VALUING OTHERS

Moral responsibility and psychopathy

James Edward Baxter

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Philosophy, Religion and History of Science

March 2017

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

Heartfelt thanks to:

Ulrike Heuer, for guidance, support and many difficult questions, always offered with patience and wisdom.

Helen Steward for positivity, encouragement and insight.

Andrew McGonigal for some helpful early advice on PhD study.

The Inter-Disciplinary Ethics Applied Centre at the University of Leeds for supporting my study both financially and through flexibility and understanding, and particularly Professor Chris Megone and Kathryn Blythe for their invaluable help and support throughout.

My parents, Eddie and Rosemary Baxter, for hours of looking after my son while I wrestled with psychopaths upstairs, and for their emotional support and good humour. Joan and Mike Abrams for the same. Half of the work for this thesis was done on a crowded train between Sheffield and Leeds; the other half in a house in Sheffield filled with laughter and fun.

Bella Abrams, for being the best person I know.

Elijah Thorbjørn Abrams Baxter, for teaching me about responsibility, emotions, empathy and value.

Abstract

The question of whether psychopaths are morally responsible is a difficult one for philosophers and non-philosophers alike. In comparison to some other forms of mental illness, it is difficult to locate intuitions concerning what our attitudes to psychopaths should be and how they should be treated. This is because, unlike people with some other forms of mental illness, psychopaths (qua psychopaths) do not appear to be mistaken about the facts bearing on their choices, but they do appear to lack understanding of the world in an important way. Working within an understanding of moral responsibility as consisting in responsiveness to reasons, I argue that psychopaths lack responsiveness to certain kinds of reasons and are therefore not morally responsible for failing to act on reasons of these kinds. Based on a review of the empirical evidence, I conclude that psychopaths experience deficiencies of emotional engagement and of empathy, which are the result of events that are not under their control. I argue that these deficiencies lead 'hardcore' psychopaths (those at the high end of the scale for the deficiencies in question) to fail to develop the capacity to recognise entities other than themselves as sources of value, and thus to recognise that the rights, interests and concerns of others provide reasons which bear on their choices. These psychopaths are therefore not morally responsible for failing to act on such reasons. Nonetheless, I argue that these reasons apply to psychopaths' choices in a way that they do not apply, for example, to the choices of non-human animals. Implications of these conclusions include 1) that some reactive attitudes, such as resentment or hurt feelings, are inappropriate when directed at psychopaths, and 2) that some justifications for punishment are unavailable in the case of psychopaths.

Contents

List of figures

Introduction

Imagine someone who has always seen the world in a fundamentally different way from other people. To this person, thoughts about lying, stealing and violence hold no intrinsic repulsion at all, and if other people have importance, it is either as obstacles in the way of their goals, or else tools to be pressed into service, through lying, persuasion, manipulation or threats, to achieve those goals. Acts which would be repellent or even unspeakable to most people are available to this person as options which can simply be weighed alongside others. This is not ordinary selfishness, but a complete, blank inability to see other people as having any importance. Now imagine that this person has acquired this view of the world either purely through their genetic inheritance, or as the result of a childhood characterised by trauma and lacking in moral guidance, or as some combination of these two factors. Imagine, further, that there is nothing that they, you or anybody else could do to change their outlook. They are incorrigible.

You would perhaps be afraid of such a person. You would probably want to avoid their company. But how else would you think they should be treated? If they perform criminal acts, should their unusual psychology affect the way they are treated by the law? Should they be *blamed* for the harm they cause (and should they be praised for any apparently good acts they perform)? Would you be inclined to remonstrate with them, or resent them, if they did something thoughtless or cruel? The central idea behind all of these questions is that of moral responsibility. Should this kind of person be held morally responsible for their actions, emotions, attitudes, or the states of affairs they bring about? Answering this question would be difficult and it would, I think, force you to think very carefully about exactly what it is to be morally responsible.

You might begin by thinking about other cases of mental abnormality where we are more inclined to think of the person as not being morally responsible, and trying to draw a conclusion based on these less controversial cases. Take, for example, those mental conditions which are characterised by delusions. Imagine someone in the grip of a paranoid delusion, for example, who encounters another person whom they believe to be a persecutor – an alien in disguise, perhaps, bent on the destruction of humanity – and harms this person, in what they wrongly believe is self-defence. In this case we would not, I think, be inclined to hold the mentally ill person fully responsible for their act and the harm they cause. Through no fault of their own, this person fundamentally misunderstands the nature of their actions, in a way which is clearly relevant to the way we should react to them and treat them. However, it is unfortunately not the case that the person in our original description is suffering from a directly analogous case of misunderstanding. They do not think the people they harm are aliens, and they are not mistaken about the nature of their actions – at least not in the way that a person who wrongly believes the person they are harming to be a hostile extra-terrestrial is mistaken about the nature of their actions. In this sense, at least, they appear to know what they are doing.

The person with a paranoid delusion is mistaken about their reasons for action. Among other mistaken beliefs, they believe that they have a reason to defend themselves from a hostile extra-terrestrial. Because of this mistaken belief, they are unable properly to respond to the reasons they do have, such as to avoid harming what is in fact an innocent person. According to the account of moral responsibility which I will endorse, it is this inability to respond to the reasons bearing on their choice that renders the person not morally responsible for their actions. However, it is not clear, given this account of moral responsibility, whether we should think of psychopaths as being morally responsible. Psychopaths appear pathologically unconcerned about, for example, the harm their actions cause to other people. But a lack of concern cannot in itself be

excusing. What we need to know is whether there is something special about the lack of concern shown by psychopaths, perhaps given the way they acquire that lack of concern, which means we should think of them as not being morally responsible for some of their actions.

This question raises a further interesting question. If psychopaths are indeed, as I shall argue, pathologically unresponsive to reasons of a certain kind, can we say that such reasons apply to them? After all, a wild animal is plausibly not responsive to any reasons arising from the claims of humans not to be harmed, but this is not primarily because they are not fully rational. Rather, it is because such reasons simply do not apply to the actions of wild animals. We might try to prevent wild animals from harming us, but we do not think they have performed an immoral act if they harm us. Are psychopaths more like wild animals in this respect, or more like people suffering from delusions?

My thesis is that psychopaths, insofar as they lack empathy and therefore also lack responsiveness to a certain set of reasons, are not morally responsible for failing to act on those reasons. I am of course not the first person to argue that psychopaths lack moral responsibility. There has been a small but substantial literature on this question and philosophers have taken several routes to arrive at the same conclusion. It has been argued variously that psychopaths are not responsible because they lack moral understanding,[1] 'moral rationality',[2] or personhood,[3] or because they are incapable of fully-fledged reactive attitudes.[4] Within the framework of responsiveness to reasons which I favour, David

---

[1] Duff (1977), Fine and Kennett (2004).

[2] Morse (2008). Morse is concerned with criminal responsibility rather than moral responsibility.

[3] Murphy (1972).

[4] Benn (1999).

Shoemaker has argued that psychopaths are not responsible because they are incapable of being motivated to comply with reasons[5] and Neil Levy has argued that psychopaths lack both responsiveness to reasons and moral knowledge based on their supposed inability to distinguish between moral and conventional transgressions.[6]

On the other side of the debate, several philosophers have argued that psychopaths are indeed morally responsible, again for various reasons. It has been claimed that psychopaths have the cognitive resources that are necessary for responsibility,[7] that their volitional and emotional deficits are not enough to render them non-responsible,[8] that they are capable of forming intentions in a way that justifies ascriptions of responsibility,[9] and that they are capable of moral understanding.[10]

My own view is that psychopaths are incapable of responding to some of the reasons that genuinely bear on their actions. However, I do not think this is because of a 'factual' delusion about the nature of the world analogous to the delusions often experienced by schizophrenics, nor do I think it is because of the inability to parse different forms of transgression which would appear to be implied by James Blair's well-known experiments into the 'moral/conventional

---

[5] Shoemaker (2009), Shoemaker (2011).

[6] Levy (2008). In a later paper, Levy (2014), Levy has also argued that psychopaths are not responsible because they are incapable of forming judgments with the necessary type of content.

[7] Zavaliy (2008).

[8] Glannon (1997), Glannon (2008).

[9] Greenspan (2003).

[10] Maibom (2005), Maibom (2008).

distinction'.[11] As I argue in Chapter 3, I do not believe these experiments are firm enough ground on which to build an argument of this kind.

My own view is that the primary capacity lacked by psychopaths which is necessary for moral responsibility is the capacity to see others as valuable. Understanding this capacity, and what shapes it, allows us to bridge the apparent disconnect between the deficits experienced by psychopaths, which I will argue are primarily emotional in nature, and the unresponsiveness to certain reasons which I will argue accounts for their lack of moral responsibility.

This kind of analysis is needed partly because it is not clear what our pre-theoretical intuitions should be about the moral responsibility of psychopaths. This is a point that has been missed by a surprising number of philosophers. For example, R. Jay Wallace includes psychopaths in his list of 'accepted exemptions' from moral responsibility,[12] before going on to try to explain, in the context of his overall theory, why this should be so. My own experience is that it is precisely the difficulty of saying whether psychopaths are morally responsible that makes this an interesting question. I have trouble locating my own intuitions on the subject, and my experience of speaking to people about this suggests that my difficulty is widely shared.

The difficulty of knowing how we should react to, and treat psychopaths, is reflected by a lack of clarity in the criminal law surrounding psychopaths and responsibility. As Bartlett[13] notes, psychopathy (in common with other personality disorders) has never been successfully cited as part of an insanity defence, and moreover, personality traits related to psychopathy, such as a lack of remorse, may be taken as evidence of bad character and therefore lead to

---

[11] Blair (1995), Blair (1997).

[12] Wallace (1994), p. 166.

[13] Bartlett (2010).

harsher sentencing. This is perhaps surprising, given that the M'Naghton standard, which is applied to insanity defences, refers to 'a defect of reason, from disease of the mind' which leads the person 'not to know the nature and quality of the act he was doing; or that if he did know it, that he did not know he was doing what was wrong'.[14] A strong case could be made that psychopaths suffer from either of the conditions described by the disjuncts of this principle. However, psychopaths are excluded from the insanity defence as a result of particular interpretations of the phrases 'defect of reason' and 'he did not know he was doing what was wrong', which may stem more from expediency than from a desire for conceptual clarity.

Perhaps partly because of a wish to justify the existing practice in the criminal law of holding psychopaths fully responsible, the early philosophical literature on psychopathy and responsibility was dominated by a debate about whether the question could be settled *a priori*, without any reference to the empirical facts about psychopaths. Barbara Wootton[15] was the originator of this view, arguing that any argument against the responsibility of the psychopath must be circular, since the diagnosis of psychopathy itself will be based on facts about criminal wrongdoing, in which case the diagnosis cannot be taken to be an excuse for wrongdoing. Vinit Haksar[16] took the contrary view, on the grounds that psychopathy is a clinical diagnosis which can be made independently, since facts not connected to criminal wrongdoing.[17]

---

[14] M'Naghton case, quoted in Bartlett (2010), p. 35.

[15] Wootton (1959).

[16] Haksar (1965).

[17] In another paper (Haksar (1964)), Haksar suggests that psychopaths may not be 'choosing agents' – they can recognise moral values but are unable to choose them - and are therefore not responsible.

Wootton's view was perhaps understandable given the unavailability at the time of robust empirical accounts of psychopathy that were not simply based on records of criminal activities. However, following the establishment of clinical tools such as Robert Hare's Psychopathy Checklist (which I will discuss in detail in Chapter 2) the existence of psychopathy as a syndrome of personality, quite separate from any criminal activity in which it might issue, is now quite well established. Many 'successful' psychopaths never come into conflict with the law at all,[18] and Hare's checklist does not depend upon facts about the subject's criminal history for its application. Furthermore, neuroscience is now making significant advances towards identifying an independent neurological basis for psychopathy. This raises the possibility of a further means of diagnosis which is independent of any criminal history the subject may have.

Summary of the argument

If we are to answer the question of whether psychopaths are morally responsible, then, we must develop a clear picture of the psychological features necessary for moral responsibility, and of the psychological features which define psychopathy as a type of personality. My overall aim is to show that psychopaths lack some of the features that are necessary for them to be morally responsible. The overall argument that I will make can be summarised in this way:

1. A person cannot be held responsible for failing to act on reasons that she is unable to recognise as reasons.
2. Psychopaths are unable to recognise reasons for action stemming from the interests, needs and concerns of others.
3. Hence, they are not responsible for failing to act on them.

[18] Hare (1995).

The aim of Chapter 1 is to defend the first premise of the above argument, on the basis that moral responsibility is a matter of being responsive to the reasons that bear on one's choices. The literature on moral responsibility has been dominated by the debate over whether or not moral responsibility is compatible with causal determinism, and providing an answer to this question may not require one to develop a fully-fledged theory of moral responsibility. Such theories are, for this reason, quite thin on the ground. There are, however, three strands within the literature which, unlike other arguments within that debate, purport to explain and justify moral responsibility as a whole. These theories therefore deserve to be considered on their merits as attempts to do this, independent of their success in defeating the challenge from incompatibilism. The first of these theories is an attempt to justify moral responsibility as an institution based on its consequences. I will argue that this attempt fails, because it provides the wrong kind of explanation for why we hold some people morally responsible and others not morally responsible, and also because it leads to implausible results. The second, originating with P.F. Strawson, is very helpful in displaying the social nature of moral responsibility, and the way it is inherent in a wide variety of attitudes and emotions, not just the Aristotelian notions of praise and blame. This theory also offers a robust justification of the practices, attitudes and emotions involved in holding people morally responsible. However, Strawson is unable to offer a complete analysis of when it is right to apply, or withhold, judgements of moral responsibility. At the end of the chapter, I will argue that an analysis of this kind can be found in the work of R. Jay Wallace, who links moral responsibility to the idea of responsiveness to reasons.

Chapters 2 to 5 then defend the second premise above in a number of steps.

In Chapter 2 I develop a picture of the psychopathic personality-type based on the empirical literature. Psychopathy is a complex diagnosis, and there are some controversies about what elements of personality should be considered

central to it. Using sources from psychiatry, psychology and neuroscience, I will gather evidence of the peculiar deficiencies exhibited by psychopaths, concluding that these are primarily emotional in nature.

In Chapter 3, I consider various interpretations of these deficiencies in terms of moral responsibility, offering as the best interpretation that psychopaths do not recognise reasons stemming from the rights, interests and concerns of other people, due to their inability to recognise sources of value other than themselves.

In Chapters 4 and 5, I seek to bolster and support this interpretation by explaining it in the light of the peculiar emotional reactions of psychopaths that I noted in chapter 2. In Chapter 4, I draw on literature from the philosophy of the emotions to make the case that psychopaths' emotional deficiencies interfere with their ability to engage evaluatively with the world. In Chapter 5, I argue that empathy has a specific role to play in the development of the ability to see others as valuable.

Finally, in Chapter 6, I will turn to the question of what reasons psychopaths actually *have*. Are psychopaths like non-human animals, who are not morally responsible for harming people because they have no reason not to harm people, or are they like people with delusions, who have reasons not to harm people but are not responsive to those reasons? Answering this question will require engaging in the debate between internalism and externalism about reasons. Ultimately, I will conclude that psychopaths do have reasons stemming from the rights, interests and concerns of other people, and are not morally responsible because they are not responsive to these genuine reasons.

Chapter 1: Moral responsibility

Introduction

The word 'responsibility' in English is used in several different ways. For example, its meaning in the sentence, 'Nigel is a pretty responsible sort of guy' is clearly different from its meaning in the sentence, 'Anastasia is responsible for the death of my rabbit', or 'Hurricane Sandy was responsible for millions of dollars' worth of damage', and so on. If we are to make inquiries into the nature of responsibility, we would do well first to clarify exactly what sense (or senses) of responsibility we are interested in.

On the other hand, while the word has several distinct meanings, it is not merely by coincidence that we use the same word in each of the sentences above, or in others in which its meaning is different again. These meanings are related, though distinct. In this initial section, I will try to put the idea of responsibility into focus, by examining some of the different ways in which it is used, and exploring the relationship between these. This groundwork will be helpful later in the thesis, because it will allow me to separate out and begin to explicate the idea of *moral* responsibility which is my primary focus. Later in this chapter I will evaluate competing accounts of moral responsibility, and some groundwork will be helpful to this project too, identifying territory that is disputed between different accounts, and illuminating some apparent features of the concept of responsibility which those accounts try to explain. I will be drawing in this section on the work of Nicole A. Vincent, whose paper 'A Structured Taxonomy of Responsibility Concepts'[1] offers an analysis of what she sees as six separate responsibility concepts. I will not be adopting the whole of Vincent's terminology, however, which I think is misleading in places, nor will

[1] Vincent (2011).

I be following exactly her account of the relations between different kinds of responsibility.

At the end of this section, having developed a picture of the different ways in which responsibility is typically understood, I will focus in on moral responsibility, and consider the question of what we should be looking for when examining theories of moral responsibility. I will identify two things that a theory of moral responsibility ought to be able to provide: a justification of the practice of holding people (and perhaps other entities) morally responsible, and an explanation of *which* entities we should hold morally responsible, in what situations.

