaging, this does not mean that we ought to throw out an emotion that is uniquely well-placed to illuminate certain types of moral value.

I conclude, then, that shame is needed to supplement guilt as it can mark areas of moral concern which guilt cannot. I must now address the second strand of the case against shame, which says that shame brings with it a whole host of undesirable reactions which are potentially damaging to both the subject of the shame and to others, as well as undermining our ability to make good moral judgements. Unless I can successfully fend off these arguments, I leave myself open to the objection that shame is simply more trouble than it is worth. According to such an argument, while it may be true that shame is capable of capturing instances of moral concern which guilt is not, this

still does not mean that we should embrace shame as a moral emotion alongside guilt, because the costs (both moral and non-moral) of doing so are just too great.

In Defence of Shame (2): Not all Types of Shame Carry the Same Risk of Negative Effects

Having addressed the first strand of the case against shame as a moral emotion, by arguing that it is a necessary supplement to guilt, I must now turn to the second strand: that shame, unlike guilt, is prone to result in negative effects, which may be both undesirable in themselves and undermining of shame's role as a moral emotion.

Limitations of Empirical Work

My first counter-thrust against these arguments is to lodge some objections to the methodology of the empirical work carried out by Tangney and her collaborators. As I mentioned above, the method used in these experiments is to divide subjects into guilt-prone and shame-prone groups by setting out different scenarios for them, and then asking them to describe how they would feel in each situation. Based on the type of language used in their descriptions, each response is classified as guilt or shame, and then all the responses of each subject are used to assign him or her to one of the groups. These groups are then investigated to see how they correlate with various negative characteristics (tendency to depression, violent rage, lack of empathy etc.) yielding the result that a tendency to shame is related to the negative characteristic, while a tendency to guilt is not. There are, however, several reasons why we ought not to accept these findings at face value.

First, the experimenters divide subjects according to whether they are apt to feel shame or guilt in response to situations where either is an intelligible response. As I have argued, however, there are situations where shame is an intelligible, and morally valuable, response, but guilt is not. Therefore, their contention that the shame-prone group is more prone to various negative outcomes is beside the mark. The purported conclusion that people who tend to feel shame when guilt is also an option are particularly likely to experience negative effects does nothing to suggest that those who are apt to feel shame where guilt is *not* an option are also particularly likely to experience these effects.

In other words, the experimental methods do not allow us to distinguish between two possible hypotheses: first, that negative effects are correlated with a tendency to feel shame full stop, and second, that they are correlated with a tendency to feel shame in preference to guilt, where both would be intelligible responses. If the latter turns out to be true and the former false, this would not

license us to discourage all cases of shame; rather it would suggest we ought to restrict shame to the cases where guilt is not a viable alternative.

Second, the experimental method fails to take properly into account the fact that guilt and shame are very often felt together. It seems clear that there are cases where we alternate between two different emotional reactions to a situation; for instance, think of the parent who vacillates between being angry at their child for having wondered off without telling anyone, and relieved that they have returned home safely. This opens the way for guilt and shame to be felt simultaneously, with the focus of the subject's attention shifting from his deed to his character. The methodology adopted in these experiments, however, does not seem equipped to pick up this possibility, as we do not know to what extent the people in the two groups would tend to feel just the one emotion, or both emotions with that one dominant[218]. Again, the experiment does not allow us to distinguish between two possible interpretations of the results: whether the negative effects are correlated with a simple tendency to feel shame at all, or with feeling shame as the dominant emotion in a mix of shame and guilt feeling. If the latter is true, it suggests not that shame leads to negative outcomes in itself, but that it should be restricted to a subordinate role to guilt when both emotions are possible.

Third, the method ignores the fact that within types of shame experiences there are a number of distinctions which could have a bearing on whether the various negative effects will in fact result. One important distinction that is ignored is that between localised and globalised shame. Tangney and Dearing state that [Third party material excised][219] (emphasis added). They ignore the possibility of localised forms of shame, for which I argued in Chapter 3.

Similarly, Tangney and collaborators ignore the distinction between autonomous and heteronomous shame, for which I also argued in Chapter 3. Without taking these factors into consideration, their experimental results cannot tell us whether it is a tendency to feel shame *tout court* that is correlated with these negative effects, or just a tendency to experience a particular type of shame.

---

[218] In this regard we should also consider the wider difficulties with interpreting self-reported, hypothetical psychological states. We might be more willing to trust people's accounts of how they would feel in a hypothetical situation when this is restricted to an either/or emotional response (*either* shame *or* guilt). When we consider that much more complex, dynamic patterns of emotional response, incorporating both types of emotion, are possible (and indeed likely) our trust in the ability of individuals to accurately forecast and describe how they would feel may be diminished.

[219] Tangney and Dearing, 53.

Alternative Outcomes of Shame Episodes

The three arguments presented in the previous section give us a reason to think that the empirical studies under discussion fail to prove the existence of a necessary causal link between feeling shame and the various negative effects under discussion. They are, however, purely negative arguments, pointing out areas where the experimenters fail to take into account considerations which may be relevant to whether shame leads to negative effects. In the absence of a strong reason to believe that the considerations are in fact relevant, however, my arguments may not be very convincing. In order to address this, I will now revisit the proposed explanations of why shame supposedly results in negative effects to explore how these considerations may be relevant to whether this relationship in fact holds good.

Central to my strategy is the claim that there is an alternative, positive reaction that individuals can adopt in response to shame. This possibility is noted by Williams when he writes that [Third party material excised][220]. Similarly, Manion writes that:

> [Third party material excised][221]

Given that shame involves a pained, negative judgement about the self, surely one intelligible response to that is to resolve to improve the self so that this judgement is no longer justified. Naturally, the course of action needed to do this is not as straightforward as in the case of guilt, where due to a transgression an obvious act of reparation often presents itself. In shame, by contrast, there will be no magic bullet, no single act that can be taken to restore the self to its proper state. Instead, shame mandates a restorative *programme*, which necessarily takes time, through which the subject takes positive steps to guard against the failings of the self without faltering.[222]

To take an example, we might imagine someone who comes to realise that through excessive commitment to his work he has been neglecting to give sufficient time and attention to his children, and feels ashamed of this[223]. Of course, there is no single act that he could decide to do to suddenly become a properly attentive father. Over time, however, he might decide to make changes such as ensuring that he does not do any work on Sundays, making time to do activities with his children on a regular basis and so on. In this way he can rebuild his image of himself as a good father, and so

---

[220] Williams, *Shame and* Necessity, 90.
[221] Manion, 81.
[222] It must be noted that Tangney and her various collaborators recognise the possibility of a constructive response to shame, but they do not appear to think it much of a match for the competing, negative responses.
[223] For the purposes of his example, we must assume (perhaps implausibly) that his decision to spend so much time at work is purely a personal matter, and his prospects would be in no way diminished by his choosing to spend less time there.

remove the basis of his shame. He is thus able to react positively to the painful experience of shame, expiating the emotion without recourse to the destructive paths open to him.

An interesting link here is with the famous 12-step programmes for recovery from alcohol and drug abuse with their "one day at a time" mantras. Just as nobody recovers from alcoholism though a single act of will, nor can someone take a similar single act to recover from shame. Instead, one must adopt a programme of changes, which over time can lead to recovery.

With this alternative possibility fixed in our minds, we can now return to the explanations of why shame supposedly leads to negative consequences. For each of the varieties of negative outcome that have been claimed to be linked with shame, I want to argue that this more positive outcome is a possibility, and that the likelihood of this possibility being realised is vastly increased when a case of shame exhibits some of the characteristics that Tangney and others have ignored.

First, consider the purported link between shame and depression. The central argument here is that shame can result in feelings of hopelessness, as it brings with it the sense that nothing can be done to repair the damaged self. This hopelessness then leads to depression, according to this explanation.

While this may be a plausible story to tell in relation to some cases of shame, it seems much less so in other cases. First, consider the distinction between cases where the shame is felt over something one cannot help, such as a physical deformity, and those where it is felt over something within one's voluntary control, such as the father who is failing to spend enough time with his children. In the former case, there seems to be no positive action one can take to rectify the situation; whatever one does, one will still be deformed[224]. Feelings of hopelessness are natural in this situation because, to a certain extent, the situation genuinely is hopeless. There is nothing that can be done to change the 'failing' that has elicited the shame; it can only be put in perspective.