In the sections which follow, I will consider three attempts to provide a full-fledged theory of moral responsibility, and will evaluate them in terms of their ability to fulfil the two desiderata in the paragraph above. I will firstly consider, briefly, an attempt to set out a theory of responsibility in consequentialist terms, which I will reject. I will then turn to P.F. Strawson's account as set out in his lecture 'Freedom and Resentment'[2], and finally to the 'responsiveness to reasons' accounts which build on Strawson's insights, focusing particularly on the account offered by R. Jay Wallace in *Responsibility and the Moral Sentiments*.[3] I will suggest that an account of this kind constitutes the best available account of moral responsibility, and will outline the version of reasons-responsiveness which I personally favour. An implication of this account is that someone cannot be held morally responsible for failing to act on reasons which she is incapable of recognising as reasons, which is the first premise of the overall argument of the thesis as summarised in the introduction.

---

[2] Strawson (2008).

[3] Wallace (1994).

1.1 Senses of responsibility

Let us start by identifying some different senses of the word 'responsibility'.

As in the example of Nigel, who is 'a pretty responsible sort of guy', the word 'responsible' is sometimes used to refer to someone who has a particular virtue which manifests in a tendency to be trustworthy, consistent, and so on. They are 'a responsible sort of person'; they take their responsibilities seriously; they do not act *ir*responsibly. To describe someone as having responsibility in this sense – *virtue responsibility* – is to praise their character.

There is a very different sense of responsibility which is purely about causation; it has no moral dimension at all. A claim of *causal responsibility* is a claim about the causal history of an event or a state of affairs. The Hurricane Sandy example above is an example of mere causal responsibility: it makes no sense to speak of holding a hurricane morally responsible for the damage it causes. Similarly, if a computer virus wipes my hard drive and destroys the only copy of my PhD thesis, I might say that the virus was responsible for this destruction, but not in a sense that implied any moral assessment of the virus itself (any moral assessment of the people who created the virus would be additional to this immediate judgment of causal responsibility).

This contrasts with the sense in which the word responsibility is employed in sentences such as 'I hold you responsible for the damage you caused', or 'through his negligence in not holding on to the lead properly, Eric was responsible for the damage caused by his dog'. If a person is responsible in this sense for an action, then the person is liable for various moral repercussions arising from the action. For example, it might be that the person can be either blamed or praised for the action. It might also legitimise other attitudes and emotions, including resentment, indignation, and so on. In some cases, it might mean that social sanctions, such as shunning, are appropriate. It might also lead to expressions of disapproval (or approval), remonstration with the person,

or 'taking them to task'. All of these crucial elements of our social interactions rely on a judgment, whether implicit or explicit, about the person at whom they are directed: that they are *morally* responsible for some relevant action or state of affairs.

This sense of responsibility is what philosophers generally have in mind when they write about *moral responsibility*, and I will hold onto this term for convenience, though it is of course not the only sense of responsibility with a moral dimension (consider for example what I have called 'virtue responsibility').

Moral responsibility has a legal parallel in the idea of criminal responsibility. To say that someone meets the criteria of criminal responsibility in relation to a particular crime is to say that they should answer to the law in respect of that crime. It may be that someone who is causally responsible for a crime may yet not be criminally responsible, for example because they are too young, or because they have a mental illness which exempts them from criminal responsibility (the 'insanity plea'). It is also possible that someone might be criminally responsible without being causally responsible, as in cases of 'strict liability'. It may also be that criminal responsibility and moral responsibility come apart in at least some cases of strict liability.

Nicole Vincent also identifies a concept, separate from moral responsibility, which she calls 'capacity responsibility', having to do with the capacities people may or may not have which would make them candidates for judgments of moral responsibility. A judgment of capacity responsibility is a judgment of the entity as a whole, not in relation to any particular act, state of affairs etc. Clearly, there are some entities that are *never* capable of moral responsibility. We might say, for example, that a stone, or a baby, lacks capacity responsibility, in the sense that there is nothing for which the stone or baby is morally responsible.

In this sense, the stone or baby lacks whatever capacities allow an entity to be 'in the game' for attributions of moral responsibility in the first place.[4]

However, there are also cases in which people can lack moral responsibility for some things, or types of thing, but not others, because of certain capacities that they lack. The parallel concept of 'capacity' in medical ethics is illuminating here. Judgments about people's medical capacity are, in practice, always judgments about their capacity to do something in particular, for example to consent to a medical intervention. In many cases it is likely that moral responsibility operates in the same way. If someone suffers from paranoid delusions, it would not be appropriate to hold him morally responsible for insulting me if I know that one of his delusions has convinced him that I am a persecutor. If, on the other hand, none of his delusions apply to me at all, a judgment of responsibility does seem appropriate - he might simply not like me! Capacities, then, enter into judgments of moral responsibility for individual acts, as well as judgments of 'capacity responsibility' in Vincent's sense.

Finally, there is a sense of 'responsibility' which is roughly equivalent to 'duty' or 'obligation' – what we might call an '*obligation* responsibility'. To say that a referee has a responsibility to ensure that a game is played fairly is just to say that she has an obligation to do so. Sometimes these responsibilities are generated by the roles – social, contractual and so on – which we occupy, but this is not always the case. It makes sense to say that I would have a responsibility to rescue a drowning child if I could do so easily, and this would not be generated by any role I occupy (I would not need to be a lifeguard, for example, or to have any familial or other relationship with the child).

---

[4] There are also controversial cases in this area. For example, there is an ongoing debate within business ethics about whether an organisation is the kind of entity that can ever be morally responsible, i.e. that has capacity responsibility in this sense.

1.2 Moral responsibility, praiseworthiness and blameworthiness

In relation to moral responsibility, I noted that this idea is linked to attitudes including praise and blame. There is clearly a link between the state of being morally responsible and the state of being either *praiseworthy* or *blameworthy* – of being a proper object of praise or blame. But are they merely linked or are they in fact the same thing? It is not clear whether moral responsibility and blameworthiness/praiseworthiness can come apart. Some examples may help here.

Imagine I am visiting your house and knock over your valuable vase, breaking it. Depending on how this comes to pass, several implications of the event may differ, including your verdict over my blameworthiness or otherwise, how I would feel about it, and whether reparations on my part would be appropriate or not. Here are some possible cases:

> Vase 1: I knock over your vase intentionally, because I don't like the vase (or maybe I don't like you).
>
> Vase 2: I blunder into the vase accidentally, because I am not being careful, because I don't really care about your possessions or about the effect of my actions on your feelings.
>
> Vase 3: I fall over into the vase because your enemy pushed me into the vase with the intention of breaking it.
>
> Vase 4: I have a heart attack and fall against the vase, knocking it over.
>
> Vase 5: Your dog jumps up at me, and, being afraid of dogs, I back into the vase and knock it over.

In Vase 1 and Vase 2, it is quite clear that I am blameworthy for breaking your vase. In Vase 1, it is my intentional act that leads to the vase being broken, and there are no special conditions that should deter you from blaming me for it. In

Vase 2, it is my negligence – my failure to act in a way in which I ought to have acted – that leads to the vase being broken, and again there are no special conditions that should deter you from blaming me for it. In both cases, while I may not, in the case as described, feel bad about what I have done, it is clear that I *ought* to feel bad about it, and all other things being equal I am presumably liable for making reparations of some kind.

In Vase 3, it seems clear that I am not blameworthy. In this case, it was not my action that caused the vase to be broken, but your enemy's. I was *used* – and the blame for breaking the vase lies with your enemy, and not me. Nonetheless, I might feel some need to apologise to you. After all, it was my body that caused the vase to break. I was *involved*. However, the appropriate response on your part is surely, 'don't be silly!' rather than, 'apology accepted.' Regardless of my involvement in the scene, it was not my fault, and you should reassure me that there is nothing to apologise for.

Something similar seems to apply in Vase 4. I am not to blame because, again, the vase did not come to be broken through any action of mine. In this case, no-one acted. An unfortunate event occurred which resulted in the vase being broken. This time (at a stretch) I can perhaps imagine being moved to apologise for the broken vase (assuming I survive the heart attack of course). You would (I hope!) move even more quickly to reassure me that there is nothing to apologise for.

In all four of these cases, it looks as though blameworthiness goes hand in hand with action. In 1 and 2, it is my actions that cause the vase to be broken. In 3 and 4, this is not the case, either because (in 3) it was *your enemy* who acted and I was merely a passive object upon which she acted, or because (in 4) nobody acted.

In Vase 5, I do act – I back into the vase – and my action causes the vase to be broken. Let us assume that my act was voluntary – not that I voluntarily broke

the vase, but that I voluntarily moved to get away from the dog. However, not only have I acted without intending to break the vase, but it would also not be right to say that I have acted thoughtlessly or without due care. Is the breaking of the vase, then, an action for which I am blameworthy? Probably not. It is an accident, and it is *my* accident, but it is not one in which I am negligent or careless. It would seem unreasonable for you or anyone else to blame me, given the way I have described the case. Even more than in 3 and 4, however, I would certainly feel the need to apologise, and to offer to make reparation for the broken vase.

What do these cases tell us? Firstly, perhaps that my feeing the need to apologise does not imply that I accept blame for the incident, and also that its being right that I should apologise does not imply that I am blameworthy, or even that you would be justified in accepting my apology. Sometimes, it would appear, at least given the cultural norms that affect my own intuitions, my proper action is to apologise, and your proper response is to reassure me that there is no need to apologise. It would also appear that, for this to be true, all that needs to be the case is that I have some place in the causal chain resulting in the event in question. This is a very minimal requirement of *causal responsibility*: not that I need to have chosen to act, or even acted at all, in such a way as to bring about the event, but merely that I am involved in some way, even if only in that my body was one of the physical objects involved in the event's coming about.

The really difficult question is where *moral responsibility* fits into all of this. Personally, I find that attempting to test my intuitions about moral responsibility against cases like Vases 1-5 to be of only limited help. In the simpler cases, it seems fairly clear that moral responsibility tracks blameworthiness: I am both morally responsible and blameworthy in 1 and 2, and neither morally responsible nor blameworthy in 3 and 4. In the more difficult Vase 5, I find it hard to discern a clear intuition regarding whether I am

morally responsible or not. This is perhaps because moral responsibility is a technical term whose meaning and application are actually somewhat unclear. If this is right, then I will have to make a decision about what I am going to take moral responsibility to mean for the purposes of this thesis, and it will be reasonable to take this decision at least partly on pragmatic grounds: what definition of moral responsibility is most likely to play a useful role in my overall argument, and confer clarity on the debate that is to come?

In the Vase cases, the vase's breaking is an event, the vase being broken is a state of affairs, and the breaking of the vase is, in some variations at least, an action. Typically, we are responsible for events and states of affairs that are the result of our actions, or sometimes of our failure to perform certain actions. And again, typically, we are responsible for events and states of affairs that are the result of actions for which we are morally responsible. There are exceptions here, as perhaps when we are responsible for an action which leads to an event or state of affairs which we could not reasonably be expected to have included in our deliberation about how to act. Nonetheless, in the typical case, if we are responsible for the act, we are responsible for its consequences – for the events and states of affairs that result from it. At least, if we are to determine whether A is responsible for some given event or state of affairs, then we will need to know what action or failures to act on A's part have led to that event or state of affairs coming about, and we will need to know whether A is responsible for those actions or failures to act (and we may need to know some other things as well). The primary locus of responsibility, in this sense, is actions.

One option, then, is to link moral responsibility to action: if an action is attributable to me as an agent – if it is *my* action - then I am morally responsible for it.[5] In Vase 5, this would mean that I am morally responsible for breaking

---

[5] There is some ambiguity here around what it means for an action to be *my* action, or to be *attributable to me* as an agent. This depends on one's understanding of action. Whether it is enough simply for me to have

the vase, because the breaking of the vase is my action – I broke the vase – in contrast to Vase 3 and Vase 4. But I would plausibly not be blameworthy for it. So linking moral responsibility to action broadly fits our intuitions about the 'Vase' cases. However, in these cases it is my control over my actions that is in question. Beginning with Aristotle, lack of control is typically thought to be an excusing condition, with another being ignorance.[6] Equating moral responsibility with action, it turns out, fares better with cases in which *lack of control* is the excusing condition (including the five 'Vase' cases') than with cases in which *ignorance* is the excusing condition. While filming the movie *The Crow*, the actor Brandon Lee was killed by a bullet from a gun which was fired by an extra – the gun was supposed to contain blanks but somehow a live round had found its way in. It is surely true to say, then, that the extra killed Brandon Lee. But was he morally responsible for doing so? To say that he was is to abandon the idea that ignorance is an excusing condition on moral responsibility, since the extra was surely blamelessly ignorant of the most relevant fact in the case – that the gun contained a live round. This would not be disastrous – we would need to talk in terms of blameworthiness and praiseworthiness, at least when discussing matters which touch on the knowledge condition rather than the control condition – but it would put us at odds with the way moral responsibility is typically discussed by philosophers, and it is not clear that there would be any advantage to make up for this.

The better option, I think, is to link moral responsibility closely to blameworthiness or praiseworthiness. We are morally responsible for something if certain conditions (the exact nature of which we have yet to determine) are met, and we are praiseworthy if these conditions are met *and*

---

performed the action, or whether some further conditions need to be met, I think the result will be too thin a concept to be equated with moral responsibility, as I hope the following discussion shows.

[6] Aristotle (1985), Book ii, Chapter 9, Section 3.1.

praise is due to someone for the thing in question, blameworthy if blame is due. Taking this option allows us to retain the traditional view that there is a knowledge, as well as a control, condition on moral responsibility. However, it does raise two issues which I will deal with in turn before proceeding.

One issue is a potential unclarity around the distinction between justifications and excuses. This distinction, which has been much discussed by philosophers,[7] would need to be clarified in any case, but I shall need to make sense of it in the context of an account which links moral responsibility closely to praiseworthiness and blameworthiness. The basic form of this distinction is that if someone has a justification, then they have done nothing wrong, whereas if they have an excuse, they have done something wrong but are not to blame for it (and, I would have to add given my understanding of moral responsibility, are not morally responsible for it). When trying to apply this distinction to cases, however, the water becomes muddied very quickly. Did the extra in the Brandon Lee case do anything wrong? The answer to this perhaps depends on how we describe the action in question. It seems odd to say that they did anything wrong *in pulling the trigger*, since that was their job, and they had no reason to think that anything bad would result from it. But did they do anything wrong in killing Brandon Lee? Well, it was surely wrong for Brandon Lee to be killed. Furthermore, if it is supposed to be the case that someone who does nothing wrong by acting in a way which might have been wrong has a *justification*, then this does not seem to be the natural way to talk about this case. The extra was not *justified* in killing Brandon Lee. For this act to be justified would require that somehow it was right for Brandon Lee to be killed, which again clearly it was not. Better, then, to say that the extra did something wrong in killing Brandon Lee, but that they were not to blame for it. Because I am linking moral responsibility closely to praiseworthiness and

---

[7] Austin (1956) and e.g. Robinson (1996), Gardner (2007), Botterell (2009).

blameworthiness, I am therefore committed to saying that they are also not morally responsible for killing Brandon Lee. Since this seems to me a perfectly natural thing to say about the case I am happy to be so committed.

The other issue raised by the strategy of linking moral responsibility to blameworthiness and praiseworthiness has to do with acts that are neither blameworthy nor praiseworthy in themselves. There are actions which are morally neutral (e.g. going to the shop to buy some milk) and for which it would make no sense to use terms like 'blameworthy' or 'praiseworthy'. In cases which are *not* morally neutral, moral responsibility on my suggestion would be the state of being the proper recipient of praise and blame. Certain conditions would need to be fulfilled (that the agent is in control of her action, knows what she is doing, etc.) for her to be morally responsible in this sense. In cases which are morally neutral, those conditions still exist, but this does not legitimise praise or blame, because neither praise nor blame is appropriate in morally neutral cases. Am I then to be described as morally responsible for going to the shop for some milk, or not?

On the one hand, I can see that there is something strange about describing someone as *morally* responsible for something which has no moral dimension to it at all. On the other hand, it is possible to describe such a case in a way that exactly mirrors how one would describe a case which did have moral implications. If I went to the shop for orange juice, and picked up a carton marked 'orange juice' which for some reason contained milk, there is a sense (other than causal responsibility, in which sense I would be responsible) in which I would not be responsible for buying the milk. In fact, not much rides on whether we choose to call this sense 'moral' responsibility or to allocate some other name to it. By definition, nothing of moral consequence depends on attributions of this kind of responsibility in morally neutral cases. However, since what is being described *is* the type of state which justifies praise and blame where there is praise or blame to be justified, and since it is only in cases where

praise or blame is appropriate that we are likely to find ourselves discussing this type of state, it at least has the virtue of simplicity to maintain the same term both for cases which have moral dimensions and for cases which do not. I will therefore take this tack, but am unconcerned if the reader would prefer to reserve the term 'moral responsibility' for cases in which there is a moral dimension.

1.3 Moral responsibility for mental phenomena

I have been talking so far largely about actions, and have also alluded briefly to events and states of affairs. However, we are also, interestingly, often thought to be responsible for mental phenomena including attitudes, emotions and beliefs. Here, briefly, is an example of each of these mental phenomena: 'Stephen thinks that Johnny takes him for a fool, and demands an explanation.' 'Dave demands an apology from Ray because he believes Ray's anger at Dave is unjustified.' 'Neil takes Chris to task for his racist beliefs'. In each of these cases, the attitude, emotion or belief in question is attributable to the relevant person – it is *their* attitude, emotion or belief. This contrasts with the case in which an attitude, emotion or belief is not really attributable to me – say I have been slandered or misquoted.

Now, as with actions, there will be cases where states of affairs, attitudes, emotions and beliefs are attributable to me, but I am not morally responsible for them. So, in the Vase 5 case, the fact that the vase is broken is a state of affairs that is due to an act of mine, but I am not morally responsible because of the excuse provided by the dog. If Stephen takes Johnny for a fool because he has (through no fault of his own) mistaken Johnny for someone else who *is* a fool, then he is not morally responsible for his misdirected attitude. If Ray is angry at Dave because he thinks Dave has burned his hat, when in fact Pete has burned Ray's hat, and created a plausible situation in which it looks as though Dave burned it, then Ray is not morally responsible for his anger at Dave. If Chris has been brought up in a very isolated community, fed propaganda about

the supposed inferiority of some races, and not been exposed either to any real members of those races or to any opposing views, then he is (plausibly) excused from moral responsibility for his racist views.

Thus in the case of attitudes, emotions and beliefs, as in that of actions, there are conditions which must be met before someone is morally responsible for the thing in question, and it is only when they meet those conditions that any praise and blame can legitimately be attached to them. The most important question of this chapter is, how should we describe those questions? I will turn to this question shortly, but firstly I would like to revisit the different senses of the word 'responsibility' I set out earlier, and consider some relations between them.

1.4 Relations between senses of responsibility

Firstly, virtue responsibility appears to be loosely related to both moral responsibility and obligation responsibilities. Someone who has virtue responsibility is likely to recognise that they have certain obligations, and to recognise their moral responsibility (to *take* responsibility) for fulfilling, or failing to fulfil, these obligations. This is partly what we mean when we say that someone is a 'responsible sort of person'. Conversely, when we say someone 'abdicates responsibility', we tend to mean that they either fail to recognise or take seriously their obligations, or to act in a way that suggests that they accept *moral* responsibility for the fulfilment or non-fulfilment of those obligations. This is a good indication that they lack virtue responsibility.

Secondly, there is clearly a link between causal and moral responsibility. In many cases, it would be strange to say that someone was morally responsible for something while maintaining that they were not causally responsible for that thing. If Lee Harvey Oswald did not fire the gun that killed President Kennedy, then he was not causally responsible for Kennedy's death, and therefore could not be morally responsible for it either. However, as some of the brief examples

I have given show, it would be wrong to think that causal responsibility is always a necessary condition of moral responsibility. The problem with Eric and his dog is not that Eric caused the damage. The dog caused the damage, but Eric is still responsible for it because he failed in a duty to prevent the dog from doing so. Generally, we tend to think that parents bear at least some of the moral responsibility for the actions, or things caused by the actions, of their children. This is why a parent might apologise on behalf of their child, or offer to pay for damage, and so on. We can also be morally responsible for omissions – for things that we fail to do. A driver who fails to signal when turning right is morally responsible for this failure. The driver in this case has not caused anything to happen (this, if you like, is the problem).

Both of these examples also highlight a relation between moral responsibility and obligation responsibilities. We hold the driver morally responsible for failing to signal, partly because we believe she had a responsibility (obligation) to do so, and we hold a dog owner morally responsible for damage caused by his dog when he fails to keep it on a lead partly because we believe he has a responsibility (obligation) to keep the dog on a lead. The fact that the protagonists in these examples are morally responsible is an indication of the existence of an obligation that each has. However, it does not seem to be the case that being morally responsible, or being in a position to be morally responsible (in cases where the act in question has not occurred yet), for an act is a necessary condition of having an obligation responsibility to perform, or not to perform, that act. Imagine a football referee who sees a player fall over after being tackled in the penalty area, but, through no fault of his own, the referee is unable to see whether the footballer was fouled or not. Perhaps another player passed through his line of sight at the critical moment. Now, it seems to me that this referee is not morally responsible for failing correctly to judge whether a foul has taken place. However, I do not think it is the best explanation of this case to say that the referee does not have an obligation to

make this judgment correctly. I would say rather that the obligation stands, but that the referee is not morally responsible for failing to fulfil it in this case. Nonetheless, I accept that the correct verdict on cases such as this one is not always obvious – it does not seem incoherent to say that the referee has no obligation – and it may not be possible to argue conclusively for either side. This disagreement runs deep between competing theories of moral responsibility: we will encounter it again at the end of the chapter when looking at an account which emphasises the importance of responsiveness to reasons in explaining why we hold some people morally responsible and not others.

Perhaps a clearer exception to the close link between obligation responsibilities and moral responsibility is in cases of moral responsibility for supererogatory actions. An ordinary member of the public who rescues someone from a burning building is morally responsible for the rescue, but clearly not because they had any obligation-responsibility to do so.

As I have said, the sense of responsibility which is central to this thesis is what I have called moral responsibility. It is this concept that is the object of the philosophical work that has already been published on responsibility and psychopaths, and I tried to show in the introduction why it is interesting to ask whether psychopaths can have this kind of moral responsibility. It is also interesting to note connections to other senses of responsibility in the case of psychopaths, however. Firstly, it is certainly true that psychopaths on the whole lack virtue responsibility. As we will see in Chapter 2, failure to *take* responsibility for one's actions is one of the features by which psychopathy is diagnosed in clinical settings. Harvey Cleckley[8] sets out a series of case studies of psychopaths who manifestly and repeatedly fail to take responsibility – to recognise that they are morally responsible, both for the consequences of their actions for other people, and for their own lives – in any meaningful way. As I

[8] Cleckley (1941).

have noted, people with virtue responsibility do recognise that there are things for which they are morally responsible. Of course, the fact that psychopaths lack virtue responsibility does not in itself tell us anything about whether they *have* moral responsibility (or for what, if anything, they have it): it may be that they do, but that they fail to recognise this.

Secondly, the link to obligation responsibilities is also interesting. If it were the case that obligations were only possible where the person concerned could be held morally responsible for breaking those obligations, then either psychopaths must be capable of moral responsibility or else they could not have obligations. The latter conclusion would be a surprising one. It would mean, for example, that a psychopathic referee, or a psychopathic teacher, had no obligations at all generated by their role. However, it would also be strange to think that this in itself settled the question of whether psychopaths can be morally responsible. Perhaps, then, the example of psychopaths gives us another reason to doubt that there is such a close link between moral responsibility and obligations.

1.5 Theories of moral responsibility

My aim in this chapter is to draw on existing work in order to identify the best available account of what moral responsibility is. In the second half of this chapter I will examine some candidate accounts. But what exactly should these accounts be trying to explain?

To say that someone is morally responsible for something (an act, state of affairs, attitude, emotion or belief) is to say, first of all, that they meet certain conditions in relation to that thing. This may include some claim of causal responsibility (though, as we have seen, there are exceptions to the link between these two ideas). It may also include discussion of the obligations that the person has. In addition to these, however, the conditions of moral responsibility are also generally thought to include two other things: a

particular kind of control and a particular kind of knowledge. This thought can be traced to Aristotle, who begins his discussion of voluntary action in the *Nichomachean Ethics* with the two central claims that 'feelings and actions…receive praise or blame when they are voluntary'[9] and that 'what comes about by force or because of ignorance seems to be involuntary'[10]. Substitute moral responsibility for voluntary action and we have the basis of much of the discussion of moral responsibility that has followed. It is interesting to note that Aristotle makes no attempt to explain why it is that ignorance and compulsion, and not other conditions, are thought to be adequate excuses. This is simply taken as given. There are also controversies around the application of these conditions. The most obvious of these is the dispute between compatibilists and incompatibilists about moral responsibility and causal determinism: can someone be said to be in control of their actions if they exist within a causally deterministic universe in which those actions are predetermined and unavoidable? There is also the difficulty of determining whether or not someone with a mental illness is in a position of ignorance in regard to the nature of their actions. It would be a strength of any theory of moral responsibility, then, if it could fill these gaps by giving us a way of deciding who, and in what circumstances, is morally responsible, and a clear explanation of why this is so.

A claim of moral responsibility is also, however, a claim about what should be *done* in relation to the person or other entity who is responsible. Aristotle links his conditions of voluntary action directly to the practices of praise and blame: when someone acts voluntarily, it is proper to praise and blame them. However, it is surely true that several other practices and attitudes are legitimised by moral responsibility. An obvious example is that the appropriateness of rewards

---

[9] Aristotle (1985), ii 9 1109b30-34.

[10] Ibid., 1110a1-2.

and punishment depends on verdicts of moral responsibility. Emotional reactions too can depend on whether we think someone is morally responsible or not. If my friend cooks a meal for me and I subsequently catch botulism, I might be upset and even angry towards her. But if it turns out that the food was contaminated by my enemy leaning in through the kitchen window when her back was turned, so that she could not have known what she was serving me, it looks as though she is not morally responsible for my illness. I would change my attitude towards her, or if I did not, my continued anger and upset would seem to be misplaced. My emotional reaction to my friend, then, depends on whether or not I judge her to be morally responsible, and the legitimacy of my reaction depends on whether she really *is* morally responsible.

Judging someone to be morally responsible also allows one to call them to account. This might mean that the person who is judged morally responsible is expected to offer an explanation for her actions, for example, or to make amends.[11]

In short, a multiplicity of social practices, emotions, attitudes and behaviours are provoked and apparently justified by judgments of moral responsibility, implicit or explicit. If moral responsibility seems to be so closely linked to these various practices, attitudes and so on, it would – to say the least – be very useful to know whether these really are justified, and if so by what. Otherwise, it would appear that an important aspect of our moral lives is without justification. It would therefore also be a strength of any theory of moral responsibility if it could offer a justification of this kind. This is not to say, however, that a successful theory needs to give a full justification of the *exact* practices, attitudes or intuitions that we find ourselves with. For one thing, it may be that these are not consistent with each other, and it is quite likely that

[11] See Watson (1996) and Oshana (2004) for discussions of this aspect of moral responsibility.

there is not a single set of generally shared intuitions, or an agreed set of appropriate practices or attitudes. It is also quite possible that common intuitions are mistaken. Nonetheless, a theory which delivered something that is at least recognisable as moral responsibility (as we know it), and which gave us good independent reason for any divergence from generally accepted norms (to the extent that such norms exist) would be a more readily acceptable one.

It is worth noting at this point that the majority of philosophical work on responsibility makes no attempt to offer anything like a comprehensive answer to the questions above. The very long-running debate about the compatibility or otherwise of free will with determinism, or of a lack of free will with moral responsibility, has not required philosophers to offer a full description of the conditions of application of moral responsibility, nor a justification of the various practices associated with it. Instead, it has to a great extent been confined to questions about the metaphysics of determinism, free will and control. For example, does a lack of alternative possibilities imply a lack of control over one's actions? Or, does determinism imply a lack of alternative possibilities? The aim of the game is to show either that determinism implies that nobody is morally responsible, or else to escape this charge. Thus, the debate is generally confined to the control condition of moral responsibility – if we can be shown to be in control, in a way that is compatible with determinism, then it follows that our practices of holding people morally responsible escape the very specific charge from determinism, and it is not necessary to look for a more general account of why control (or knowledge, or any other condition) is important in the first place, or of how our practices as a whole might be justified.

Nonetheless, relatively recently, this debate has inspired some philosophers to develop theories of moral responsibility that have attempted to answer these more general questions. In the remainder of this chapter, I will look at these types of theory in turn. I will first look at a consequentialist theory-type. This has the appeal of seeming to offer a relatively simple answer to the questions I

have posed above, but is unconvincing for some obvious reasons, which I will explain. From here, I will turn to two developments in the history of theories of moral responsibility which I see as making significant advances on what has gone before. By the end of the chapter, I will have developed an initial account of moral responsibility upon which I will begin in Chapter 2 to bring to bear some empirical findings about psychopathy.

1.6 Consequentialism

Consequentialist theories of morality are based on the desirability of maximising the total amount of some good in the world. It is fairly simple to see how a theory of moral responsibility could be constructed on the same basis. A theory of this type would allow us to discriminate between those who should be held morally responsible and those who should not, based on whether holding each person morally responsible would result in a maximal amount of the good. The practices, attitudes and emotions involved in holding people morally responsible would be justified by their role in maximising the good (and only those practices, attitudes and emotions that did in fact tend to maximise the good would be justified, so that we would have a principled way of choosing which practices, attitudes and emotions to maintain, and which to jettison). It is also certainly plausible to think that the practices, attitudes and emotions involved in holding people morally responsible do have consequences that we would generally think beneficial. Most of us do not like being blamed or censured, and we tend to avoid things that are likely to result in these outcomes. On a deeper level, being taken to task for our actions, attitudes and emotions provokes our conscience, forcing us to admit our own true motivations. Both of these processes are likely to have the effect of changing behaviour and, as long as those applying these sanctions are doing so in pursuit of socially beneficial goals, they are likely to have socially beneficial effects. Similarly, praising or expressing approval of socially beneficial behaviour is likely to inspire more of that behaviour in the future. It is also apparent that broadly

consequentialist thinking does enter our ordinary thought about moral responsibility to some extent. For example, if I were to harbour resentment towards someone for something that they did years ago, I could imagine a friend admonishing me on the basis that 'it does no good' to hold that person morally responsible for whatever it was, and I would take this as a valid criticism. For these reasons, consequentialist theories of moral responsibility do have some initial appeal.

Despite this, I do not think that the consequentialist approach to responsibility is the right one. In this section, I will attempt to show why this is so, first outlining a simple version of a consequentialist theory of moral responsibility, raising some obvious objections to it, and then describing an attempt by Richard B. Brandt[12] to respond to these objections with a more sophisticated theory. The views I will describe tend to be identified as utilitarian, but since it seems to me that these views could easily be adapted to fit other consequentialist theories by substitution of some other good in place of utility, I will talk in terms of 'consequentialism' and in terms of maximisation of 'the good', whatever that good may turn out to be.

It is worth noting first of all that I have so far been speaking in terms of the morality of *holding* people morally responsible, and that this is distinct from their *being* morally responsible. To construct a consequentialist theory of moral responsibility, we want to evaluate the consequences of being morally responsible. But it is not clear how our being morally responsible can have consequences, except through the actions that it inspires in others or ourselves: through the actions of those (including ourselves) who *hold* us responsible. If consequentialism is to have anything to say about what it is to *be* responsible, it will need to explain this in terms of when it is *right to be held* responsible: it

[12] Brandt (1969).

is right to hold someone responsible when to do so will maximise the good, and someone *is* responsible just when it is right to hold them responsible.

Brandt claims that the 'simple' consequentialist theory of moral responsibility can be found in the work of Henry Sidgwick, G.E. Moore, Hastings Rashdall and John Laird. We can summarise this general view thus: someone should be held morally responsible for some act, state of affairs, attitude, emotion or belief, if and only if to do so would maximise the good. J.J.C. Smart's 1961 paper 'Free Will, Praise and Blame'[13] is perhaps the most recent prominent publication to defend a version of this type of view. In this paper, Smart is concerned mainly with attacking 'libertarian' conceptions of free will, and the idea that moral responsibility is incompatible with determinism. He wants to show that 'threats and promises, punishments and rewards, the ascription of responsibility and the non-ascription of responsibility, have… a clear pragmatic justification which is quite consistent with a wholehearted belief in metaphysical determinism'.[14] For Smart, these practices are justified by their socially useful consequences. Praise is also justified in similar terms, because 'to praise a class of actions is to encourage people to do actions of that class. And utility of an action normally, but not always, corresponds to utility of praise of it'.[15] Blame, which Smart characterises as 'a grading plus an ascription of responsibility', is also justified based on its socially useful consequences. Thus, Smart develops a justification of a whole range of practices and judgments, including ascriptions of moral responsibility and many of the social practices that are typically thought to hang on these ascriptions, based on their consequences. This is an *act*-utilitarian

[13] Smart (1969).

[14] Ibid., p. 302.

[15] Ibid., p. 304.

account: the moral status of each act – an ascription of moral responsibility, an act of praising or blaming, etc. – is to be judged on its consequences.

In general, Smart provides two reasons why holding people responsible would tend to maximise the good. Firstly, many of the practices which depend on ascriptions of moral responsibility, including praising or blaming people for their acts, are informative: they spread information about the character of the person concerned and about their tendency to act in certain ways. It is plausible that the availability of this information to others would be socially useful in various ways, particularly in supporting the institution of trust.

Secondly, as I have already noted, blaming people deters them from acting in ways which are socially undesirable (in consequentialist terms, ways which tend not to maximise the good), and praising them encourages them to act in more desirable (good-maximising) ways. Similarly, holding someone to account can be thought of as a way of eliciting certain forms of behaviour from them: an undertaking to make amends, perhaps, or to behave differently in the future. Accountability is closely linked to moral responsibility: the act of holding someone to account for X seems inappropriate in the absence of a judgment that they are morally responsible for X.