In the case of the workaholic father, by contrast, we have already seen how he could adopt a programme aimed at restoring himself to the status of doting father. Of course, while it is possible that in this situation the man might see his situation as hopeless, and fall into a depression, this seems much less likely than in the physical deformity case. On considering the situation, it is easy to see what might be done to remove the basis of the shame (although that is not to say it can be done quickly or easily). In other words, one is less likely to see such a situation as hopeless simply

---

[224] Of course, one could try to convince oneself that being deformed is nothing to be ashamed of, but in many cases this is likely to be psychologically difficult, if not impossible.

because it is clearly not hopeless. This suggests that the empirical work needs a clearer picture of *what* the shame-prone people are feeling ashamed of, and whether this is within their voluntary control or not.

Second, whether or not an ashamed person sees their position as hopeless might also be affected by whether their shame is localised or globalised. As I argued in Chapter 3, shame does not need to focus on the state of the whole self; rather it can also focus on some aspect of the self – shame about one's failings as a teacher, say, rather than for generally being a lazy or bad person. When shame is globalised, it seems plausible that it would be difficult for any positive plans to gain a foothold; when one sees oneself wholly negatively it may be more difficult to formulate a plan to get back on track. When the shame is more localised, however, it seems this would be easier.

To see how this might work, consider again the case of the workaholic father. If, on realising the type of neglectful parent he has become he feels ashamed of himself in a globalised sense, simply feeling that he is a terrible person, it may be difficult to come up with a workable plan to put this right. Spending more time with his children will not necessarily remove the shame, as this is seen as just one symptom of his underlying bad character. If, on the other hand, he feels a more localised shame, focused specifically on his failings as a father, then this tighter focus makes it easier for a positive course of action to emerge. In turn, this may make it less likely that the father would regard his situation as hopeless. For these reasons, then, we might think that globalised shame is more likely to lead to depression than localised shame, and this is a factor that needs to be taken into account in the empirical work.

In this regard, we also need to take into account the impact of the particular instance of shame, which I also discuss in Chapter 3. The type of consideration I am getting at can be expressed in terms of how hard the emotion hits someone, and the extent to which it claims their attention. As I noted above, a person can feel ashamed over an extended period without constantly attending to the focus of her shame. The more impactful a particular emotion is, the more forcefully and regularly the subject's attention will be dragged back to the focus of that emotion. While this may be related to the globalised/localised distinction, with globalised cases of shame tending to be more impactful, the two concepts can be distinguished. We can make sense of them coming apart in the real world, with very impactful cases of localised shame, and less impactful cases of globalised shame, certainly when comparing across different people.

We might think, then, that more impactful cases of shame are more likely to lead to feelings of hopelessness and depression than less impactful ones. When faced with a deep experience of

shame, one may be more likely to brood on what a bad person one is, rather than come up with a plan to rebuild the self back to how it should be. In fact, when shame is very impactful, it may seem that the self is so sub-standard that there is no course of action one could take to rebuild it, the damage to the self being seen as irreparable. In less impactful cases of shame, however, these pathologies may be less evident. Again, then, we should demand that empirical work on the proposed link between shame and depression takes into account the impact of individual instances of shame, as it seems this could be a relevant factor in determining the association with negative emotions such as anger.

Moving now to the supposed link between shame and anger or aggression, the proposed story is that, in order to escape painful shame feelings, the subject may transmute his shame into anger; the focus is shifted from a negative judgement about the self to a negative judgement about someone else. Again, however, we might think that certain sorts of shame are more amenable to this reaction than others.

The same arguments that I have just set out in regard to shame and depression can be carried across to the case of anger. Again, an alternative to the anger response is the positive, rebuilding response. As I argued above, this seems more likely to be adopted if the shame is focused on an aspect of the self that can be altered through a rebuilding project, if the shame is localised, and if it is less impactful.

In addition, we should also consider whether the shame is heteronomous or autonomous. When a case of shame is of the immature heteronomous kind, where the subject feels shame in response to negative judgements other make about her, which need not be shared by the subject, an angry response seems more likely. The shame is being imposed on the subject from the outside, and this may motivate anger towards those who are imposing this painful emotion. In cases of autonomous shame, however, it is the subject of the shame who passes the negative judgement on herself. While working oneself up into an angry state may mask the shame, by focusing her attention elsewhere, it cannot replace it altogether, as deep down the agent knows that shame is justified and anger is not.

Similar arguments can be made with regard to the supposed link between shame and withdrawal. As well as the same considerations about the availability of a positive alternative, the question of heteronomous versus autonomous shame is also relevant. When shame is imposed on us by the judgements of others, withdrawing from their view may be an effective means of escaping from shame feelings. When it is we who pass the negative judgement on ourselves, however, this seems less of an option, as it is impossible to withdraw fully from one's own gaze.

Finally, turning to the supposed link between shame and lack of empathy, we may again think that certain types of shame reaction are more likely to lead to reduced empathic feeling than others. The lack of empathy, however, seems to be a consequence of the other negative reactions (depression, withdrawal and anger). Where, by contrast, shame facilitates a positive, rebuilding reaction, this may take the form of coming up with ways in which we can be more responsive to the feelings of others – it may, in fact, spur us to adopt a project of building up our empathic capacities.

The standard criticisms that attribute bad effects to shame do not give sufficient weight to the possibility of an alternative, positive reaction, where the subject adopts a programme of rebuilding the self to a non-shameful state. I have argued that there are a number of considerations that may make this positive reaction more likely than any of the negative ones: when the shame is about something under the subject's voluntary control, when it is localised rather than globalised, when it is relatively low-impact and when it is autonomous rather than heteronomous. The empirical studies I have quoted ignore all of these distinctions; we should therefore be wary of accepting at face value their claim that shame (without qualification) has damaging effects.

Instead, we may posit that certain types of shame are liable to lead to negative effects, but for other types of shame these reactions seem much less intelligible. While this claim would, of course, be open to refutation by empirical study, I have yet to come across any study that takes all these distinctions fully into account; and we have ample reason to believe that any which did would come to a more nuanced conclusion. Both guilt and shame permit degenerate or pathological forms. There seems no reason to those associated with shame as licensing a general rejection of shame as a moral emotion.

It is important to be clear at this point, however, exactly what I take this argument to have established. I certainly do not intend to make the claim that, because I have argued that alternative forms of shame are possible, this therefore means that they are likely, or more likely than the negative or damaging forms of shame discussed in the empirical work I have discussed. Rather, my claim is that due to their failure to investigate these alternative possibilities, the empirical work so far completed does not make a wholly convincing case that shame is inevitably or usually a damaging emotion. Of course, further empirical study could show that the more positive possibilities I posit for shame are extremely rare, and negative forms are overwhelmingly the norm; I would then have to decide if I wanted to bite the bullet and say that the role of shame in picking out moral value was important enough that we should try to retain it despite these practical

difficulties. However, for studies to be able to do this, they must first contemplate the positive forms as a conceptual possibility; to my knowledge, none have yet fully done so.

If we accept that some cases of shame are damaging and some are not, this places shame on a par with guilt (we need only look to Nietzsche for ample illustration of the pathologies of the latter emotion). The question therefore moves from whether we should get rid of shame entirely, to which forms of shame are positive, and therefore to be encouraged, and which damaging, and therefore to be discouraged or mitigated. This is exactly as is the case with guilt.

My case in defence of shame is therefore complete. Contra Gibbard, we need shame as a moral emotion to supplement guilt. Contra Tangney et al, we need not accept that all forms of shame carry a substantial risk of damaging effects. Shame performs a unique and valuable function as a moral emotion, and it need not be damaging to oneself and others. For these reasons, we should accept shame as a moral emotion alongside guilt.