To these we can add a third consideration: holding people responsible plausibly leads to desirable changes in behaviour from third parties. Censuring people for their wrong acts can be a deterrent to others, and praising can be an encouragement to others. I want to avoid the treatment being meted out to a wrongdoer, so I avoid repeating their crimes, or I act well in order to get some of the praise I have seen being given to others.

Unfortunately, Smart's act-utilitarian account does not stand up well under scrutiny. Firstly, it appears to lead to highly counter-intuitive consequences when applied to certain kinds of scenario. The problem is that there are likely to be situations where the act of holding someone morally responsible for an

act will plausibly be good-maximising for reasons that have nothing to do with the conduct or psychology of the person in question, and where holding them responsible simply seems enormously unjust. This will include scenarios where the person in question is not causally responsible for the act, and where we would expect there to be a link to causal responsibility (i.e. they do not look like the exceptions to this link noted earlier). These scenarios tend to rely on the third of the positive effects of holding people responsible listed above: the effect on third parties. It is easy to imagine, for example, a corrupt police officer presenting a consequentialist argument in favour of framing someone for a crime they did not commit. Perhaps there is little hope of finding the real suspect, and the crime is of such a nature that if it were seen to go unpunished this would be likely to lead to a slew of similar crimes committed by others. The negative consequences suffered by the person being framed could plausibly be outweighed by the positive effects for society generally, leading the consequentialist to the strongly counter-intuitive result that framing someone would be the right thing to do.

This kind of problem has, of course, afflicted utilitarianism generally, not just in the debate around moral responsibility. Many philosophers prefer rule-utilitarian theories in order to avoid these difficulties. Brandt's 'rule-utilitarian theory of excuses' is an attempt to derive a theory of moral responsibility from an overall rule-utilitarian theory of right action:

> 'If a rule-utilitarian affirms that an act is objectively right if it would be permitted by the moral code which will have the best consequences, then, since the best moral system will also contain a system of excuses, the utilitarian will presumably say that behaviour in some way out of line *should be excused* if its excuse

> would be provided for in the total moral system which would have the best consequences.'[16]

We can see how this proposal might be thought to avoid the problem presented by the framing case: while it might be true that a single act of framing someone might sometimes have consequences which are positive on balance, it is perhaps less likely that a *rule* in favour of framing people could ever maximise the good. This refinement therefore may be an improvement on the simple act utilitarian account, although it may also be vulnerable to challenges which affect rule-utilitarian theories in general.

The really insurmountable problem for any consequentialist theory of moral responsibility, however, whether based on acts or rules, is surely that they make responsibility reactive to the general effect on the world of attitudes towards the person, rather than on features of the person themselves. An example will show what I mean by this. Imagine a trustee of a charity finds out that, due to an accounting error - his own fault - he has lost a significant amount of money. Instead of owning up to this, he decides simply to pick a volunteer at random and accuse them of stealing from the collection boxes, and to back up this accusation with fabricated evidence. Now, a consequentialist might argue that the trustee's doing this would not lead to good consequences in general, and therefore this case can be accommodated under their theory. But there are problems with this response. Firstly, it is certainly not obvious that the consequences would be as suggested. Perhaps, for example, making the accusation will enable the charity to recover the money through insurance, meaning that a great many people can benefit from the charity's work. This might plausibly outweigh the bad consequences for the volunteer. More generally, at the beginning of this section I suggested that, in order to have anything to say about whether or not a given person *is* morally responsible, they

---

[16] Brandt (1969), p. 350. (italics author's own).

would need to make this depend on the rightness of that person being *held* morally responsible, something which in turn would be judged on the likely consequences of doing so. To return to our example, is it really plausible to claim that the negative consequences of holding the charity volunteer responsible (or of a rule in favour of holding people responsible in circumstances like this) explain why the person is not morally responsible? Surely the kind of reason we should be looking for to explain this should take as central the fact that the volunteer simply did not steal the money, and perhaps also the injustice *to her* of holding her morally responsible for the loss. On the consequentialist theory under discussion, these facts are only relevant to the explanation of moral responsibility insofar as they bear on the consequences of a responsibility attribution – they have only secondary relevance.

A theory which makes ascriptions of moral responsibility primarily a matter of weighing the future consequences for the human race in general, then, is ill-suited as a means of determining when people are morally responsible and when they are not, not only because it is likely to lead to some implausible results, but more importantly because it appeals to the wrong kind of reason in doing so. When we say that someone is morally responsible for something, it is most natural to think of this as a descriptive statement about them which should be responsive primarily to facts about their actions, character and circumstances. Instead, the consequentialist approach makes judgments of moral responsibility contingent on the general effect of our making those judgments. Because of this, it risks justifying judgments of moral responsibility which depart radically from the pattern of such judgments which I sketched in Section 1.2. Even when it does not do this, it provides an explanation which does not call on the kinds of consideration that would normally enter into explanations of this kind. A theory which better fits these explanations would surely be preferable.

1.7 The reactive attitudes

Smart and Brandt's utilitarian accounts sought to explain and justify a range of practices and attitudes that are legitimised by moral responsibility: centrally praise and blame, but also punishment and rewards, etc. On the other hand, as I noted at the start of the chapter, it is not only these practices and attitudes that are closely linked to judgments of responsibility. There are also a range of emotional attitudes that only seem appropriate if we think of the people at whom they are directed as morally responsible. The strand in philosophy in which these attitudes are taken seriously when talking about moral responsibility begins with P.F. Strawson's 1962 lecture, 'Freedom and Resentment'.[17] Strawson's account was formulated as a compatibilist response to the incompatibilist position which states that moral responsibility depends on libertarian free will, and is therefore inconsistent with causal determinism. This response results in a genuine theory of moral responsibility, which deserves to be considered on its own merits. 'Freedom and Resentment' is rich and provocative, and Strawson's argument in it is complex and open to interpretation. In broad strokes, it proceeds as follows.

Holding people morally responsible, for Strawson, is not a simple, unitary practice, but is inherent in a complex, variable set of attitudes which include praise and blame, but also resentment, gratitude, forgiveness, love and hurt feelings, as well as self-directed attitudes such as pride, guilt and shame. These attitudes, which Strawson calls the 'reactive attitudes', are 'something we are given with the fact of human society'[18], and as such are a basic, inescapable part of our nature, though we are capable of withholding them towards specific kinds of people, or in specific circumstances. We do this firstly when actions have been performed through ignorance, compulsion, lack of choice, etc. In

[17] Strawson (2008).

[18] Ibid., p. 25.

such circumstances, we do not 'view the agent as one in respect of whom these attitudes are in any way inappropriate'[19]. Rather, we view the specific action as one in reaction to which such attitudes held towards the agent would be inappropriate. Secondly, we sometimes do withhold reactive attitudes towards the agent as a whole, but only in unusual circumstances, such as when they are under abnormal stress, or under hypnotic suggestion, and are temporarily 'not themselves' in some way, or because they are abnormal in a relevant way (for example mentally ill, or a child). Finally, we are able to withhold the reactive attitudes voluntarily and temporarily towards someone, 'as a refuge, say, from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity'[20], and not because of any fact about the person who is the object of the attitudes, or because of any fact about any action which they have performed.

Having set out these three categories of situation in which we are capable of suspending the reactive attitudes, Strawson goes on to claim that no thesis which applies to people indiscriminately – including the incompatibilist thesis – either could or should lead us to withhold reactive attitudes in any of these ways. The incompatibilist thesis could not imply that all human interactions would fall into the first category (ignorance, compulsion, lack of choice, etc.) because we ought to be looking for a justification for suspending reactive attitudes towards the agent, not towards the act. Nor could any thesis (including the incompatibilist thesis) ever show that all people are always 'not themselves', or that all agents, at all times, are abnormal (the second category). This leaves only the voluntary suspension of reactive attitudes, which Strawson believes is 'practically inconceivable'[21] as a long-term, general strategy, because

---

[19] Ibid., p. 8.

[20] Ibid., p. 10.

[21] Ibid., p. 12.

of the strain of withholding reactive attitudes in this way, and the way in which attempting to do so would impoverish our lives. Further to this, although Strawson's overall approach is in some sense to eschew discussion of the rationality of responsibility attributions – he claims that the full set of reactive attitudes 'as a whole … neither calls for, nor permits, an external "rational" justification'[22] – he apparently does believe that it is rational to hold the reactive attitudes in broadly the circumstances in which it is natural to do so. It could not be rational, according to Strawson, to behave in a way that is so unnatural as to be practically impossible and which, were we to attempt it, would impoverish our interpersonal relationships to the point where they would become unbearable: 'we could choose rationally only in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment; and the truth or falsity of a general thesis of determinism would not bear on the rationality of this choice.'[23]

There are a number of claims put forward in this argument, some empirical, some conceptual and some normative, which would benefit from some unpacking and consideration.

One aspect of Strawson's ideas that can be somewhat difficult to disentangle is the exact nature of the relationship between reactive attitudes and moral responsibility. In fact, the word 'responsibility' is significantly absent from the main part of Strawson's discussion of the reactive attitudes, and only enters the text when Strawson relates his own position back to the debate between incompatibilists and compatibilists. Nowhere is there anything that looks like a definition of moral responsibility. It is therefore tempting to interpret Strawson as rejecting the idea that such a definition is required. This fits with

[22] Ibid., p. 25.

[23] Ibid., p. 14.

his dismissal, later in the lecture, of 'over-intellectualising'[24] approaches to the debate. Taking this thought further, we may wish to read Strawson as thinking of moral responsibility as an 'umbrella term': a term under which are sheltered the objects of a range of attitudes, rather than a single, unitary concept which could be given a precise definition.

On the other hand, it cannot be the case that reactive attitudes are always appropriate. Otherwise we would be left with the absurd claim that no-one is ever inappropriately angry, resentful or grateful. We therefore need a way of deciding when the reactive attitudes would be appropriate, and indeed, the idea of appropriateness is central to Strawson's discussion. I stated earlier that it would be an advantage of a theory of moral responsibility if it could explain why some people are morally responsible and others are not, and the notion of appropriateness plays a key role in Strawson's attempt to provide such an explanation.

In fact, the idea of appropriateness takes up roughly the same conceptual space in Strawson's discussion as that of moral responsibility takes up in the work of other philosophers. Strawson no doubt has his reasons for avoiding use of the term 'moral responsibility'; he is attempting to reframe the debate and to get away from the metaphysical controversies that have traditionally formed the battleground of debate around moral responsibility. Better to avoid a term which might come loaded with too many preconceptions. Nonetheless, there can be little doubt that, at a basic level, the idea at stake in Strawson's account is moral responsibility. The position occupied by moral responsibility in previous philosophical enquiry has been as the quality or set of qualities which render praise and blame appropriate. Broadening this out to include the other reactive attitudes does not remove the necessity for determining what these qualities are, and indeed Strawson attempts to do so. Moral responsibility forms

[24] Ibid., p. 25.

one of the conditions which together legitimise – i.e. render appropriate – not only praise and blame, but the full range of reactive attitudes as described by Strawson. This is, I think, the natural way to think about moral responsibility: if I resent someone for an action, there is implicit in my attitude of resentment a judgment that the person is morally responsible for the action. If it turns out that they are *not* morally responsible, then my implicit judgment is incorrect, and my resentment is therefore inappropriate. For example, I can resent you for pushing in front of me in the queue, but if it turns out that you had no idea that there *was* a queue then it is no longer appropriate to resent you. For Strawson, this is because your actions did not manifest a bad quality of will, but this is to say that moral responsibility is a question of qualities of will. It remains true that it is inappropriate to resent you because you were not morally responsible in this case.

I have said that Strawson has 'broadened out' the discussion from focusing solely on praise and blame to include also the reactive attitudes. It is worth briefly considering how we should understand the relation of the reactive attitudes to praise and blame. Consideration of this question, I think, reveals an interesting asymmetry between these two ideas. It is natural to think of blame as inherent in those reactive attitudes which involve negative assessment of their object, and difficult to think of counter-examples to this thought. It is difficult to see how I could resent someone for an action, have hurt feelings in reaction to their performance of the action, be indignant towards them, and so on, if I did not think they were to blame for that action. On the other hand, it does not seem to me that *praise* is inherent in the *positive* reactive attitudes in the same way. Imagine my parents sent me to a private school – they worked hard and made sacrifices to do so, and as a result I received a better education than I would have otherwise. As an adult, I have come to believe that all education should be publicly funded, and that those who send their children to private schools are perpetuating a social evil. It seems to me perfectly coherent

that I could be grateful to my parents for sending me to a private school, while simultaneously not thinking that their doing so was a praiseworthy act, or that they were praiseworthy for it. Another example: Strawson includes certain kinds of interpersonal love in the set of reactive attitudes. Now it seems to me perfectly possible that I could see someone I care about doing something that I do not think is particularly praiseworthy, and yet love them for it, perhaps because it is so perfectly expressive of their personality – the personality that I love.

The natural way to think of the relationship between the reactive attitudes and blame and praise, it seems to me, is simply to think of both blame and praise as examples of reactive attitudes. Moral responsibility, then, is a necessary condition of those attitudes' being appropriate (of the act, etc. being blameworthy or praiseworthy) in the same way as it is of the other reactive attitudes. This is simply to extend the analysis of the relationship between moral responsibility, praiseworthiness and blameworthiness I set out earlier. We might say, for example, that someone who is morally responsible for an unworthy act, is not only therefore blameworthy, but also 'resentment-worthy' in the sense that an attitude of resentment towards them may also be appropriate, depending of course on the nature of the act and the circumstances of the person doing the resenting.

What does Strawson have to say about the question of when the reactive attitudes are appropriate, and when they are not? Strawson begins the part of the essay that deals with this question by talking about the types of situation in which we typically withhold reactive attitudes, describing categories of cases in which this typically occurs. The categories that Strawson gives are, I think, open to question. Strawson writes in terms of a distinction between withholding reactive attitudes towards the *act*, and towards the *person*, and includes cases of mental illness in the second category. This is similar to the distinction made by Nicole Vincent between 'capacity responsibility' and what I have called moral

responsibility. I noted earlier that this distinction is not a simple one, and it is not entirely clear what to make of Strawson's reading of this. Strawson presumably cannot, for example, mean that suspending reactive attitudes towards the person involves suspending those attitudes in relation to every act by that person, because this is rarely what happens in the types of case described. It is only in very extreme cases of mental illness, for example, that we suspend *all* reactive attitudes in this way. Surely, in most cases, we suspend reactive attitudes towards the mentally ill person only with regard to those actions which we can attribute to the mental illness in some way. To adapt the example I used earlier, if someone suffers from paranoid delusions, it would not be appropriate to resent his insulting me if I know that one of his delusions is that I am a persecutor. On the other hand, if none of his delusions apply to me at all, I might have a different attitude. He might simply not like me, and if so, I might be justified in resenting him. How, in general, would one go about deciding whether to take personally an insult from someone with a psychological or neurological disorder? One might look for evidence that his insult was caused by some delusion that denied him full knowledge of what he was doing (e.g. he is paranoid and thought he was insulting his nemesis; in fact, he was insulting his friend). Alternatively, one might look for evidence that he lacks control over his actions in some relevant way (e.g. he has a form of Tourette's syndrome which manifests in coprolalia – the condition which causes involuntary swearing). Either way, we would be looking at conditions relevant to the *act*, and not the person generally. The capacities of the person are only relevant insofar as they bear on the person's responsibility for the act.

Related to this is another problem with what we might call Strawson's 'categorisation strategy': his strategy of setting out categories of situation in which the reactive attitudes are withheld, and arguing that the truth of determinism would not render it the case that all situations fall into one or other of these categories. Strawson claims that the withholding of reactive attitudes

for incompatibilist reasons could not fall into the category of normal practice which has to do with particular acts done under compulsion, ignorance and so on, since the incompatibilist ought to be looking for a justification for suspending reactive attitudes towards the *agent*, not towards the act. But in fact, the majority of incompatibilists do focus their arguments on acts rather than agents, apparently with good reason. Causal determinism is a thesis about the nature of action, and incompatibilism can be defined broadly as the claim that causally determined action is not compatible with someone's being morally responsible for that action. Incompatibilists typically proceed by claiming that the type of control necessary for moral responsibility is precisely control over individual actions, not general capacities of control, the latter being more commonly appealed to by compatibilists.[25] If all action can be shown to be outside the realm of moral responsibility in this way, then the incompatibilist has won. There is no need to talk about general capacities of control at all.

I think, then, that there are problems with Strawson's categorisation strategy, but it is possible that these could be addressed by changing the specific categories used. If not, and there is a deeper problem as suggested above, this applies to Strawson's argument against incompatibilism, and not to the ability of his theory to account for moral responsibility outside the boundaries of this debate. A more relevant question for our purposes is *why* Strawson thinks it is in these specific categories that we withhold the reactive attitudes. This is where the notion of appropriateness comes into play. Strawson's explanation is that some actions are expressive of 'goodwill, its absence or its opposite', while others are not:

> If someone treads on my hand accidentally, while trying to help me, the pain may be no less acute than if he treads on it in contemptuous disregard of my existence or with a malevolent

[25] See the discussion of Jay Walllace and Fischer and Ravizza in the next section.

> wish to injure me. But I shall generally feel in the second case a kind and degree of resentment that I shall not feel in the first. If someone's actions help me to some benefit I desire, then I am benefited in any case; but if he intended them so to benefit me because of his general goodwill towards me, I shall reasonably feel a gratitude which I should not feel at all if the benefit was an incidental consequence, unintended or even regretted by him, of some plan of action with a different aim.[26]

In other words, part of what separates those actions that are appropriate targets of reactive attitudes from those that are not is that the actions in question are expressive of some quality of will on the part of the agent: either goodwill, ill will, or an absence of the ordinary level of regard that we demand from people as part of normal human relationships.

I have interpreted Strawson's account in terms of moral responsibility by claiming that moral responsibility is one of the conditions that renders the reactive attitudes appropriate in certain cases, with another condition being that the agent has performed an act (or held an attitude, etc.) that is either morally worthy or unworthy. Might the expression of a certain quality of will (or its absence) though an act (attitude, etc.), as described by Strawson, be what makes it the case that the person is morally responsible for the act (attitude, etc.)? This, it seems to me, is a promising suggestion and, as an explanation of why some people in some cases are morally responsible while others are not, it has some appeal.

One advantage of this suggestion is that it provides a ready explanation for why we hold people morally responsible not only for actions, but also for emotions and attitudes. Good and ill will are themselves attitudes, and other attitudes

---

[26] Strawson (2008), p. 6.

can be partly constituted by good or ill will. Emotions too can be expressive of attitudes towards others. For example, we might think it praiseworthy that Patti frequently feels compassion for her friends when they are undergoing some hardship or other, implying that she is morally responsible for her emotion, because it is expressive of a general attitude of goodwill towards her friends.

The way to test the adequacy of 'qualities of will' as an explanation of why we hold some people responsible and others not would be to look for cases in which we hold people responsible regardless of their quality of will. A potential category of such cases would be one we have already discussed, namely the category of cases where the action, attitude or emotion concerned is neither morally worthy nor unworthy. In the case in which I buy a carton of milk from a shop, this act expresses no quality of will, either good or bad, on my part, and it does not express a lack of good will towards anybody which they might have expected from me. Therefore, if being morally responsible is purely a matter of qualities of will, we must conclude that I am not morally responsible for buying the milk. As I noted above, this is probably an acceptable conclusion, after all, there is something strange about saying I am *morally* responsible for something that has no moral dimension at all. On the other hand, it does mean that the suggested account has nothing at all to say about the difference between this case and the alternative case in which I accidentally buy mislabelled orange juice. This type of case so closely mirrors cases having to do with moral responsibility – shares so many of the important features of such cases – that it would be somewhat surprising if there were no theoretical connection between them.

As well as morally neutral cases, there is also a category of cases in which we might want to hold someone morally responsible for *good* acts, states of affairs, etc., not based on any particular quality of will on their part, or on an absence of a quality of will which they have an obligation to show. As an example, imagine an entrepreneur who creates a business for purely selfish reasons, but

as a result creates a number of jobs, which have beneficial consequences for those employed in them, and perhaps for the local economy. It seems to me that the entrepreneur is morally responsible for creating the jobs, even though she may not have any particular feeling of goodwill towards her employees. If this is right, then cases of this kind apparently cannot be explained by an account purely based on qualities of will.

Because of the two types of case I have discussed, I think the 'qualities of will' account is incomplete. However, it is also worth noting that consideration of how this would apply to psychopaths calls into question the intuition upon which the 'qualities of will' account is based. The problem is that in the vast majority of cases, we can assume that the person in question is perfectly capable of understanding that other people are *due* some degree of good will, or at least an absence of ill will. Because of this, it makes intuitive sense to hold them morally responsible when they fail to exhibit this. However, it is not at all clear that psychopaths do understand that people are due particular qualities of will. In turn, it is not at all clear that they can be held morally responsible when they fail to exhibit those qualities of will. Again, what seems like a solid intuition in the vast majority of cases becomes harder to discern when applied to psychopaths.

In addition to explaining why some people, in some circumstances, are morally responsible for some actions, attitudes and emotions while others are not, I suggested earlier in the chapter that it would be an advantage of a theory of moral responsibility if it could provide a justification of the practices, attitudes and emotions involved in holding people responsible. Strawson's argument in this area is interesting and arises from his general project of arguing against incompatibilism. It is expressed in answer to 'a question about the rational justification of ordinary inter-personal attitudes in general.' Is it rational to hold such attitudes?

> To this I shall reply, first, that such a question could seem real only to one who had utterly failed to grasp the purport of… the fact of our natural human commitment to ordinary inter-personal attitudes. This commitment is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework. And I shall reply, second, that if we could imagine what we cannot have, viz, a choice in this matter, then we could choose rationally only in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment; and the truth or falsity of a general thesis of determinism would not bear on the rationality of this choice.[27]

Strawson makes two claims here. Firstly, that suspending the reactive attitudes is not available to us as a real option because they are so deeply embedded in 'the general framework of human life', and secondly that it could never be rational to suspend the reactive attitudes because of a general thesis such as that of incompatibilism, because to do so would all but completely destroy our interpersonal and social relationships, and would make life unbearable.

There can be little doubt that a consequence of the wholesale abandonment of the reactive attitudes would be a profound impoverishment of human relationships. In this key passage, Strawson characterises the suspension of reactive attitudes as 'taking the objective attitude' towards someone:

> To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of senses, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of;

---

[27] Ibid., p. 14.

> to be managed or handled or cured or trained; perhaps simply to be avoided.... The objective attitude may be emotionally toned in many ways, but not in all ways: it may include repulsion or fear, it may include pity or even love, though not all kinds of love. But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships; it cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other. If your attitude towards someone is wholly objective, then though you may fight him, you cannot quarrel with him, and though you may talk to him, even negotiate with him, you cannot reason with him. You can at most pretend to quarrel, or to reason, with him.[28]

Strawson appeals to our intuitions to support his claim that for this to become the normal way of treating human beings would greatly impoverish interpersonal relationships, and this is an intuition that I certainly share. A world in which the 'objective attitude' described above became the norm, in which the full range of emotional connections between people, as well as the possibility of reasoning with each other in order to influence each other's behaviour, was abandoned, would, I think, be clearly much worse than the world we have now.

However, it is one thing to show that the consequences of withholding the reactive attitudes would be very bad, and even that to do so is practically impossible, and another to establish that to do so is not rational. There is nothing inconsistent about the conclusion, catastrophic though it may be, that we have no choice but to act irrationally. Perhaps Strawson is right that our commitment to the reactive attitudes (and thus, according to my interpretation

---

[28] Ibid., p. 10.

of Strawson at least, to the practice of holding people morally responsible) is so deep as to make abandoning them a practical impossibility. Unless Strawson is also right about the type of rational assessment which is appropriate here, they may nonetheless be rationally unjustified.

The question of rationality is key to Strawson's contribution to the compatibilism/incompatibilism debate. However, leaving this debate aside, we can ask how well Strawson's claim about rationality functions as a general justification of the practices and attitudes involved in holding people morally responsible.

Strawson implicitly recommends assessing the rationality of holding a reactive attitude according to a standard of rationality that has to do with the consequences of holding an attitude. This need not imply a maximising approach to consequences of the type found in consequentialist theories of responsibility; Strawson is not committed to the idea that the rational attitude is always the one which has the best consequences overall. It does however imply that the likely consequences of holding an attitude can be decisive in determining the rationality or otherwise of holding that attitude. But there are reasons to doubt that we should primarily be looking at consequences in order to answer these questions. As A.J. Ayer comments in his reply to Strawson:

> There is another sense of 'rational' in which the rationality of an attitude is measured not by the probable consequences of adopting it but by the standing of the beliefs that enter into it. In this sense, an attitude is irrational if it rests on a belief which we have no good reason for accepting.[29]

One way in which it would be irrational in Ayer's sense to adopt a reactive attitude toward an agent with respect to an act, would be if it were the case that

---

[29] Ayer (1980), p. 11.

reactive attitudes in fact rest on beliefs about the type of control possessed by the agents who are their objects, and in this case the agent did not possess that type of control. If, further, no agents possess the relevant type of control with respect to any act (which the incompatibility thesis purports to show), then reactive attitudes would always be irrational in Ayer's sense. To my mind, there is nothing in 'Freedom and Resentment' which should convince us that judgment of consequences is the *only* standard of rationality that should be applied to reactive attitudes or to the practices involved in holding people morally responsible.

However, this is a criticism that only really has teeth if one believes some version of the incompatibilist thesis, that moral responsibility is incompatible with causal determinism, or with some other metaphysical thesis that one takes to be true. I am agnostic about this question, which is tangential to my aims. Leaving this aside, then, Strawson's account does offer a justification of the practices and attitudes involved in holding people responsible, based on the psychological strain and impoverishment of human relationships which would result from abandoning them.

However, this account is starting to look suspiciously consequentialist. Should we be worried about this? It is certainly true that Strawson's justification of the reactive attitudes, taken as a whole, at least as I have interpreted it, rests on the consequences of holding them, and in this sense his is a consequentialist theory. However, his theory is different in several respects from the somewhat cruder consequentialist theories I considered in the previous section. For Strawson, the consequentialist justification applies to the reactive attitudes taken as a whole, and not to each individual judgment of responsibility. It might be thought that this simply makes Strawson's account similar to rule utilitarian theories as opposed to act utilitarian theories. Can Strawson's account escape the criticisms of these theories that I identified in the previous section?

One of these was that applying a consequentialist account is likely to lead to situations in which very counter-intuitive instances of holding someone responsible would appear to be good-maximising, and therefore would be the right thing to do. The case of the falsely accused charity volunteer is an example of this. I argued that even the rule utilitarian has problems here, since they need to give an account of why it would not be preferable in situations like that of the charity volunteer to make exceptions to the rule, since it would seem that utility would be best served by doing so. However, in contrast to the consequentialist accounts examined earlier, there are considerations inherent in Strawson's account which are enough to suggest that it would be wrong even to be thinking in these terms in the first place. It would be counter to the whole substance of Strawson's justification of the reactive attitudes if considerations of consequences were to form a central part of our deliberative process in applying the reactive attitudes.

There are two test cases here. The first is one in which, if we were to deliberate about consequences, we would decide, on the grounds of the general consequences of doing so, to withhold a reactive attitude, even though a reactive attitude would be natural and fitting. The second is the mirror image of this: a case in which we would be led by considerations of consequences to try to hold a reactive attitude when to do so would be wholly *un*natural. What the picture offered by Strawson shows is that to do this would be a matter of psychological strain on our own part, and would lead to strained and unnatural social relations with others. These in themselves would be highly undesirable consequences.

It might be argued that our own psychological strain, along with considerations such as the resentment of the falsely accused, would still in some cases be outweighed by the positive consequences of holding someone morally responsible when they are not, or of not holding someone morally responsible when they are. In these cases, the argument would go, we would do better to

think like consequentialists. However, in order to identify these rare cases, we would need continually to subject our reactive attitudes to the same kind of assessment in terms of their general consequences, something that would take us very far from our ordinary deliberative process. Seeing the issue through a Strawsonian lens allows us to see just how far it would take us. It would mean, in effect, trying out the 'objective attitude' in every social interaction, in order to see whether the consequences of holding this attitude would be socially beneficial. Once we are doing this, the psychological strain and impoverishment of relationships that Strawson warns of is already largely in place. Strawson's account therefore shows why the overall beneficial consequences to which he appeals as a general justification should not enter the deliberative process at the level of individual judgments of responsibility.

However, the deeper point was not about whether we should deliberate in a consequentialist manner, but rather about what it is that, as a matter of fact, makes someone morally responsible or not. As I noted above, consequentialist theories must account for this in terms of when it is right to hold someone morally responsible. The point about psychological strain and impoverished relationships suggests that we should hold people morally responsible when it is natural to do so, but this is something of an unsatisfactory answer. A criticism I made of consequentialist theories was that they make moral responsibility depend on considerations of general consequences when they should, rather, directly track facts about the person concerned and their actions. This is a criticism that holds even when we consider cases – unlike the charity volunteer case – in which the consequentialist verdict is in line with our intuitions. Even in these cases, it should not be because of considerations of general consequences that the person is morally responsible, but because of facts about them and their actions.

The best answer to this that one can derive from Strawson's account is, as I have argued, to make moral responsibility depend upon the quality of will expressed

in the actions concerned. Someone who steals from someone else thereby expresses ill will towards them, whereas the wrongly accused charity collector has expressed no ill will, and therefore cannot be morally responsible. On this model, facts about moral responsibility do indeed track facts about the person and their actions. However, as I have argued, the 'qualities of will' idea gives an incomplete account of what it is to be morally responsible.

In addition, it would be preferable if the central concepts in a theory, whether or not these relate to qualities of will, could be both part of what justifies holding people responsible and a way of distinguishing between those who are morally responsible and those who are not. It may well be 'natural' to hold people responsible when their actions express particular qualities of will, and it may be true that the consequences of abandoning this practice would, because of this, be bad enough that we are justified in continuing with it. However, a justification based on the naturalness of holding people responsible can only justify our holding people responsible when it is natural to do so, and it is not yet clear why it would (usually) be more natural to hold people responsible when their actions express particular qualities of will. A preferable justification would be one which makes it clearer why we are justified in holding people responsible precisely when we do. I will now turn to a strand in philosophical thinking about responsibility which I think can provide this kind of justification.

1.8 Responsiveness to Reasons

Beginning after Strawson, and partly inspired by Strawson, there have been a number of attempts to elucidate moral responsibility in terms of the ability to recognise and respond to reasons. The first complete theory of this kind, and one of the most influential, is that of R. Jay Wallace. Wallace's theory is presented as a development of Strawson's ideas, and Wallace follows Strawson by beginning with a discussion of what it is to *hold* someone morally responsible, which he also understands in terms of the reactive attitudes. He goes beyond Strawson however by presenting a more complete analysis of what

it is to *be* morally responsible, which he understands in terms of the agent's ability to grasp, respond to, and control her behaviour in the light of certain kinds of reason. By making this the central condition of moral responsibility, Wallace hopes to avoid the contention that moral responsibility requires the kind of freedom that might be thought to be incompatible with causal determinism. For Wallace, 'the "can" that matters in moral responsibility is not the "can" of alternate possibilities, or strong freedom of the will, but the "can" of general rational power."[30] Whether he is ultimately successful in this project does not need to concern us here, however. Instead, I will sketch Wallace's account simply as an attempt to explain what it is to be morally responsible, and particularly as an attempt to fill the gaps left by Strawson's account.

Broadly, Wallace's argument proceeds like this. Firstly, building on Strawson, he contends that,

> to hold someone to… an expectation is to be susceptible to the reactive emotions in the case that the expectation is breached, or to believe that the reactive emotions would be appropriate in that case…. To hold a person morally responsible… is to hold the person to moral expectations that one accepts.[31]

The conditions for someone's being an appropriate target for reactive attitudes should, in Wallace's view, be subject to a *normative* analysis; we should be looking for whatever conditions make it *fair* to adopt these attitudes towards them. Now, it cannot be fair to expect someone to fulfil an obligation unless they can firstly recognise that an obligation exists, and secondly control their behaviour in such a way that ensures that they fulfil the obligation. Since moral obligations are, for Wallace, a matter of there being moral reasons in favour of

---

[30] Wallace (1994), pp. 7-8.

[31] Ibid., p. 51.

or against our performing certain actions, the kind of control that someone must have in order to be morally responsible is bound up in 'the powers of reflective self-control: (1) the power to grasp and apply moral reasons, and (2) the power to control and regulate [one's] behaviour by the light of such reasons'.[32]

A particularly interesting aspect of Wallace's analysis is his insistence on tying the practices of holding people responsible closely to the existence of moral obligations and the person's success or failure in fulfilling these. While this approach promises to make sense of cases in which someone is held morally responsible for an action by which they have, apparently, broken an obligation, there are many cases which are not like this. One such type of case is that of omissions – cases where someone has apparently broken an obligation not by performing a certain action (or by bringing about a certain state of affairs, or by having a certain attitude, emotion or belief) but by *not* performing a certain action which they had an obligation to perform (or by not bringing about a state of affairs that they had an obligation to bring about, etc.). In fact, it is an advantage of Wallace's account that he can make sense just as easily of responsibility for omissions as for actions, since an act of culpable omission is still a failure to fulfil an obligation, and can be assessed in the same way as any other such failure.

On the other hand, responsibility for *good* acts is a little harder to bring under Wallace's general scheme, especially in those cases where the act in question does not constitute the fulfilment of any obligation. In such cases the usual conditions of moral responsibility, whatever these are, are presumably fulfilled, and the agent can surely be held morally responsible for the act. To deal with this difficulty, Wallace presents an overall explanation of moral responsibility for morally worthy acts which is conditional and dispositional: to hold a person

---

[32] Ibid., p. 157.