CHAPTER SIX: THE SITUATIONIST CHALLENGE TO SHAME AS A MORAL EMOTION

In this chapter I will consider one final source of possible objections to shame's status as a moral emotion: situationism. At the root of the situationist critique is a scepticism about the ascription of robust character traits to people, arising from a claim that what people do is much more likely to be determined by the situation in which they find themselves than by any facts about their character. Given that guilt tends to focus on individual actions, and shame on a person's character or status, situationism has the potential to challenge the place of shame as a moral emotion, at the expense of guilt. By arguing that situational factors affect our actions much more than we may otherwise have thought, situationism creates problems for guilt's status as a moral emotion; but by apparently licensing the conclusion that there is no stable self, situationism threatens to undermine the entire basis of appropriate shame, which would clearly be devastating to any moral theory that wanted to claim an important role in moral life for appropriately-felt shame.

Social psychology and the Fundamental Attribution Error

The basis of the situationist critique is that we are prone to commit the Fundamental Attribution Error ("FAE"). The central point here is that in trying to explain people's actions and predict how they will act in the future, we systematically place too much weight on our beliefs about their supposed character traits, and too little on situational factors. Gilbert Harman, identifying Nisbett and Ross as the pioneers in recognising the prevalence of the FAE, quotes from them as follows:

> [Third party material excised] [225]

Clergy behave in similar ways to other clergy, criminals behave in similar ways to other criminals, and clergy behave in different ways to criminals. In explaining these patterns, the FAE thesis[226] goes, we rely too heavily on character traits (e.g. the clergymen are honest

---

[225] Gilbert Harman, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error", *Proceedings of the Aristotelian Society* 99 (1999), 315-331 (320).
[226] i.e. the thesis that we are systematically prone to committing the FAE.

while the criminals are dishonest). We should, instead, be relying more on situational factors (e.g. the clergymen are comfortably-off while the criminals are poor).

According to proponents of the FAE thesis, then, we are too quick to ascribe robust character traits to people on the basis a few instances of observed behaviour, so just because a colleague tends to have a kind word for me, I characterise him as a kind person when in reality he might be a tyrant in his home-life. To see this type of error in action, we need only think of the now-clichéd comments of neighbours interviewed on television when someone is arrested for a shocking crime: "He seemed like such a nice man".

Importantly, the effects of the FAE are not supposed to be all one way; as well as being mistaken in our inferences from behaviour to trait, we also have a tendency to be mistaken in our inferences from trait to behaviour. Thus once we have attributed to someone a particular character trait, we place undue weight on that trait in predicting how they will act in a given situation. We end up paying too much attention to supposed character traits, and not enough to situational factors, which makes us bad at predicting how a person will behave.

The contention that the FAE is widespread is not, however, the same thing as accepting situationism. The FAE thesis is an empirically-based claim about our epistemic practices; it says that we are prone to accept that somebody's behaviour implies possession of a character trait (and, secondarily, that a person's supposed possession of a character trait can explain/predict their behaviour) on the basis of too little evidence, systematically undervaluing the effects of variation in context. Situationism, on the other hand, involves making a negative ontological claim about the types of character traits that people can have, arguing that either character traits in general, or a subset of character traits, do not exist at all. The link between the two is that situationists draw on the experimental results that are taken to establish the truth of the (epistemic) FAE thesis, and from these infer the further (ontological) conclusion that situationism is true. In the realm of character traits, then, the FAE thesis claims that we over-generalise, while situationism claims that all, or an important subset, of our generalisations are baseless.

The two philosophers who have played the most important role in setting out the situationist challenge are Gilbert Harman and John Doris. They draw on a wide range of experiments that they claim show the truth both of the FAE thesis and of situationism. This literature is far too extensive to cover in any detail here, but we may divide the experiments into two categories: those that suggest that in some situations, people will act much worse than we would have

expected, and those that suggest that seemingly insignificant situational factors can have a much greater effect on behaviour than we would expect. For each of these categories, I will describe one paradigm example.

The first category of empirical work cited is exemplified by the infamous Milgram experiment[227]. In the original experiment, subjects were assigned the role of 'teacher' and a confederate of the experimenter was assigned the role of 'learner'. The subjects believed that the assignment of these roles was random, but in reality it was rigged. The learner went out of sight into another room. The teacher was told that every time the learner got an answer wrong on a word-pairing task (the experiment was set up so that about 3 out of every 4 answers were incorrect), he must move a lever to give an electric shock to the learner. The electric shocks were to increase in intensity from 60 volts (marked on the panel as 'mild shock') through to 420 volts ("danger: severe shock") and finally 435 and 450 volts ("XXX"). A recording was set up to lead the teacher to believe that the learner was responding to the shocks with increasing distress, begging to halt the experiment, refusing to participate and, finally, falling silent. If the teacher refused to continue, the experimenter used four escalating prompts in turn, ranging from "Please continue" to "You have no other choice, you must go on". If the 'teacher' refused to continue after the fourth prompt, the experimenter allowed them to leave.

Before the experiment took place, Milgram told another group about the scenario and asked them what they thought they and others would do if placed in the teacher role. Each person said that they would stop before they inflicted serious pain, and it was thought to be extremely rare that someone would go to the full 450 volts, with subjects predicting that this would happen in less than one per cent of cases. In reality, however, none of the experimental subjects stopped before reaching 300 volts, and 65 per cent (26 out of 40) continued all the way to 450 volts. Experiments of this type have been recreated in numerous different countries with many different groups of people, with similar results each time.

In terms of the argument for the FAE thesis, the Milgram experiment is taken to show that there are circumstances where a large majority of people will act in ways that are contrary to character traits such as kindness, or at least lack of callousness, that we would expect most people to possess. It is important to note that this is not a case of extreme circumstances,

---

[227] Both the Milgram and Samaritanism studies are widely discussed in the literature. My factual summaries are taken from the discussion of them in Goldie, *op. cit.*

where we would expect even a kind person to crack and perform a cruel act.[228] The subjects of the Milgram experiment were not threatened with torture themselves if they refused to comply, for example.

The second type of experiment involves scenarios where a minor change to the situation results in a significant change in the behaviour of subjects, suggesting that the details of the situation are of much more importance in determining how people will act, relative to whatever character traits they may possess, than previously thought. My paradigm example of this group of cases is the Princeton Samaritanism study.

In this experiment, Princeton seminarians were told that they needed to prepare a lecture. Some of them were told that their topic would be possible career paths for people on their course of study, and others that it would be the parable of the Good Samaritan. The experimental subjects were asked about their views of religion to ascertain whether their primary concern was with their own salvation or helping others. Finally, as they were told to walk across campus to deliver their lectures, some were told that they had plenty of time, some that they had just enough time, and others that they needed to rush or else they would be late.

As the subjects were walking to deliver the lecture, they walked past a confederate of the experimenter, who was pretending to be in some distress (it was supposed to appear ambiguous whether this person was ill, or had been in an accident, or was drunk etc.); it was observed whether the subjects stopped to help this person. Of the three variables (topic of lecture, attitude towards religion and amount of time they had) the only one that had a significant relationship with whether or not they helped was the amount of time they had: 63 percent of those with plenty of time helped, as opposed to 45 percent with just enough time and just 10 percent of those that were running late. When we think of a person who is kind and decent (or at least not cruel) we expect them to reliably engage in kind and decent acts (or avoid cruel ones) when this is appropriate, but the results of the Princeton Samaritanism experiment, along with many other similar experiments, would appear to show that this is not justified.

What are we to take, then, as the lesson from experiments of these two types? Situationist philosophers have argued that we ought to conclude that our claims about character traits are

---

[228] In other words, it seems no valid excuse based on the extremity of the circumstances is in play – see my discussion of excuses later in this chapter for more on this possible line of defence.

systematically mistaken. While we think that character traits, or character traits of a certain type, exist, the situationist argues that this is not true. In the next section, I shall set out the main situationist arguments for this conclusion, before offering my response to them.

From Empirical Results to Situationism

Harman argues that empirical findings from the type of experiments under discussion license a radical conclusion:

> [Third party material excised][229]

Harman claims that the experimental evidence shows that there are no broad character traits; that is, there are no character traits that reliably deliver trait-consistent behaviour across time and across a wide range of different circumstances. People cannot be said to be simply kind or honest, as whether one will display behaviour characteristic of kindness or callousness, honesty or dishonesty, depends heavily on the situation in which one finds oneself. Harman offers no real argument for drawing this conclusion from the experimental findings, apparently taking it as obviously warranted. People we normally think of as kind behave in ways characteristic of cruelty in the Millgram experiment so they can't have been kind at all; people we think of as kind in the Princeton Samaritanism experiment behave in ways characteristic of kindness when they are on time and unkindness when they are late, so they could not have been kind at all. The conclusion we ought to draw from this goes beyond the epistemological (we were wrong to think of these people as kind) to the ontological (there is no such thing as kindness or, by extension, any other broad character trait).

Harman considers a claim by Owen Flanagan that this does not mean there are no character traits. Rather, Flanagan suggests, it merely means that there are no character traits of unrestricted application, or that are totally context-independent. While it may be true that there are no broad character traits, such as honesty or kindness, Flanagan's argument goes, there may be narrower character traits, such as honesty to one's mother or kindness to small children. Such an argument would moderate Harman's position, so that instead of arguing that we were systematically mistaken in thinking that there are any character traits, we would instead be mistaken merely in thinking that there are broad character traits.

Harman acknowledges that experimental research does show regularity of behaviour when circumstances are very similar. For instance, one study showed that students who took the

---

[229] Harman, 316.

opportunity to cheat on an exam by surreptitiously looking at an answer key were disposed to do the same thing when, subsequently, similar opportunities presented themselves; these students were not, however, disproportionately disposed to cheat in other ways. Another study showed that talkativeness at lunch was a fairly stable disposition across time, but was not a good predictor for talkativeness in other situations.[230]

However, for Harman this type of narrow disposition cannot constitute a character trait: [Third party material excised][231] Harman gives the example of somebody who refuses to ride any roller-coaster, but is not disposed to react with cowardice or undue fear in other situations. This person, Harman claims, cannot be described as a coward, and his refusal to ride the roller coaster is not a manifestation of cowardice. Similarly, if this person develops an aversion to acting in other ways, such as speaking up in History class (but not other classes), then, unless there is a shared explanation of the two dispositions (anti-roller-coaster-riding and anti-speaking-up-in-History-class), the two cannot be explained by possession of a single, overarching character trait, such as cowardice.

Harman gives us two reasons why we should not accept narrow dispositions as genuine character traits:

> [Third party material excised][232].

If Harman is correct, then this is potentially extremely troublesome for my advocacy of shame as an important moral emotion. As I argued in the previous chapter, shame is to be distinguished from guilt by its focus on the self, as opposed to particular actions. We feel shame because of what we *are*, or what we have *become*, rather than because of what we have *done*. If Harman is correct that there is no such thing as character, however, then talk of what we are or have become is otiose. For my feeling of shame to be appropriate, it would need to be true that I have fallen short of being the type of person I ought to be. If Harman is correct, however, I could never be an aretaically substandard type of person, because I am, in character terms, no *type* of person at all. If Harman's position can be sustained, then, as I have characterised it, feeling shame over one's character could never be appropriate, and so shame

---

[230] c.f. Harman, 326.
[231] Harman, 318.
[232] Harman, 326.

could not serve as a moral emotion, at least in many of the central cases I have assigned to it.[233]

The other leading advocate of situationism, John Doris, takes the more moderate approach suggested by Flanagan. He argues that we need not reject the existence of character traits as a whole, but that we do need to think about the type of character traits we are justified in positing. In particular, he argues that we must move away from a "globalist" conception of character. Doris defines globalism as the view holding the following three theses:

> [Third party material excised][234]

The first two theses are about the robustness of traits across different types of situation, and the same type of situation across time, respectively. On a globalist reading, if someone is kind when faced with a crying child, we would expect them to be kind in response to an injured animal. Similarly, if someone is kind to a crying child today, the globalist would expect them, all other things being equal, to be kind to crying children they meet in the future. The third thesis is about how traits go together in a single person's character, suggesting that a person who has one positive trait is more likely to have other positive traits than negative ones; someone who is kind, the claim goes, is more likely to be honest than someone who is not kind, all else being equal.

Doris argues that the first and the third of these theses are unsustainable. On consistency, he claims, the experimental evidence suggests that human traits tend not to produce trait-relevant behaviour across a variety of trait-relevant situations. For example, the Princeton seminarians who were in a hurry should, if they possessed the character trait of kindness, have stopped to help the person in trouble (i.e. they were in a trait-relevant situation) but they did not (i.e. trait-relevant behaviour was not produced). The stability condition, Doris argues, is met within a limited domain: evidence suggests that some traits are reliable across time, such as cheating by looking at an answer key when available (see discussion above). Finally, Doris argues that the evaluative integration condition is not borne out by the facts; people can quite easily be very courageous, for example, without being particularly compassionate.

---

[233] It could, for instance, still serve as a moral emotion in terms of being an appropriate response to group wrongs for which we are personally culpable, as shame's appropriateness here would depend on our relational properties (i.e. our membership of the group in question) rather than any intrinsic characteristic we are thought to possess.
[234] Doris, 22.

While Doris would therefore reject both broad character traits and evaluative integration of character at a meta-trait level, he leaves the door open for the existence of narrow traits, which Doris refers to as "local trait attributions". On this view, a behavioural disposition can be stable over time (thereby being fit to play the narrative and predictive roles in our commonsense psychology that I suggested were necessary to qualify a disposition as a character trait), but such a trait will not be consistent across a wide range of possible trait-eliciting circumstances (i.e. will not satisfy the "consistency" thesis of globalist psychology). As Doris points out:

> [Third party material excised][235]

While arguing that we should reject globalism about character traits, then, Doris is willing to accept, pace Harman, that narrow, or local, character traits do exist; or rather, that narrow behavioural dispositions can constitute genuine character traits.

While Doris, in accepting the existence of narrow character traits, does not go as far as Harman, his position may still pose a threat to shame's status as a moral emotion. This is because, Doris argues, shame may rely on illegitimate globalist assumptions about psychology:

> [Third party material excised][236]

If this were true, it would indicate a problem for shame as a moral emotion. Shame would never be appropriate, as the globalist judgements at its centre ("I am thoroughly selfish/cruel/bad") would themselves never be correct. As a result we could not place appropriate shame at the centre of our moral psychology, as no such thing as appropriate shame would exist.

We are left then with two different versions of situationism, each of which suggests a different argument against shame as a moral emotion. Harman's situationism claims that there are no such things as character traits. If Harman is correct, then character-based shame can never be appropriate, so shame's capacity to act as a moral emotion is undercut, across a wide range of circumstances.

Doris' situationism, on the other hand, claims that there are no broad character traits, only narrow ones. He suggests that shame may necessarily rely on an assumption that character

---

[235] Doris, 66.
[236] Doris, 155.

traits are broad, so that one is ashamed of being bad or cruel in general. Again, if Doris is correct, such emotions would never be appropriate, as one cannot truly be said to be bad or cruel tout court. As a result, our shame feelings about our characters would never be appropriate, and so shame would be unable to function as a moral emotion, across a wide range of cases.

Responding to the Situationist Critique of Shame as a Moral Emotion

I will respond to the two types of situationist critique in turn. First, in response to Harman's claim that there are no character traits, I will argue that narrow traits, which Harman acknowledges to exist, deserve to be classed as genuine character traits. This is because the features that Harman claims would rule them out from being classed as genuine character traits are in fact common to broad traits (which Harman says do not exist, but would be genuine character traits if they did). Second, I will argue that for each of the features of existing character traits that Doris argues for either (a) shame can function with traits like that or (b) we can resist Doris's claim that traits necessarily have that feature.