A morally responsible for a morally worthy act X is, firstly, to believe that A is the 'sort of agent' to whom the reactive emotions would be an appropriate response to their nonfulfillment of an act which they had an obligation (which we accept) to perform, and secondly, to believe that A 'has done something that meets or exceeds the moral obligations one accepts'[33].

The two categories of cases which I claimed in the previous section cause problems for an account based purely on qualities of will are 1) cases where people are morally responsible for positive states of affairs which are unintended by-products of their actions and 2) actions (or states of affairs, emotions, attitudes) which are neither morally worthy nor unworthy, i.e. they are morally neutral. How well can Wallace's account cope with these cases?

Cases in which people are morally responsible for positive states of affairs which are unintended by-products of their actions (the selfish entrepreneur who is morally responsible for creating jobs, for example) would presumably need to be brought under the same kind of analysis as Wallace uses for all morally worthy acts. The selfish entrepreneur is, by hypothesis, the 'sort of agent' to whom the reactive emotions would be an appropriate response to their nonfulfillment of an act which they had an obligation (which we accept) to perform. Has she done something that meets or exceeds the moral obligations we accept? This phrase requires interpretation. One might naturally interpret the idea of 'exceeding obligations' as meaning the performing of an action which one has an obligation to perform, to a greater degree than that to which one has an obligation to perform it. For example, if I work through the weekend to do a particularly good job on a piece of work, I could be said to have exceeded my obligations to my employer. I had an obligation to do the work to an acceptable level, within my contracted hours, but I have acted beyond this obligation in terms of quality and time. However, this cannot be what Wallace is referring to

---

[33] Ibid., p. 71.

in his account of morally worthy acts, and it cannot be the requirement in the types of case I have been discussing. Both of these types of case frequently involve actions which have nothing to do with any really existing obligations of any kind.

Perhaps a better way to interpret Wallace's idea of exceeding obligations, then, is simply as referring to the performance of morally worthy acts which one has no obligation to perform. The trouble with this is that there are some such acts for which one is *not* morally responsible, despite being the 'sort of agent' to whom the reactive emotions would be an appropriate response to their nonfulfillment of an act which they had an obligation (which we accept) to perform.

As an example, imagine a slapdash chef serves undercooked shellfish to a customer. By sheer coincidence, the customer turns out to be a terrorist who is planning to hijack a plane the following day. Because the terrorist is incapacitated by severe food poisoning, the hijacking never takes place, and many innocent lives are saved.

What separates the slapdash chef from the selfish entrepreneur, who creates jobs purely by accident when her sole motivation is enriching herself? One answer might be that job-creation is a predictable outcome of entrepreneurship. It is the kind of consideration to which the entrepreneur would be entitled to appeal in order to justify her activities. It supplies a good *reason* to start a business and, even if it is not a reason that influenced this particular entrepreneur, it is a reason of which she can be expected to be – and presumably is - aware. Preventing terrorist attacks, however, is not a predictable outcome of undercooking seafood (when one does not know that it is going to be served to a terrorist), and therefore it does not supply a reason of which the chef could be expected to be aware. Moral responsibility for a morally worthy act, therefore, would appear to rely on being responsive to the specific reasons that

are generated by that act – the considerations that make it morally worthy – even if it does not require that those reasons actually influence the agent.[34]

More obviously there are, again, also some acts which are neither morally worthy nor unworthy, but for which we are nonetheless morally responsible. This is the second category of actions which I mentioned in the previous section. I have no obligation to buy milk from the shop, and am not breaking any obligation by doing so. Again, as with the examples above, what separates this from morally neutral actions for which I am *not* morally responsible is plausibly my awareness and receptivity to the reasons that apply to this particular action, though in this type of case they will not be moral reasons. (A relevant reason in the milk buying case might be the fact that I have no milk to pour on my cornflakes.) If the 'milk' in the shop is actually mislabelled orange juice, I have not, through buying it, exhibited awareness of or receptivity to any of the reasons which might have spoken in favour of buying orange juice at that moment, and this is why I am not morally responsible for this action.

There is a possible confusion raised by cases such as the 'slapdash chef' and 'milk' cases which it will be helpful to clear up here. As we saw, Wallace's description of responsiveness to reasons as a condition of moral responsibility has it consisting in 'the powers of reflective self-control: (1) the power to grasp and apply moral reasons, and (2) the power to control and regulate [one's] behaviour by the light of such reasons'.[35] Yet the slapdash chef does not lack a

---

[34] I should note that this example, and much of the discussion to follow in this chapter, assumes that one's reasons are, broadly, the facts that bear on one's choices, as opposed to what one *takes* to be the facts that bear on one's choices. Thus one can believe oneself to have reasons that one does not in fact have, and one can be unaware of reasons that one does have. This view, which I endorse, is subject to some controversy, but defending it would require a lengthy diversion for which I do not have the space. For an opposing view see Gibbons (2010).

[35] Wallace (1994), p. 157.

general rational power of this kind, and nor do I in buying mislabelled orange juice.

Nonetheless, it is clear that Wallace's general picture is supposed to make sense of cases in which the agent acts through ignorance, as well as those in which she acts through a lack of control. This is shown by Wallace's 'typology of excuses', one of the categories in which is 'inadvertence, mistake or accident':

> Suppose I do something that happens to be of kind *x*. The first class of excuses [having to do with inadvertence, mistake or accident] defeats a presumption that I did *x* intentionally, by showing that I did not know that I would be doing something of kind *x* at all when I chose to do whatever it was that turned out to be of kind *x*. Thus, if I tread on *s*'s hand inadvertently, while walking to the refrigerator to get a beer, then I must not have anticipated that I would be treading on *s*'s hand when I made the choice to get a beer. If I tread on *s*'s hand by mistake, I may have known that I would be treading on a hand, but not that it was *s*'s hand that I would be treading on (perhaps I took the hand for *p*'s, where *p* was a thief trying to reach for the weapon on the floor). And if I tread on *s*'s hand accidentally (say, while trying to stomp out the flames), I may know that I am treading on a hand at the time when my treading motion occurs. But again, I will generally have lacked the foreknowledge that I would be treading on *s*'s hand, at the time when I made the choice that led to the treading activity.[36]

Wallace's three examples illustrate types of case in which the agent is either unaware of, or mistaken about, the reasons that bear on their choice to move in

---

[36] Ibid., pp. 136-7.

a particular way. As Wallace points out, for the conditions in these cases to be genuinely excusing, it must not be the case that 'the ignorance that makes what one did unintentional is itself culpable.'[37] Assuming it is not, then the agent is not aware of the true reasons which bear on her choice, and cannot be expected to be aware of those reasons, and therefore she is not morally responsible for failing to act on those reasons. Thus, the aspect of the global condition of responsiveness to reasons which has to do with knowledge or understanding – the 'ability to grasp and apply' reasons – is continuous with local examples of ignorance or lack of understanding in cases such as the slapdash chef, or Wallace's case of treading on someone's hand on the kitchen floor. A general inability to 'grasp and apply' reasons is excusing because it renders the agent non-culpably ignorant of the reasons that bear on individual choices that she makes.

Something similar applies in cases in which someone's moral responsibility or otherwise depends on whether they have control of their actions in the case. This includes cases of coercion and of involuntary movement. For example, in the variation of the vase case in which you push me into the vase, I am not morally responsible for breaking the vase because I am not in control of the action which leads to the vase's being broken. More generally, someone who lacks control over their bodily movements (say because they suffer from a neurological condition involving violent 'tics') might lack moral responsibility for a broad range of things resulting from those movements.

Both the 'knowledge' and 'control' conditions of moral responsibility, then, can apply either to specific actions, or generally to an agent in a way which renders that agent morally responsible (or not) for a broad range of actions, and the agent's position with respect to the reasons that bear on her actions is what makes the difference between responsibility and non-responsibility. The agent

---

[37] Ibid., p. 138.

is either in a position to engage with these reasons in her actions or she is not, because of conditions which apply either specifically to the case in question or generally across a range of cases. Wallace's formulation of responsiveness to reasons has to do with 'general rational powers', and so is focused on the latter conditions. However, given that the former conditions also have to do with whether the agent *can respond to* the reasons that bear on her choice, I see no reason not to refer to these conditions also in terms of responsiveness to reasons. Therefore, I will use the term to refer to both types of condition in this thesis, distinguishing when necessary between *global* responsiveness to reasons and *local* responsiveness to reasons.[38]

The idea of 'culpable ignorance' alluded to above would also bear some exploration. Wallace states that excuses arising from ignorance 'may not be accepted at all if the ignorance that makes what one did unintentional is itself culpable'[39]. However, this turns out not to be a complete description of what he has in mind:

> In that case it will be taken not for a valid excuse, but for evidence of one of a different family of moral faults that includes negligence, carelessness, forgetfulness, and recklessness. Thus, if the *s* whose hand I tread on is a baby I am supposed to be looking after, then I am presumably under an obligation to keep track of where the child is and what he is up to, and so my ignorance that I would be treading on the child's hand by going to the

[38] The use of the word 'global' should not be taken to imply that someone lacking one of these conditions must be unresponsive to *all reasons*. Clearly it is possible for an agent to possess general qualities which make one unable either to recognise, or to control one's behaviour in the light of, some reasons or kinds of reason but not others. Indeed this is precisely my conclusion in this thesis.

[39] Wallace (1994), p. 138.

> refrigerator would not excuse my treading on his hand. More precisely: it might excuse me from responsibility for directly treading on the child's hand, but only by making me vulnerable to the different charge of negligence, which led to the hand's being damaged.[40]

However, it seems to me that Wallace's first description of the case is actually more accurate than that following the phrase 'more precisely'. Surely in this case I am indeed morally responsible for treading on the child's hand, and not just for the negligence which led to my treading on his hand. In contrast to the slapdash chef case (in which the chef could not be expected to know that the customer was a terrorist), the fact that I have a specific responsibility to look after the baby in this case means that, although I may not be aware of the presence of the baby's hand, and the reason this supplies which bears on my choice to put my foot there, I *can reasonably be expected* to be aware of this. We are, it seems to me, morally responsible for failing to act on those reasons, and only those reasons, of which we can reasonably be expected to be aware. This result has implications for my broader project, since to show that psychopaths are not morally responsible for failing to act on a certain class of reasons will involve showing, not just that they are unaware of these reasons, but also that they cannot be expected to be aware of them.

Another way in which my view of moral responsibility differs from that given by Wallace is in respect of the range of cases in which people can be morally responsible, and the class of reasons on which we can hold people responsible for acting or failing to act. For Wallace, we hold people morally responsible primarily when we believe they have broken an obligation which we accept, and cases where there is no specific obligation in the case are defined as special cases, and related back to obligations by the insistence that the people involved

---

[40] Ibid.

must be the kind of people whom we *would* hold responsible if they had broken an obligation. Wallace gives a particular analysis of the relationship between obligations and 'moral reasons' in order to justify this position, but there is no need for me to endorse this analysis given my purposes here. I have argued for a broad conception of the range of reasons on which one can be morally responsible for acting or for failing to act. Carrying this through to the discussion of responsiveness removes the need for Wallace's formulation for special cases, involving obligations. Obligations, of course, generate a particular kind of reason, but there is nothing special – in this context – about reasons generated by obligations. The person would still need to be globally responsive to reasons, of course, and it might be implied by this that they would be the kind of person whom we would hold responsible if they broke an obligation, but this would be a secondary implication of the central requirement of moral responsibility – that the person is responsive to the reasons that bear on their actions.

One important implication of the responsiveness to reasons account, which is made clear by cases like the slapdash chef and the selfish entrepreneur, is that one can be responsible for some consequences of one's actions, and not for others, depending on what particular reasons bearing on that act one is responsive to. In fact, one can be responsible for an act construed in one way and not for the same act construed in another. In the slapdash chef case, imagine the chef serves the seafood to a number of different customers, only one of whom is a terrorist. In this case, the chef is still morally responsible for harming the other customers, because he can reasonably be expected to be aware of the reasons that speak against his doing this. Similarly, he is morally responsible for harming and incapacitating the terrorist. He is not, however, morally responsible for preventing the hijacking since, not knowing that his customer is a terrorist, he is not aware – and cannot be expected to be aware –

of the reasons bearing on this (preventing the hijacking) as a construal of his actions in serving undercooked seafood to this particular customer.

In the previous section I noted that Strawson's account links moral responsibility closely with qualities of will: we hold people morally responsible when their actions express a particular quality of will, or the absence of a particular quality of will which we expect from them. I also argued that this approach leaves Strawson with an incomplete answer to the question of why some people are morally responsible in some cases, and other people are not morally responsible in other cases. In my modified version of Wallace's position, being morally responsible for an act is a matter of being responsive (globally and locally) to the reasons that bear on that act. This account allows us to see why qualities of will are important indicators of moral responsibility. If we are unable to control our behaviour in the light of the reasons that bear on an act, then we have not exercised the kind of choice to perform that act that would demonstrate a quality of will, either good or bad, or the absence of a quality of will that was rightly expected of us. However, in some cases (the selfish entrepreneur and buying milk would be two examples), I would be able to exhibit the powers of reflective self-control in the choice without having a relevant quality of will, or lacking a quality of will that was expected of me. Thus, the link between moral responsibility and qualities of will is indirect and defeasible. We would expect qualities of will to figure in very many cases of moral responsibility, but not in all, and indeed this is what consideration of cases reveals.

For convenience, I have so far in this section been talking about moral responsibility for actions, but it is worth noting that the responsiveness to reasons account can also make sense of the other things for which I argued that we can be morally responsible in Section 1.4, namely states of affairs, attitudes, emotions and beliefs. There are reasons that speak in favour of or against our bringing about states of affairs, having certain attitudes and emotions, and

holding certain beliefs. For each of these, responsiveness to reasons represents a plausible way of distinguishing between cases where we are or are not morally responsible, in the same way as for actions. So, to develop the three cases I outlined when discussing this issue earlier:

> (1) If Stephen knows Johnny well, he is responsive to facts about Johnny's character which generate reasons which bear on Stephen's choice about whether or not to take him for a fool. If he does not know Johnny, then he is responsive to reasons that bear generally on the choice one has to take someone for a fool when one does not know the person in question. If, however, he has (through no fault of his own) mistaken Johnny for someone else who *is* a fool, then he is responsive to none of these reasons – to the reasons that bear on this particular case.
>
> (2) If Dave has burned Ray's favourite hat, and Ray knows about it, then Ray is responsive to the reasons that speak in favour of his being angry with Dave. The same is true if Dave has not burned the hat or done anything to incur Ray's wrath, and Ray is well aware of the situation. However, if Pete has burned Ray's hat, and created a plausible situation in which it looks as though Dave burned it, then Ray is not responsive to the actual reasons for and against anger directed at particular people with regard to the burned hat.
>
> (3) If Chris's racist beliefs are simply the result of his own irrational hatred and prejudice, then he is responsive to the reasons that bear on whether one should hold such beliefs. If, however, he has been brought up in a very isolated community, fed propaganda about the supposed inferiority of some races, and not been exposed either to any real members of those races or to any opposing views, then he is not responsive – because he cannot reasonably be expected to respond – to

> the relevant reasons and is therefore (plausibly) not morally responsible for holding beliefs that are contradicted by those reasons.

By modifying Wallace's theory so that moral responsibility is a matter of responsiveness to the specific reasons (not just those generated by obligations) that bear on a choice, then, we are left with a theory which allows us to fill the gaps left by the Strawsonian account. It provides a plausible explanation of why we hold some people, in some cases, morally responsible while others we do not, and it explains why there is frequently, but not always, a close link between moral responsibility and qualities of will.

The idea of responsiveness to reasons offers what I think is the best analysis of how we, as a matter of fact, naturally and instinctively arrive at ascriptions of responsibility. This is why it gives the most intuitively plausible results in the range of cases I have been discussing. It is the best analysis of what we mean when we say that someone *is* morally responsible for something, which is distinct from their being blameworthy or praiseworthy, but also from that thing's merely being attributable to them as an agent, since there are many cases in which an act, say, is attributable to someone as an agent, without their being responsive to the reasons that bear on that act. Vase 5 would plausibly be an example of this. In this case, the act of breaking the vase is an act which is attributable to me as an agent, and I am aware of the reasons which bear on that act, including the fact that it is your vase and an expensive one. However, I did not in this instance have the power to regulate my behaviour by the light of these reasons, and therefore I lack local responsiveness to reasons with regard to the act of breaking the vase.

The responsiveness to reasons account shows how the distinction between those who are morally responsible and those who are not is related to the question of what we hold people responsible *for*. We hold people responsible either for acting on (or holding beliefs based on, etc.), or for failing to act on (or

to hold beliefs based on, etc.) reasons of which they can be expected to be aware, and to which they can be expected to control their actions (beliefs, etc.) in response. If we do not hold someone responsible for failing to respond to a particular reason or set of reasons in this way, it is because they could not reasonably be expected to respond to that particular reason or set of reasons, either because of local conditions in the case, or because they are not globally responsive to a set of reasons that includes this particular set of reasons.