Turning first to Harman, then, I want to deny the strong claim that there are no such things as robust character traits. My strategy to do this involves invoking defeaters: contextual factors that explain why, contrary to initial appearances, possession of a particular character trait does not require a particular course of action.[237] For instance, suppose I come across a young woman, obviously lost and in distress. I might suppose that a kind person would offer to help the woman find her way, and so the people who I see who are in a position to intervene, but do nothing, cannot be kind. Unbeknownst to me, however, those people know that the woman is a dangerous criminal, attempting to flee from the police. The onlookers' failure to act in no way indicates that they are cruel, as kindness does not (usually) extend to helping wanted criminals evade capture. While we might think that kindness requires us to help someone who is lost and in distress, the fact of that person being a criminal on the run acts as a defeater for this claim.

The existence of such defeaters throws up numerous difficulties for us in using character traits to play the narrative and predictive roles reserved for them in our commonsense

---

[237] The sense of "require" I intend here is a definitional one; kindness requires a response A in situation-type B, because anyone who did not do A in B could not possess the trait of kindness.

psychology. First, in any given case it could be true that some known defeater was in play without us knowing it. Second, while we know in advance that defeaters exist, and what some common defeaters to the requirements of certain traits in certain situations might be, we will not be able to state, in advance, all possible defeaters because the list is potentially endless. Third, we may not even be sure whether some consideration ought or ought not to count as a defeater (part of being kind is recognising a genuine defeater to kindness when presented with one, but not all of us are perfectly kind, and even if we are, we have no way of knowing for sure that we are). Finally, there is such a thing as a counter-defeater. In the example of the young woman on the run from the police, the defeater against kindness demanding that one intervenes was that she was a fugitive from the law. A counter-defeater might be that the police were corrupt and would violate her human rights if they caught her. Obviously, any of the three foregoing considerations about defeaters could also apply to counter-defeaters.[238]

Defeaters, then, are invoked to show why something that looks like an action that possession of a trait requires is not in fact required by possession of that trait. Excuses, on the other hand, are invoked to show why we should not rescind our former ascription of a particular trait to somebody, despite observing some behaviour that is (under ordinary circumstances) contrary to that required by possession of the trait. Take the example of a kind person who fails to help somebody in difficulty (presuming the difficulty to be inconveniencing, but not running the risk of serious harm) because she is distracted by the news of a recent bereavement. Ordinarily, failing to help the person in distress might justify us in saying that the non-helper could not really be a kind person. In this case, however, the non-helper has the excuse of being distracted by her own troubles. It is true that, in general, kindness requires helping in the sort of situation under discussion; the existence of the excuse, however, means that the non-helper's failure to help need not be taken as evidence that she is not a kind person.

Similar considerations come into play in regard to excuses as I discussed around defeaters. In any given case, there may be an excuse in play that we have not taken into consideration (perhaps because we do not know about it). We cannot know in advance all the possible excuses for a given situation. We may be unsure whether or not some consideration genuinely counts as an excuse in a given situation. Finally, there may be counter-excuses (considerations which render what would ordinarily be an excuse incapable of functioning as such), to which any of the three foregoing considerations might also apply.

---

[238] There is obviously also such a thing as a counter-couter-defeater, to which similar considerations apply.

The point of this brief discussion of defeaters and excuses is that although we use character traits in the traditional, broad sense to perform narrative and predictive functions, the way in which they do so is not unproblematic. The potential existence of defeaters and excuses in any given situation, and the epistemic relations in which we stand to them, mean that our deployment of character traits must always remain contingent, and subject to revision. This is not so much a point about the type of character traits we deploy (broad or narrow), but about the way in which we deploy them, and our attitude towards them after they are deployed.

All this means that even broad character traits do not play their role in our commonsense psychology in a straightforward or unproblematic manner. Turning to narrow dispositions of character, it is clear that they would be even more problematic than their broad counterparts. As well as suffering the same difficulties I have already described for broad traits, narrow traits bring the further problem of leaving us unsure how far they are supposed to extend. Take somebody whom we observe to cheat from an answer key when the opportunity arises, but who does not cheat when opportunities to cheat in different ways present themselves; it will be difficult to tell what character trait we ought to ascribe to him. Is he dishonest only when it comes to cheating from answer keys, or when he thinks he will not get caught, or some other option? We may need further information to make this decision, and any decision we do make would be open to being over-turned by further evidence (e.g. observation of further pertinent behaviour).

This is a genuine difficulty for us in using narrow dispositions to play a narrative or predictive role in our commonsense moral psychology – it means that any ascriptions of narrow dispositions we make would need to be contingent, pending further information. However, given that we do not see a problem with using broad character traits in this way, despite the difficulties posed by defeaters and excuses, it is hard to see why we should see this further difficulty brought by the move to narrow dispositions as the tipping point. Unless some argument to the contrary is forthcoming, we should accept that if broad character traits can play narrative and predictive roles in our commonsense psychology, so too can narrow dispositions. As such, we should accept that these narrow dispositions deserve to be called "character traits". We should therefore reject Harman's claim that the empirical evidence licenses the conclusion that there are no such things as character traits of any kind.

Having rejected Harman's claim that there are no character traits, I must now address Doris's claim that shame implicitly relies on an unsustainable globalist notion of what character traits

must be like; there are character traits, but they are not of the right kind to ground appropriate feelings of shame. It is important to note at the outset that there are two levels of globalism that Doris may be attributing to shame, corresponding to the two globalist theses that he regards as unsustainable: the consistency and the evaluative integration theses. Shame would be bound up with the higher level of globalism if it necessarily involved negative judgements about one's whole character, stretching across a range of different traits (i.e. acceptance of the evaluative integration thesis). This appears to be the claim that Doris is explicitly making ("shame's characteristic thought is that I am bad/weak/small"). On the other hand, shame might involve a lower level of globalism, by requiring implicit acceptance of the consistency thesis; in other words, if the negative judgements central to shame are necessarily based on broad character traits (I am totally selfish/cruel/cowardly).

The first of these alternatives can be dealt with rather quickly. Regardless of whether Doris is correct to say that the evaluative integration thesis is unsustainable, this does not pose a problem for shame's role as a moral emotion because it is not true that the appropriateness of all forms of shame depends on assuming that the thesis is correct. As I argued in Chapter 3, shame does not need to be globalised in terms of involving negative judgements about one's self as a whole (e.g. "that one is bad through and through"). It can instead take a localised form, and so be restricted to individual traits (e.g. "I am ashamed of myself for my rudeness").

Given the possibility of both globalised and localised forms of shame, the first, and most natural, interpretation of Doris's argument only bites on certain cases of shame, rather than on shame *per se*; as a result, it does not threaten to derail my project of establishing appropriate shame as an essential moral emotion. If anything, the situationist argument strengthens my case. In response to the arguments from Tangney et al, above, I said that the distinction between globalised and localised versions of shame was vital, as some of the supposed negative effects of shame seemed much more intelligible in the globalised case than the localised. In accepting Doris's argument that globalised shame is necessarily inappropriate (because it fails to track the facts), I am left saying that the best moral reactions will not include feeling shame over all aspects of oneself, so we are automatically protected from the damaging effects that may flow from experiencing shame of this kind.

The second possible interpretation of Doris's argument - that shame is unsustainably globalist in involving judgements about broad character traits – offers more of a challenge. If Doris is

correct that the only character traits that exist are extremely narrow ones, then this leaves those who would promote shame as a moral emotion on the horns of a dilemma. On the one hand, shame could involve judgements that invoke broad character traits, and so will be systematically mistaken. This would erode the distinction between appropriate and inappropriate shame that is at the centre of the case for shame as a moral emotion.

Alternatively, shame might be restricted to cases involving judgements about very narrow character traits, such as being prone to dishonesty in exams when an answer key is readily available. This, however, would have a severe effect on the capacity of shame to act as a moral emotion. As I have argued in earlier chapters, a major part of the role of moral emotions is motivational. In attempting to make a case for shame as a central moral emotion, I have argued that part of the value of shame is its capacity to drive us to improve our character, to avoid in the future failings that we have found shameful in the past. If shame were restricted so that the only character traits over which it could be felt were very narrow ones, it would be much less able to perform this role, as it could only lead us to improve in the future across an extremely restricted range of possible circumstances.

Both of these possible accommodating approaches create serious difficulties for the advocacy of shame as a moral emotion. In order to defend shame's position as a central moral emotion, then, I will need to adopt a more oppositional approach. This will involve arguing that Doris is incorrect to place such a tight limit on the breadth of character traits that can be said to exist.

As I have argued, of the three claims that Doris attributes to globalist views of character, it is the second, the consistency thesis, that is the source of potential problems for shame's role as a moral emotion. Recall that the consistency thesis requires that: [Third party material excised][239] What is in question is the range of application of the trait or disposition; broad traits would remain operative across the whole range of situations, while narrow ones operate only within a sub-section of them.

A putative broad trait, like kindness, would satisfy the consistency condition, as a kind person would react kindly in all circumstances where a kind reaction is called for (barring situations where a valid excuse is in operation or occasional lapses of character). A narrow trait, on the other hand, would not satisfy this condition; someone who has the narrow trait of being kind

---

[239] Doris, 22.

to small children would not necessarily be disposed to act kindly to help an adult who was obviously lost in a strange city, although this latter is a "trait-relevant eliciting condition" for kindness conceived as a broad trait.

The key question, then, is whether the empirical work Doris invokes genuinely shows a failure of traits to apply in trait-relevant eliciting conditions. I would argue that there is room to question the extent to which this is the case. As I suggested in my discussion of defeaters, what constitutes a trait-relevant eliciting condition is extremely complex. In addition, as my discussion of excuses suggested, the failure of a trait to be elicited in trait-relevant eliciting conditions does not necessarily imply non-possession of that trait, as a valid excuse may be in operation. To take a simple example, kindness clearly does not require leaping in to relieve the pain or peril of everyone one comes across who is in some kind of discomfort or danger, as any number of defeaters could also be present. The phrase "cruel to be kind", though misleading, illustrates this point quite well; the kind thing to do is not to fulfil the drug addict's immediate wishes by giving him money to buy drugs, but rather to refuse his request, no matter how much this pains him in the short term.

Turning to the two paradigm cases for situationism, which I recounted earlier, is it possible to make the argument that they provide evidence not that broad traits are absent, but that broad traits are extremely complex in their application and attribution conditions? If so, it could be possible that the subjects of the experiment do have the broad trait of kindness (or at least do not have the trait of callousness), but the situations in question contain some kind of defeater or excuse such that their behaviour is not inconsistent with such character traits.

In the Samaritanism case, is it plausible that lateness for an appointment is a defeater for kindness requiring that one helps someone who has fallen down? Unless being right on time for the appointment is of massive importance, probably not, on its own. Combined with the thought that others may be well-placed to help, however, it may be. Perhaps in such a situation, what kindness requires is that one ensure that others are better-placed to help than oneself. It is true that in the case where the subject was not running late, the presence of other people clearly did not prevent the subjects from thinking they should help the person, as most of them did so. The fact that they are not running late changes the import of the presence of other people, however. In their case, they are as good a candidate to help out as anybody else, whereas in the running-late case, others are better-placed to help. We do not need to say that it is simply that it is a case of plenty-of-time versus running-late that made the difference, and

this is not a genuine defeater. Rather we might say that it is plenty-of-time-and-in-as-good-a-position-to-help-as-any versus running-late-and-others-better-placed-to-help, and that this difference could constitute a plausible defeater in the latter case.

Similarly, one could argue that even if kindness *does* require helping when one is running late (i.e. the running-late subjects are in a trait-relevant eliciting condition for kindness), there is some excuse open to them which means that even though they did not, on this occasion, do what kindness requires, they still possess the trait of kindness. I do not wish to go into the many argumentative avenues around whether or not this is the case, but it is sufficient for my purposes if it can be agreed that this is at least a possibility.

Moving now to the Milgram experiment, it is much harder to make the case that there may be a defeater in play; it seems pretty clear that, barring science-fiction scenarios, acting so as to administer a potentially fatal electric shock to a stranger is never a kind thing to do. The terrain, however, seems more receptive to the claim that an excuse is in operation, and so the fact that the subjects do something diametrically opposed to the demands of kindness does not necessarily mean that they are not kind (or that they are callous).

Perhaps the most promising approach in this regard is to say that the particular set-up of the situation manages to short-circuit the heuristics people tend to use to decide if some action is permissible or not. As I have already discussed, the requirements of particular character traits are complex. In order to work out quickly what is required of one in a particular situation, one might therefore use heuristics to reach a conclusion. In the Milgram experiment, normally useful heuristics, such as that someone from a respected institution would not carry out an experiment that would seriously endanger any of the participants, fail to work. Of course, the kind thing for a subject in the experiment to do would be to notice the dissonance between the heuristic and the particular situation in front of them, and refuse to participate. The fact that in a stress situation someone falls back on a heuristic that happens not to work and so fails to act as kindness requires, may excuse them from having to relinquish their claim to be a kind person.

In these two cases, then, it is at least arguable that we can tell a different story from Doris. On Doris's account, the running-late subjects of the Samaritanism study and the subjects of the Millgram experiment lack the broad character trait of kindness, as they fail to act as kindness requires in a particular type of situation. On the alternative story, we may argue that these cases show not the non-possession of these broad character traits, but rather the complexity of

their requirements and conditions of attribution. In each of the two paradigm experimental cases I have discussed, we can point to the possible presence of defeaters and excuses that will allow us to argue that, in spite of their actions, we can maintain the attribution of the broad character trait of kindness to the subjects.

Perhaps an alternative way of making this argument is to say that our everyday understanding of character traits involves seeing them as an ideal to be attained, rather than as a descriptive and explanatory concept that must be displayed exceptionlessly by an individual, if they are to be said to possess the trait in question. Thus, we might argue that even if there are situations where somebody behaves in a way that is not consistent with kindness, this does not mean that they do not possess the trait of kindness; rather it could simply mean that they do not possess it as fully as could be the case.

This type of response to the situationist challenge is taken by a number of philosophers in the literature. For instance, Miller argues that failure to act in the way a trait would seem to require "should not necessarily be seen as evidence that the participants do not possess any global character traits whatsoever"; rather, the most this can show is that such agents [Third party material excised][240]. He goes on to argue that:

> [Third party material excised][241]

Miller's proposed strategy for virtue ethicists holds more generally for those who wish to resist situationist claims about character traits.

If this broad strategy could be used to counter all the empirical studies used by the situationists, it would be possible to argue that broad character traits do exist as in commonsense psychology, and so there is no problem for shame as a moral emotion. Despite the fact that this is a promising strategy in some areas, however, applying it across the board is not sustainable. One experiment in particular has convinced me that the complexity defence will not work as a blanket strategy: Isen and Levin's "dime-finding" experiment.[242]

In this experiment, subjects are people who are found making a telephone call from a public payphone in a shopping mall. The experimenters have rigged the phones so that some people will find a ten-cent piece in the phone's coin-return slot, while others will receive nothing. As

---

[240] C, Miller, "Social Psychology and Virtue Ethics", *Journal of Ethics* 7 (2003), 365-392 (379).
[241] Miller, 379.
[242] Details of this experiment are taken from Doris, *op. cit.*

the subject walks away from the payphone, a young woman (in reality a confederate of the experimenters) drops a folder full of papers, which scatter in front of the subject. The experimenters found that, overwhelmingly, those who found a dime helped the young woman pick the papers up (14 to 2), whereas those who found nothing did not (24 to 1). The implication here is that the subjects did not possess the broad character trait of kindness/compassion as whether or not they acted as was required depended primarily on whether or not they found some small change.

In this situation, it does not seem possible to argue that failure to find a ten-cent piece constitutes a defeater or an excuse, as it is such a seemingly insignificant factor. Someone who is genuinely kind would not be dissuaded from doing what kindness required by such a minor consideration. In other words, we do not seem here to be looking at effects of the complexity of the requirements and conditions of attribution of a trait, but rather at a genuine limitation on the broadness of the character trait; the trait simply does not seem to have effect in some situations. If we adopt the thesis offered by the experimenters, that finding the money changed the subjects mood, and so also affected their behaviour, this is potentially worrying for the defender of broad character traits. As Doris notes, [Third party material excised][243] We have no choice but to take mood effects seriously as a limitation on the breadth of character traits.