Building on this, the responsiveness to reasons account also provides what I hoped for in the previous section: not simply a justification for targeting the practices and attitudes involved in holding people responsible wherever happens to be natural, but a justification for targeting them precisely where we do naturally target them. If holding people responsible implies believing them to be responsive to reasons, then when we hold someone responsible who is not responsive to reasons, we are being irrational, and we are treating them unfairly (assuming we are aware, or should be aware, of the fact that they are not responsive to reasons). But when we hold them responsible and they *are* responsive, the attitudes we hold, and the practices we engage in, have the chance of being justified, assuming they are themselves sensitive to the relevant set of reasons arising from whatever the person at whom they are directed has done.

Let us explore how this works for the reactive attitudes, using anger as an example. If A is morally responsible for ϕ-ing, an action which they had an obligation not to perform, then B may be justified in being angry with A for A's ϕ-ing, assuming B is in a position to be angry with anyone for ϕ-ing. This is because A either was, or should have been, aware of the reasons arising from her obligation not to ϕ. It is fair to expect A to take proper account of reasons arising from obligations which she genuinely has, if she is responsive to those reasons, and anger can be partly a matter of believing someone not to have taken proper account of reasons arising from their obligations. (That anger has

cognitive content is a controversial idea, but it is one with which I agree and for which I will argue in Chapter 4.) Thus, particular reactive attitudes can be justified in the sense that the beliefs upon which they depend are justified beliefs, and part of what makes them justified beliefs is the fact that the person concerned is responsive to the reasons that bear on whatever it is about that person that is prompting the reactive attitude (i.e. they are morally responsible for it).

A similar story can be told about practices that depend upon responsibility ascriptions. So, for example, having justification for punishing someone for a crime may depend on having a justified belief that they are morally responsible for that crime. Whatever it is that justifies punishment (and there are of course conflicting accounts of this), the reasons justifying punishment of the individual will be related to the reasons to which the person must be responsive if they are morally responsible. If they did not have the powers of reflective control with regard to the crime, then punishing them is unjustified because the reasons that would normally justify punishment do not apply.

In short, the justification for each case of performing a practice or holding an attitude which is involved in holding people responsible is to be found in that particular practice or attitude, and is sensitive, in the right kind of way, to considerations about the person whom one is holding responsible, and whatever it is one is holding them responsible for. Strawson's point about the psychological strain and impoverishment of relationships which would result from abandoning the reactive attitudes is a plausible additional justification of them taken as a whole, but this is not the whole justification. The refinement offered by the responsiveness to reasons approach is therefore an improvement on Strawson's account.

Conclusions

Between Strawson's approach and Wallace's development of that approach in terms of responsiveness to reasons, we have, I think, a complete account of moral responsibility which is able to justify the practices and attitudes involved in holding people morally responsible, and to explain why these are applied in some cases but not in others. These were the two tasks which I set out for a theory of moral responsibility towards the beginning of the chapter.

The particular consequence of the account I have endorsed which is relevant to the central argument of this thesis is that someone cannot be held morally responsible for failing to act on reasons which she is incapable of recognising as reasons. As I have argued, to be morally responsible for an action, it must be the case that one can reasonably be expected to recognise those reasons. Whatever makes it the case that one can reasonably be expected to recognise a reason, one condition must surely be that one is capable of recognising that reason or, if not, that the conditions which make it the case that one cannot recognise the reason are not themselves within one's control. This second condition excludes cases where someone has, either intentionally or through negligence brought it about that she is incapable of recognising an important reason – for example, I have blindfolded myself while driving my car and, as a result, cannot see the child in the road or recognise that I have a reason to turn the steering wheel. As we will see in later chapters, this is relevant to the case of psychopaths because, if psychopaths are to be judged non-responsible, the conditions which lead to their unresponsiveness to reasons must not be under their control in an analogous way.

I will go on to argue that psychopaths are indeed unresponsive to a particular class of reasons in a way that renders them not morally responsible for failing to act on those reasons. However, before we can see why this is, we first need to have a good understanding of what is unusual about psychopaths, and particularly of what it is about them that might lead us to doubt that they are

morally responsible for the normal range of actions. Beginning to develop understanding is the aim of the next chapter.

Chapter 2: Psychopathy

Introduction

In the introduction, I gave a very brief sketch of the psychopathic personality-type. I suggested that such a person would be remorseless, cunning, and selfish. The function of this second chapter is to flesh out this picture. In doing so I will draw on literature from the fields of psychology, psychiatry and neuroscience. Some of these areas of study are better developed than others, and as with most scientific fields there are controversies about some central questions. It will not therefore be possible to give definitive answers to most of the questions which we would want to ask about psychopaths. However, it seems clear that an enquiry about the moral responsibility of psychopaths, a category of person which exists in reality, should be informed by empirical evidence as far as possible, and to this end I will present my own interpretation of the evidence as it stands.

My conclusion, based on evidence of the peculiar deficiencies exhibited by psychopaths, will be that these deficiencies are primarily emotional in nature. In the later chapters, I will attempt to show that this diagnosis supports the overall conclusion that such psychopaths (or at least 'hardcore' psychopaths – those at the upper end of the scale for emotional deficiencies and for deficiencies of empathy in particular) would be unable to respond to reasons in a way that would qualify them as morally responsible.

2.1 Diagnosis

The serious study of the phenomenon of psychopathy as it is now understood begins with the psychiatrist Harvey Cleckley's seminal 1941 study, *The Mask of Sanity: An attempt to clarify some issues about the so-called psychopathic*

personality.[1] Prior to this, in the nineteenth and early twentieth centuries, there had been a number of attempts to describe those psychological conditions that were associated neither with delusions of any kind, nor with intellectual impairment, and yet which affected the subject's social functioning. Indeed, the term 'psychopath' was originally intended to cover *all* members of this very broad category, which is why the word's etymological meaning is so vague: literally, 'psychopath' means nothing more specific than 'diseased mind'. Other terms, including the 'moral insanity' generally thought to have been coined by James Cowles Prichard,[2] and the 'moral imbecility' favoured by Henry Maudsley[3] and Havelock Ellis,[4] are slightly more specific, but still include a much greater variety of phenomena than would now be categorised as psychopathic. Prichard's concept of moral insanity, for example, included mental conditions that would today be classed as depressive or bipolar.[5]

Cleckley, a practising psychiatrist who had worked in an asylum for many years before writing his book, used the term psychopath, apparently derived from 'the vernacular of the ward or the staff room',[6] to refer to a class of psychiatric patients who, having been committed to the asylum because of a clear inability to function within society – manifesting in a series of typically petty and impulsive criminal acts – failed to show any evidence of psychosis or neurosis once admitted. These patients, though apparently lucid and rational, had failed

---

[1] Cleckley (1941).

[2] Pritchard (1835).

[3] Maudsley (1873), Maudsley (1874).

[4] Ellis (1890).

[5] A very helpful discussion of the diagnostic history of psychopathy and similar conditions is provided by Ward (2010).

[6] Cleckley (1941), p. 20.

'to translate [their] apparent rationality into the successful conduct of life'.[7] Cleckley combined detailed case studies of thirteen such patients, with a careful description and analysis of what he saw as the common condition which afflicted them. He characterised this condition as a 'pathologic general devaluation of life, a complex deficiency, confusion, or malfunction in what chooses aims and directs impulse'[8], and coined the term 'semantic disorder' to refer to the absence of meaning he perceived in the worldview of his patients.

Stemming from the 'general devaluation', Cleckley identified a number of observations about psychopathic lifestyle and personality, which were later adopted by R.D. Hare as the basis of his Psychopathy Checklist (PCL) and its later revised version (PCL-R).[9] Since this checklist is the central diagnostic tool for psychopathy, and can therefore be taken to define the concept of psychopathy as it is applied in clinical contexts, it is worth reproducing it in full here.

---

[7] Ward (2010), p. 21.

[8] Cleckley (1941), p. 172.

[9] Hare's construct of psychopathy is generally thought to have 'drifted' somewhat from Cleckley's description. Hare and Neumann (2008) are happy to accept this claim, noting among other factors the relatively small sample size of Cleckley's work compared to Hare's own. Nonetheless, Hare's PCL-R retains significant similarity to the phenomenon described by Cleckley.

Figure 1: The Hare Psychopathy Checklist, Revised Version (PCL-R)[10]

Factor 1: personality "Aggressive narcissism"	**Factor 2: case history "Socially deviant lifestyle"**
1. Glibness/superficial charm	9. Need for stimulation/proneness to boredom
2. Grandiose sense of self-worth	10. Parasitic lifestyle
3. Pathological lying	11. Poor behaviour controls
4. Cunning/manipulative	12. Lack of realistic/long-term goals
5. Lack of remorse or guilt	13. Impulsivity
6. Shallow affect (genuine emotion is short-lived and egocentric)	14. Irresponsibility
7. Callousness; lack of empathy	15. Juvenile delinquency
8. Failure to accept responsibility for own actions	16. Early behaviour problems
	17. Revocation of conditional release
Neither factor	
18. Promiscuous sexual behaviour	
19. Many short-term marital relationships	
20. Criminal versatility	

The checklist is applied by clinicians on the basis of file information and – usually – a semi-structured interview with the subject.[11] The subject is given a score of 0, 1 or 2 against each of the 20 items, reflecting the extent to which he or she demonstrates the given trait. A score of 30 or more overall (out of a possible 40) is typically used as the cut-off point for a diagnosis of full-fledged psychopathy.

[10] Hare (1998).

[11] According to Hare, 'the PCL-R can be scored on the basis of file information alone, provided that the material contained in the files is extensive and detailed, and that the rater acknowledges the limitations of the procedure' (Hare 1998, p. 101).

It is notable that, while many of the items in the PCL-R are directly observable facts about the subject's lifestyle and behaviour (e.g. 'juvenile delinquency', 'promiscuous sexual behaviour'), others refer to personality traits which must be inferred by the person applying the checklist (e.g. 'callousness, lack of empathy', 'grandiose sense of self-worth'). The apparent element of subjective judgment introduced by this aspect of the PCL-R worried the compilers of the American Psychiatric Association's Diagnostic and Statistical Manual (DSM), currently in its 5th edition (DSM-V).[12] This manual, which is the global standard diagnostic tool for psychiatrists, stipulates that classification in psychiatry should not include reference to underlying causes or inferred psychological traits. For this reason, the DSM's compilers did not adopt Hare's construct, replacing it instead with 'Antisocial Personality Disorder' (APD), which is applied on the basis of observed behaviour only. As Minzenberg and Siever note, DSM 'criteria for APD consist almost exclusively of behavioural indicators, neglecting the affective-interpersonal features that appear to reflect much of the notion of a distinct personality type as described by Cleckley'.[13] Though the DSM states that APD 'has also been referred to as psychopathy, sociopathy, or dissocial personality disorder',[14] it is clear that Hare's construct of psychopathy and APD are not the same thing. Importantly, APD bears significant relation to Factor 2 of PCL-R, as opposed to Factor 1.[15] The emphasis on quantifiable behavioural tendencies has had the effect of creating a new construct that shares many of the elements of psychopathy, but favours those of the lifestyle/antisocial type over those of the interpersonal/affective type. Given the

---

[12] American Psychiatric Association (2013).

[13] Minzenberg and Siever (2006), p. 251.

[14] American Psychiatric Association (2013), p. 645.

[15] Hare and Neumann (2010).

similarities and differences between the two constructs, APD is often seen as a rival to PCL-R.

I will take the PCL-R diagnosis to be the central one. This decision is based on a number of factors, including three interrelated worries about the usefulness of APD as a construct in grounding judgments of responsibility. Firstly, since the question of responsibility will depend on judgments about the psychological make-up of individuals – about their rational and emotional deficits and so on, a diagnosis which makes explicit reference to these psychological features is likely to serve as a stronger ground for such judgments, and APD does not do this. Secondly, PCL-R has been proven to be a better predictor of behaviour, including criminal recidivism,[16] than APD. In addition, APD applies to a much wider class of people than PCL-R. For example, the majority of prison inmates have APD,[17] whereas PCL-R diagnoses only around 20%.[18] In civil populations too, there is 'a prevalence of APD… that is at least three times the prevalence of psychopathy (based on the PCL-R and PCL)'.[19] Together, these facts suggest that PCL-R picks out a much more tightly defined set of personality traits, whereas the antisocial behaviour of those diagnosed with APD may have its roots in more disparate aspects of personality, or in environmental or social factors. Though there will be considerable variation even within the PCL-R diagnosis of psychopathy, the more closely related are the individuals picked out by that diagnosis, the more likely are judgments of responsibility to apply to a greater number of those individuals. Thirdly, there is reason to question whether a diagnosis of APD truly functions as an *explanation* of behaviour in

---

[16] Hemphill, et al. (1998).

[17] Hare (1995), p. 25.

[18] Ibid., p. 87.

[19] Hare and Neumann (2010), p. 131.

the same way that a PCL-R diagnosis of psychopathy does. Indeed, since the diagnosis of APD is based entirely on observed behaviour, and the diagnosis does not include inferred personality traits, it is difficult to see how a person's having APD can explain their behaviour in a non-circular way. Discussions of attenuated responsibility owing to mental disorders often use the language of explanation (e.g. 'he killed her because he's a schizophrenic'). If this option is not available in the case of APD, a discussion about responsibility may have trouble getting off the ground. Interestingly, this relates to Barbara Wootton's attempt to settle the question of psychopaths' responsibility *a priori* that was discussed in the introduction. It may be that the charge of circularity that Wootton brought against any attempt to prove non-responsibility on the basis of a diagnosis of psychopathy, which fails when PCL-R is used, might have more traction in the case of APD.

One final pragmatic reason for favouring PCL-R is that it is the dominant diagnostic tool used by researchers examining the psychological and neurological mechanisms underlying psychopathy. The vast majority of the literature in these fields uses PCL-R to identify research participants. This may be partly because of the worries noted above. Though PCL-R, like APD, is not immune to worries about aspects of its validity as a construct,[20] it is seen by most psychopathy researchers as the most useful diagnostic tool available.[21] It is therefore possible to build up a more detailed and nuanced picture of what is at stake using PCL-R rather than APD.

2.2 Emotional deficiencies

It is immediately noticeable how many of the items in the PCL-R might be explained by means of specifically emotional deficiencies. A lack of remorse or

[20] See for example the correspondence on psychopathy and anti-social behaviour in the *British Journal of Psychiatry* 191 (2007), pp. 357-365.

[21] E.g. Blair, et al. (2005).

guilt, shallow affect and callousness/lack of empathy all naturally fit into the category of emotional deficits. Glibness and superficial charm perhaps suggest a lack of deep emotional engagement. Many of the items which involve manipulating or generally mistreating others – pathological lying, cunning/manipulative, parasitic lifestyle, criminal versatility – might be thought to be the result of an emotional lack, particularly a lack of empathy, or of fear or guilt, emotions which, it might be thought, regulate our behaviour and prevent us from harming or taking advantage of others when it might otherwise be in our interest to do so. There are also a number of aspects of lifestyle in the list which look like facets of a general inability or unwillingness to order one's behaviour and one's life in the pursuit of general goals: failure to accept responsibility for one's actions, lack of realistic/long-term goals, irresponsibility, juvenile delinquency, early behaviour problems, revocation of conditional release, many short-term marital relationships. Recent developments in neuroscience suggest that a function performed by emotions is to shape our lives by imposing checks on certain forms of behaviour, and encouraging others.[22] If this is true then a mental condition which lacks the emotional richness of ordinary human life might plausibly be expected to result in impulsivity and poor behaviour control, which might in turn manifest in the aspects of lifestyle referred to above. All of this points to a disorder which is essentially emotional in character. As we will see, this suggestion is backed up by neurological evidence: the parts of the brain affected in people with psychopathy are primarily the parts involved in emotional processing.

2.3 A distinct condition?

If psychopathy is a construct made up of twenty separate personality and behavioural traits, it might be wondered whether we are discussing a distinct condition at all. It might be that 'psychopath' is simply a word for someone

[22] See e.g. Damasio (2006).

who, for a disparate and random combination of reasons, happens to demonstrate a large number of these traits to a high degree. Indeed, it is generally accepted that psychopathic traits, in common with those associated with other personality disorders, exist on a continuum. That is, these traits are not unique to psychopaths, but are found also in the general population to a greater or lesser extent. We all know people who lack empathy, or are impulsive, or have trouble working towards long-term goals. We may indeed observe some of these traits in ourselves at times. On the other hand, it is apparently the case that psychopathic traits tend to cluster together, suggesting that they are related, or perhaps the product of an underlying cause. Interestingly, a large-scale study[23] found that all of the factors in the PCL-R construct correlate positively with a single, 'superordinate' factor, suggesting that the lower-order factors are related by a common theme, which Hare and Neumann characterise as 'the broad dissocial nature of psychopathic traits'.[24]

Perhaps the best reason for regarding psychopathy as a distinct condition is given by the promising attempts (to which I will turn shortly), to identify a neurological basis for the disorder. Antonio Damasio[25] for example, discusses several cases where lesions to the amygdala and frontal regions of the brain have resulted in symptoms very close to those found in psychopaths. Coupled with the numerous studies showing reduced activity in these same regions in psychopaths' brains, it is plausible to suppose that (partly genetically determined) reduced *functioning* in specific brain regions, as well as that which is the result of injury, might be the *cause* of psychopathic symptoms (rather than reduced *activity* being merely a *correlate* of these symptoms). If the cluster of symptoms associated with psychopathy has a single neurological cause, or a

---

[23] Hare and Neumann (2008).