How are we to deal with this? One option might be to argue that this limitation is relatively minor. We can accept this limitation on the breadth of character traits without saying that all existing character traits are therefore narrow – it is just that they are not quite as broad as we might otherwise have thought. Character traits can still be said to be consistent, to deliver trait-relevant behaviour across a wide range of trait-relevant circumstances; what we cannot say is that they are *maximally consistent* as mood effects can prevent the appropriate behaviour from being displayed.

However, taking this approach would leave an opening through which defenders of a situationist approach could drive a coach and horses. As Arpaly suggests, [Third party material excised][244]. The same worries ought to be shared by anyone who wishes to resist a situationist take on character traits, for at least two reasons. First, the situationist could argue

---

[243] Doris, 30.

[244] N. Arpaly, "Comments on *Lack of Character* by John Doris, *Philosophy and Phenomenological Research* 71 (2005), 643-647 (646).

that mood effects are likely to be ubiquitous, so simply accepting that mood effects can limit the breadth of character traits would be to accept a significant limitation on the breadth of character traits in general. Second, if we simply accept that mood effects place a real limitation on the breadth of character traits, we may be led to accept similar limitations from other sources; before we know it, our notion of a character trait could end up salami-sliced through a number of such concessions. These two dangers mean that I would be running the risk of ending up with a notion of a character trait that is too thin to support the weight I wish to place on it.

One possible alternative would be to attack the robustness of the experimental evidence for such mood effects. This option taken by Christian Miller and Julia Annas, who argue that the effects of Isen and Levin's original dime-finding experiment have not been replicated in subsequent attempts to re-run the experiment.[245] While I note this possible response, I will not adopt it here; instead, I want to know whether, taking the original experiment's results at face value, they strike a fatal blow against non-situationist views of character traits.

Sabini and Silver offer two good reasons for thinking that the answer ought to be "no". First, they argue that while finding a dime may be sufficient to alter a person's mood so that she does not help with picking up papers, they [Third party material excised][246]. Second, and relatedly, they argue that mood effects work by broadening (in the case of good moods) or narrowing (in the case of bad moods) attention, and that [Third party material excised][247].

In effect, what these arguments are saying is that even if we take the results of the dime-finding experiment at face value, they are simply not about a serious enough matter to underwrite a claim that mood effects can lead us to act in a way that is incompatible with possession of a particular character trait. If the situationist could present us with evidence that finding a dime led supposedly normally compassionate people to ignore the screams of someone who was apparently in great pain, say, then this would be a significant challenge to those who seek to resist the situationist view of character traits. To my knowledge, however, no such experimental evidence exists.

Doris argues that there are hundreds of experiments supporting a situationist approach to character traits; clearly it will not be possible to address each of them here. Nonetheless, in

---

[245] Miller, 390. J. Annas, "Comments on *Lack of Character* by John Doris, *Philosophy and Phenomenological Research* 71 (2005),636-642 (637).
[246] J. Sabini and M. Silver, "Lack of Character? Situationism Critiqued", *Ethics* 115 (2005), 535-562 (539-540)
[247] Sabini and Silver, 540.

this chapter for each of the main examples found in the literature I have explained how they can be interpreted in a way which does not commit us to a situationist reading. It is my belief that, in general, the type of experiments the situationist relies on will all be susceptible to one of these strategies for non-situtionist interpretation. Of course I may be wrong about that, but the onus lies with the situationist to bring forward examples that are not susceptible to such strategies.

I conclude, therefore, that the situationist challenge to shame as a moral emotion, in cases where the shame is felt about the subject's character, is ultimately unsuccessful. Harman's radical situationism fails outright because he cannot sustain his contention that narrow dispositions do not count as character traits. Doris's more moderate situationism also fails to damage shame, as the empirical studies he quotes do not support his conclusions. Correctly interpreted, these experimental results show the complexity, rather than the narrowness, of our traits Rather than posing a threat to shame's status as a moral emotion, then, situationist considerations provide a clarification of how shame is able to play this role.

CONCLUSION

It is beyond question that shame can be a painful, crushing, isolating and damaging emotion. It can leave its subjects feeling depressed, isolated, and even suicidal. Throughout history and – all too often - even in contemporary society shame is visited on members of certain groups in virtue of factors such as their ethnicity, gender, socio-economic status, sexual identity, or physical or mental disability. Shame can blight people's lives.

Despite this dark side, however, we ought not to write off shame just yet. Through appropriately felt shame we can gain access to a set of important moral values that may otherwise have remained invisible to us. Shame can spur us to go beyond the minimum requirements to perform supererogatory actions, by making us reflect on the type of person we want to be. Shame can also help us recognise wrongs done in our name by groups of which we are members, leading us to demand that restitution is made and say "never again".

Nor is shame necessarily the wholly negative experience that it sometimes can be. A calmer, more civilised shame is possible - a shame that is focused on localised, reparable shortcomings rather than damning the subject's entire self character as defective or substandard, that is appropriately impactful but not pathologically so, that is based on the subject's own, autonomous assessment of herself rather than 'what the neighbours think', and that is focused on aspects of the self that are within the subject's control.

What we need, then, is not a concept of the ideal emotional repertoire that banishes shame completely. Rather, we ought to seek to cultivate the right type of shame, purging damaging and pathological forms of the emotion at the expense of shame that deepens and enriches our moral understanding. This type of shame can help us find an acceptable moral path and either stick to it, or get back on it when we falter. As a component of our emotional repertoire, shame is neither wholly good nor wholly bad, but to lose the capacity to experience shame is to lose a vital part of our moral capacity in search of an unearned, and ultimately unattractive, emotional quietude.

BIBLIOGRAPHY

Aristotle, *Rhetoric*, Revised Oxford Translation, vol. 2, trans. by W. Rhys Roberts (Princeton: Princeton University Press, 1984)

Annas, Julia, "Comments on *Lack of Character* by John Doris, *Philosophy and Phenomenological Research* 71 (2005), 636-642

Arpaly, Nomy, "Comments on *Lack of Character* by John Doris", *Philosophy and Phenomenological Research* 71 (2005), 643-647

Bartky, Sandra Lee, *Femininity and Domination: Studies in the Phenomenology of Oppression* (New York: Routledge, 1999)

Benedict, Ruth, *The Chrysanthemum and the Sword* 2nd edn (New York: Mariner, 2005)

Blackburn, Simon, *Ruling Passions* (Oxford: Clarendon Press, 1998)

Blum, Lawrence A., *Moral Perception and Particularity* (Cambridge: Cambridge University Press, 1994)

Burnyeat, Myles, "Aristotle on Learning to be Good" in *Essays on Aristotle's Ethics* ed. by A. Rorty (London: University of California Press, 1980), 69-92

Catechism of the Catholic Church 389, www.vatican.va.

Calhoun, Cheshire, "Cognitive Emotions?" in *What is an Emotion? Classic and Contemporary Readings* ed. by Robert C. Solomon 2nd edn. (Oxford: Oxford University Press, 2003), 236-47

Dancy, Jonathan, *Ethics without Principles*, (Oxford: Oxford University Press, 2004)

D'Arms, Justin and Jacobson, Daniel, "Sensibility Theory and Projectivism", in *The Oxford Handbook of Ethical Theory*, ed. by David Copp (Oxford: Oxford University Press, 2005)

— 'Sentiment and Value', *Ethics* 110 (2000), 722-48

Deigh, John, "Shame and Self-Esteem: A Critique", *Ethics* 93 (1983), 225-245

Dennett, Daniel C., *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge [Mass]: MIT Press, 1978)

Deonna, Julien, Rodogno, Raffaele and Teroni, Fabrice, *In Defense of Shame: The Faces of an Emotion* (Oxford: Oxford University Press, 2011)

Deonna, Julien and Terroni, Fabrice, "Distinguishing Shame from Guilt", *Consciousness and Cognition* 17 (2008), 725-40

Doris, John M., *Lack of Character: Personality and Moral Behaviour* (Cambridge: Cambridge University Press, 2002)

Dworkin, Ronald, *Law's Empire* (Cambridge [Mass.]: Harvard University Press, 1986)

Ellis, Joseph J., *Founding Brothers: The Revolutionary Generation* (London: Faber and Faber, 2002)

Forrest, Peter, "Collective Guilt; Individual Shame", *Midwest Studies in Philosophy* 30 (2006), 145-153

Gaita, Raimond, *A Common Humanity: Thinking About Love and Truth and Justice*, 2nd edn. (London: Routledge, 2000)

— *Good and Evil: An Absolute Conception* (Abingdon: Routledge, 2004)

Gehm, Theodor L. and Scherer, Klaus R., "Situation Evaluation/Emotion Differentiation" in Scherer (Ed.) *Facets of Emotion: Recent Research* (Hillsdale: Lawrence Erlbaum Associates, 1988)

Gibbard, Allan, *Wise Choices, Apt Feelings* (Oxford: Clarendon Press, 1990)

Gilbert, Margaret, "Collective Guilt and Collective Guilt Feelings" in *The Journal of Ethics 6* (2002), 115-143

Goffman, Erving, *Interaction Ritual*, 2nd edn (New Brunswick: Transaction Publishers, 2005)

Goldie, Peter, *The Emotions: A Philosophical Explanation*, (Oxford: Oxford University Press, 2000)

Greenspan, Patricia, *Practical Guilt: Moral Dilemmas, Emotions and Social Norms* (Oxford: Oxford University Press, 1995)

Harman, Gilbert, "Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error", *Proceedings of the Aristotelian Society* 99 (1999), 315-331

Helm, Bennett W., *Emotional Reason: Deliberation, Motivation and the Nature of Value* (Cambridge: Cambridge University Press, 2001)

Isenberg, Arnold, "Natural Pride and Natural Shame" in *Philosophy and Phenomenological Research* 10 (1949), 1-24

Kenny, Anthony, *Action, Emotion and Will* (London: Routledge, 1963)

Kutz, Christopher, *Complicity: Ethics and Law for a Collective Age* (Cambridge: Cambridge University Press, 2000)

Lewis, Helen Block, *Shame and Guilt in Neurosis* (Madison: International Universities Press, 1971)

McDowell, John, *Mind, Value and Reality* (Cambridge [Mass.]: Harvard University Press, 2001)

McFayden, Alistair, *Bound to Sin: Abuse, Holocaust and the Christian Doctrine of Sin* (Cambridge [UK]: Cambridge University Press, 2000)

MacKay, Donald G., Shafto, Meredith, Taylor Jennifer K., Marian, Diane E., Abrams, Lise and Dyer, Jennifer R., "Relations between Emotion, Memory and Attention: Evidence from Taboo Stroop, Lexical Decision, and Immediate Memory Tasks", *Memory and Cognition* 32 (2004), 474-488.

Mackie, John L., *Ethics: Inventing Right and Wrong* (London: Penguin, 1977), 38-42

Manion, Jennifer C., "The Moral Relevance of Shame" in *American Philosophical Quarterly 39(1)*, 2002, 73-90

Matsumoto, David and Ekman, Paul, "American-Japanese Cultural Differences in Intensity Ratings of Facial Expressions of Emotion", *Motivation and Emotion* 13 (1989), 143-157

May, Larry, *Sharing Responsibility* (London: University of Chicago Press, 1992)

Mellema, Gregory, *Beyond the Call of Duty: Supererogation, Obligation and Offence* (Albany: State University of New York Press, 1991)

Miller, Christian, "Social Psychology and Virtue Ethics", *Journal of Ethics* 7 (2003), 365-392

Mollon, Phil, *Shame and Jealousy: The Hidden Turmoils* (London: Karnak, 2002)

Monteith, Margo, "Self-Regulation of Prejudiced Responses: Implications for Progress in Prejudice-Reduction Efforts", *Journal of Personality and Social Psychology* 1993 (65), 469-485

Murdoch, Iris, *The Sovereignty of Good*, 2nd edn (London: Routledge, 2001)

Nichols, Shaun, *Sentimental Rules* (New York: Oxford University Press, 2004)

Niebuhr, H. Richard, Man the Sinner", *The Journal of Religion* 15 (1935), 272-80 (278)

Nussbaum, Martha, *Hiding from Humanity: Disgust, Shame and the Law* (Princeton, Princeton University Press, 2004)

— *Love's Knowledge: Essays on Philosophy and Literature*, (Oxford: Oxford University Press, 1990)

Prinz, Jesse, *The Emotional Construction of Morals* (Oxford: Oxford University Press, 2007)

Rawls, John, *A Theory of Justice* (Cambridge [Mass]: Harvard University Press, 1971)

Ricoeur, Paul, "'Original Sin': A Study in Meaning" in *The Conflict of Interpretations: Essays in Hermeneutics* 3rd edn (London: Continuum, 2004), 266

Roberts, Robert C., *Emotions: An Essay in Aid of Moral Psychology* (Cambridge: Cambridge University Press, 2003)

Rorty, Amelie Oksenberg , "Explaining Emotions" in *Explaining Emotions* ed. by Amelie Oksenberg Rorty (Berkeley: University of California Press, 1980)

Sabini, John and Silver, Maury, "Lack of Character? Situationism Critiqued", *Ethics* 115 (2005), 535-562

Sartre, Jean-Paul, *Being and Nothingness* trans. by Hazel E. Barnes (Abingdon: Routledge, 2005)

Shakespeare, William, "The Life of Henry the Fifth" in *Complete Works* ed. by J. Bate and E. Rasmussen (Basingstoke: Macmillan, 2007), 1033-1097

de Sousa, Ronald, *The Rationality of Emotion* (Cambridge [Mass.]: MIT Press, 1987)

Solomon, Robert, *The Passions: Emotions and the Meaning of Life* (Indianapolis: Hackett Publishing Company, 1993)

Stoker, Michael with Hegeman, Elizabeth, *Valuing Emotions* (Cambridge [UK]: Cambridge University Press, 1996)

Tangney, June Price and Dearing, Rhonda L., *Shame and Guilt* (New York: Guilford, 2002)

Tangney, June Price and Mason, George, "Moral Affect", *Journal of Personality and Social Psychology* 61 (1991), 598-607

Tangney, June Price, Wagner, Patrica and Gramzow, Richard, "Proneness to Shame, Proneness to Guilt, and Psychopathology", *Journal of Abnormal Psychology* 101 (1992), 469-78

Taylor, Gabriele, *Pride, Shame and Guilt: Emotions of Self-Assessment* (Oxford: Oxford University Press, 1985)

Trianowski, Gregory, "Supererogation, Wrongdoing, and Vice: on the Autonomy of the Ethics of Virtue", *Journal of Philosophy* 83 (1986)

Velleman, J. David, "The Genesis of Shame", *Philosophy and Public Affairs* 30:1 (2001), 27-52

Wallbott, Harald G. and Scherer, Klaus R., Cultural Determinants in Experiencing Shame and Guilt" in *Self-Conscious Emotions: The Psychology of Shame, Guilt, Embarrassment and Pride* ed by June Price Tangney and Kurt W. Fischer (New York: The Guilford Press, 1995), 465-487

Walker, Margaret Urban, "Moral Luck and the Virtues of Impure Agency" in *Moral Luck* ed by Daniel Staman (Albany: State University of New York Press, 1993), 235-50

Wiggins, David, *Needs, Values, Truth* 3rd edn (Oxford: Clarendon Press, 1998)

Williams, Bernard, "Moral Luck", *Proceedings of the Aristotelian Society* Supp. 50 (1976), 115-35

— *Shame and Necessity* (Berkeley: University of California Press, 1993)

Wolf, Susan, The Moral of Moral Luck" in *Setting the Moral Compass: Essays by Women Philosophers* ed. by Cheshire Calhoun (Oxford: Oxford University Press, 2004), 113-127

Wollheim, Richard, *On the Emotions* (New Haven: Yale University Press, 1999)

Ziv, Anita Konzellman, "Collective Guilt Feeling Revisited", *Dialectica 61:3*, 2007, 467-493