[24] Ibid.

[25] Damasio (2006).

small set of closely related causes, then it is not a random cluster of unrelated traits.

Nonetheless, it is important to remain aware of the fact that not all psychopaths are alike, and not every psychopath will demonstrate every item in the PCL-R to a high degree. This has important consequences for the question of moral responsibility. Discussions of moral responsibility can sometimes give the impression that it is a binary concept – one is either morally responsible or one is not. However, this is probably not the case. The law recognises not just a complete *lack* of responsibility, but also *diminished* responsibility, and it is likely that this idea has an ethical parallel, so that there are degrees of moral responsibility in some cases. If I am a minor shareholder in a company that dumps hazardous waste in the sea, I am presumably not as responsible as the executive who ordered the dumping, but I am also plausibly more responsible than someone who has nothing at all to do with the company (is not aware of the dumping, and so on). It may also be the case that the degree of moral responsibility I have for an action tracks other concepts, such as my understanding of the facts, or of the reasons bearing on decisions I have made. If so, conclusions about the moral responsibility of psychopaths may not apply to all psychopaths equally. This fact will need to be borne in mind when considering arguments in the following chapters. It might be thought that psychopaths will lack responsibility *insofar as* they lack the relevant traits, with only the most hardcore, high-scoring psychopaths, or perhaps only those psychopaths lacking the relevant traits to a very high degree, lacking moral responsibility completely. However, we should not leap to the conclusion that degrees of responsibility will straightforwardly track degrees of possession of the relevant attributes. It might be that possession of a given attribute to *any* degree is enough of a window to allow responsibility to enter the picture.

2.4 'Successful' and 'unsuccessful' psychopaths

Despite the apparent links between what we might term the *moral* and *prudential* aspects of the condition, both at the level of psychology and of neurology, some scientists have recently begun to cast doubt on the idea that these two families of traits should each be thought of as necessary conditions of a single, overall diagnosis of psychopathy. Gao and Raine[26] present a fascinating review of recent studies which have sought to distinguish between 'successful' and 'unsuccessful' psychopaths. This distinction has been made in the past, but several recent studies have suggested that it may be more important than previously thought. Successful psychopaths are defined as those who have little or no history of criminal conviction and incarceration. In fact, the vast majority of our evidence concerning psychopaths comes from *unsuccessful* psychopaths, because of the relative ease of identifying and gaining access to psychopaths among prison populations. Among the most striking of the results surveyed in Gao and Raine's paper is evidence that successful psychopaths may not exhibit the same neurological deficits associated with psychopathy generally, which will be summarised in the following section.[27] Successful psychopaths have also been found to have executive functioning which is not only unimpaired, but may actually be improved in comparison to non-psychopathic controls.[28] As Gao and Raine point out, 'our research knowledge based on incarcerated psychopathic offenders may not be generalisable to psychopaths in the general population'.[29] Studying successful as well as unsuccessful psychopaths is essential if we are to understand the class of psychopaths as a whole. In the scientific literature, the research summarised by Gao and Raine has led to a

---

[26] Gao and Raine (2010).

[27] E.g. Yang, et al. (2005), Yang, et al. (2011).

[28] Ishikawa, et al. (2001).

[29] Gao and Raine (2010), p. 196.

debate about whether violent, criminal and anti-social behaviour should be considered an intrinsic element of the psychopathic personality. Cooke and Mitchie[30] have proposed that the behavioural (Factor 2) aspects of Hare's checklist should be considered a contingent effect of psychopathy, rather than an inherent aspect of it.

This research may have implications for judgments of moral responsibility too: Sifferd and Hirstein[31] argue that only unsuccessful psychopaths can be said to have reduced moral responsibility, on the grounds that successful psychopaths have unimpaired executive function and are capable of contravening moral norms intentionally. This conclusion is not obviously correct however: if, as many philosophers have argued, the emotional deficits of psychopaths are in themselves enough to deliver a verdict of non-responsibility, then this verdict will apply equally to successful and to unsuccessful psychopaths. One characteristic which is not generally thought to differ between successful and unsuccessful psychopaths is their lack of emotional empathy.[32] If this lack of empathy is indeed shared by successful psychopaths as well as unsuccessful ones, and is the result of factors beyond their control, such as neurodevelopmental factors, then this may be enough to ground a verdict of non-responsibility.

2.5 Psychopathy and the brain

As noted above, recent attempts to establish a neurobiological basis for psychopathy have proved somewhat fruitful. On the other hand, the techniques used to examine structural and functional aspects of the brain are still developing rapidly, the relationship between different regions of the brain

---

[30] Cooke and Michie (2001).

[31] Sifferd and Hirstein (2013).

[32] Gao and Raine (2010), p. 204.

and different psychological phenomena is only partly understood, and the business of relating personality traits to neurological phenomena is a complex one. Therefore, any conclusions drawn from neurological studies must be highly tentative. Still, it is possible to discern patterns in the results which are worth discussing here. Overall, a picture is beginning to emerge of a neurodevelopmental disorder with a significant genetic basis, though there are probably environmental factors involved in producing its full clinical manifestation. [33]

Studies have been carried out investigating two separate aspects of the neurology of psychopaths: brain activity and brain structure. These are importantly distinct because they point towards two distinct overall types of conclusion. On the one hand, to show that psychopaths have particular patterns of activity in the brain, perhaps when performing particular kinds of task, is broadly to provide evidence that they are using particular regions of the brain as opposed to others. This might lead one tentatively to conclude, for example, that psychopaths tend not to engage their emotions as much when performing certain types of task, compared to normal agents. It says nothing about *why* this is the case. Showing that psychopaths have differently *structured* brains, on the other hand, provides evidence that aspects of their psychology might have a particular neurological *cause*. If a psychopath's brain shows reduced volume in a region associated with a particular kind of emotional processing, then it might be possible to conclude that unusual patterns in their experiencing of the relevant kind of emotion are due to their not having the same neurological resources as normal agents. This in turn might lead them to

---

[33] Gao, et al. (2009). In this section I will be relying heavily on this and another review of the neuroscientific literature: Seara-Cardoso and Viding (2014). Blair (2010) provides a further useful review of studies, focusing on those concerned with structural and functional differences, and the particular issue of instrumental vs reactive aggression.

compensate by engaging other regions of the brain when performing tasks for the completion of which normal agents would be likely to engage their emotions, which might manifest at the psychological level through the use of, for example, cognitive strategies rather than emotional ones. Ultimately, this may have implications for things such as understanding and value, which are interesting in relation to moral responsibility. However, it is worth reiterating that such conclusions would only be tentative given the inexactness and incompleteness of the science in this area.

Seara-Cardoso and Viding[34] reviewed studies which used functional magnetic resonance imaging (fMRI) technology to assess whether psychopaths (diagnosed using PCL-R) showed decreased activity in specific areas in the brain correlating with their performance of specific tasks, split into three groups: tasks designed to stimulate emotional processing in general, tasks designed to provoke empathy, and moral judgment tasks.

Turning to the basic emotional processing studies first, Seara-Cardoso and Viding reviewed studies that involved different tasks: tasks involving passively observing photographs designed to stimulate particular emotional responses;[35] memory tasks involving remembering words that have either a neutral or a negative emotional valence;[36] tasks involving the recognition of faces that have either an emotionally neutral or emotionally aroused expression;[37] tasks involving passively observing faces, again with expressions that are either emotionally neutral or emotionally aroused.[38] The studies primarily examined

---

[34] Seara-Cardoso and Viding (2014).

[35] Muller, et al. (2003).

[36] Kiehl, et al. (2001).

[37] Deeley, et al. (2006).

[38] Decety, et al. (2014).

brain regions associated with emotional processing, including the amygdala, the anterior insula and various portions of the prefrontal cortex. They found that psychopaths showed consistently less activity in these brain regions when performing these tasks, compared with non-psychopathic controls.

The studies designed to find neural correlates to empathy-based tasks again used a number of such tasks, including observing pictures of people apparently in pain,[39] observing videos of people's hands in situations with emotional implications (e.g. a hand being hit, or caressed, by another hand),[40] and trying to guess the emotional state of protagonists in a cartoon story.[41] The brain regions studied included the amygdala, anterior insula, inferior frontal gyrus, and dorsal anterior cingulate, all of which are associated with empathy-related tasks. Again, psychopaths showed consistently lower levels of activity in these regions when performing the tasks relative to non-psychopathic controls.

Two specific results are interesting enough to be worth noting here. Firstly, Decety, Skelly et al[42] showed a group of psychopaths pictures of people apparently in pain, and found, as expected, reduced activity in relevant brain regions in these subjects relative to controls. However, in a follow-up study[43] it was found that manipulating the instructions given to subjects had an effect on the level of brain activity displayed. When instructed to imagine the person in the picture being in pain, they continued to display reduced activity. However, when asked to imagine *themselves* in similar pain, they showed *increased* activity. This suggests that psychopaths' own pain may even be more salient to

---

[39] Decety, et al. (2013b).

[40] Meffert, et al. (2013), Decety, et al. (2013a).

[41] Sommer, et al. (2010).

[42] Decety, et al. (2013b).

[43] Decety, et al. (2014).

them than normal subjects' own pain is salient to them, while others' pain is less salient to the psychopath.

Another interesting result was discovered in Sommer, Sodian et al's study.[44] When trying to guess the emotional state of a cartoon character, as well as showing reduced activity in brain regions associated with emotional processing (superior temporal sulcus, supramarginal gyrus, frontal gyrus), the psychopathic group showed increased activity in regions associated with 'processing the value of an outcome and mentalising efforts'.[45] This, conclude Seara-Cardoso and Viding, 'may reflect additional efforts in computing the emotion attribution due to an inability to automatically simulate the emotional state of the cartoon character'.[46] This suggestion, which invites us to picture psychopaths as detectives, using inductive reasoning to arrive at a conclusion about the mental state of a person ('I've seen people pull facial expressions like that before, when they were in pain… so that's probably what's going on here') where a normal agent would simply *see someone in pain*, raises interesting questions about what it is to empathise with someone. To acquire knowledge about their mental state? To perceive them directly as experiencing a particular mental state? To feel some of what they are feeling? We might also ask what implications not having access to the full range of processes of this kind would have for a person's attitude and behaviour toward others. I will return to these questions in Chapter 5.

The final set of studies reviewed by Seara-Cardoso and Viding concerned psychopaths' brain states when engaging in moral judgment tasks. In these tasks, psychopaths were presented with a series of moral 'dilemmas' (in fact

---

[44] Sommer, et al. (2010).

[45] Seara-Cardoso and Viding (2014), p. 5.

[46] Ibid.

strictly speaking these were not dilemmas but situations requiring a difficult moral judgment to be made by a protagonist – the experimental subject is asked what the protagonist should do in the situation). The results from these experiments were also extremely interesting. Glenn, Raine et al.[47] found that, when presented with a) moral dilemmas designed to illicit strong emotional reactions, b) moral dilemmas designed not to illicit strong emotional reactions and c) non-moral dilemmas, psychopaths showed no significant difference from non-psychopathic controls in their responses to the dilemmas. However, the psychopaths showed less activity than the non-psychopaths in brain regions associated with emotional processing (amygdala, medial prefrontal cortex, posterior cingulate, and angular gyrus) when considering the emotional dilemmas, and increased activity in brain regions associated with cognitive control (dorsolateral prefrontal cortex). Very similar results were also shown by Pujol, Batalla et al.[48] As with the empathetic tasks, these results suggest that psychopaths may be pressing into service non-emotional, cognitive processes in order to perform tasks which normal agents would perform using emotion. Moreover, given the fact that the psychopaths did not differ from the non-psychopaths in the actual responses given to the dilemmas, it would appear that this strategy on their part is successful, at least for the range of moral dilemmas included in the experiments. The relative lack of emotional engagement shown by the psychopaths was not a disadvantage in performing the tasks.[49] This is suggestive in relation to psychopaths' ability to make moral judgments. In Seara-Cardoso and Viding's words, 'these results suggest that moral judgment ability may be spared in individuals with psychopathy but that they may use

---

[47] Glenn, et al. (2009).

[48] Pujol, et al. (2012).

[49] Neither was it apparently an advantage: the balance between emotional and cognitive resources dedicated to the moral judgment task by different subjects simply made no difference to the answers given.

different strategies, or different brain regions, to compute their judgments'.[50] Philosophers should be more cautious, however: similarity in the verdicts given does not prove that the psychopaths in the studies were indeed making moral judgments. We might tentatively conclude, however, that at least for the range of scenarios tested, the psychopathic subjects were capable of either making, or successfully *faking*, normal moral judgments.

Summarising the studies in their review, Seara-Cardoso and Viding state that, 'although the direction of the findings is not entirely consistent across studies, overall, these studies seem to point to reduced response in regions typically associated with affective processing and increased activity in regions typically associated with cognitive control during processing of emotional and salient stimuli.'[51] Interestingly, many of the studies that were reviewed used 'community samples', i.e. psychopaths drawn from the general population and not only from prisons and psychiatric institutions. The conclusions canvassed above, therefore, would appear to apply to 'successful' as well as to 'unsuccessful' psychopaths. This would still be consistent with the idea that these two categories represent distinct personality-types, but with overlapping emotional deficits.

In Gao, Glenn et al.'s earlier review,[52] studies were chosen that focused on investigating possible structural and functional correlates associated with psychopathy (diagnosed using PCL-R). As noted above, work of this kind suggests a neurological cause of psychopathic traits, as opposed to merely supporting the hypothesis that psychopaths use different psychological strategies when performing different tasks. As such it is particularly interesting

---

[50] Seara-Cardoso and Viding (2014), p. 6.

[51] Ibid., p. 7.

[52] Gao, et al. (2009).

in the context of the present enquiry. Nonetheless it should be emphasised that identifying an associated structural or functional deficit in psychopaths is not the same as identifying a genetic cause to psychopathy, since brain structure and function develops from childhood onwards. It could be that structural and functional abnormalities in adult psychopaths are the result of abnormal neurological development, rather than an abnormal genetic inheritance.

Past support for an underlying role of functional connectivity in the manifestation of psychopathic traits has come from research involving subjects who have suffered injury to specific brain regions, and as a result have developed traits similar to those found in psychopaths. The phenomenon of 'acquired sociopathy' in patients suffering damage to the ventromedial prefrontal cortex (vmPFC) was explored in a seminal study by Eslinger and Damasio.[53] More recently, using advanced brain imaging techniques such as voxel-based morphology, researchers have turned to psychopathic subjects who have not suffered from brain injury, and attempted to discover whether they too have structural deficits in comparison to non-psychopathic controls.

A correlation between high scores on PCL-R and reduced volume in the vmPFC has been shown repeatedly, including by Yang, Raine et al.,[54] Muller, Sommer et al.[55] and de Oliveira-Souza, Hare et al.[56] A significant correlation between psychopathic traits and reduced volume in the amygdala has also been shown, by Yang, Raine et al.[57]. Both regions are associated with emotional processing.

---

[53] Eslinger and Damasio (1985).

[54] Yang, et al. (2005).

[55] Muller, et al. (2008).

[56] de Oliveira-Souza, et al. (2008).

[57] Yang, et al. (2009), Yang, et al. (2010).

Yang, Raine et al.[58] also distinguish between successful and unsuccessful psychopaths, finding differences in the specific regions affected. In the words of Gao et al., 'findings suggest that neuropathological characteristics such as abnormal hippocampal asymmetry and reduced prefrontal grey matter volume may contribute to the emotional dysregulation and poor fear conditioning in unsuccessful psychopathic people, and consequently render these people less sensitive to environmental cues predicting danger and capture'.[59] This interesting result suggests that the prudential deficits exhibited by unsuccessful psychopaths may have a separate neurological basis from the moral deficits found in both successful and unsuccessful psychopaths.

Overall, the evidence appears to suggest that several brain regions associated with the emotions are typically underdeveloped in people with psychopathic tendencies:

> Overall, brain imaging studies have suggested that: the orbitofrontal, ventromedial prefrontal, and the cingulate cortex are crucial in decision-making, behavioural control, and emotional regulation, and that deficits in these regions may contribute to features such as impulsivity and impaired moral judgment in psychopathic people; and, the medial temporal regions, particularly the amygdala and hippocampus, are critical for emotional processing, and thus, when impaired, predispose to a shallow affect and lack of empathy in psychopathic people. Findings also suggest that no one single region, when impaired, will result in psychopathy.[60]

---

[58] Yang, et al. (2005).

[59] Gao, et al. (2009), p. 814.

[60] Ibid., p. 815.

As noted above, this evidence is particularly important because it points towards underlying neurological deficits as a possible *cause* of psychopathic traits. In particular, this has potential implications for moral responsibility: if the structure of one's brain predisposes one to have certain psychological traits, and if the structure of one's brain is largely established by adulthood[61], then we might be less inclined to hold people with the relevant psychological traits responsible for the possession of those traits. If those traits turn out to be incompatible with moral responsibility themselves, then it becomes much more plausible to conclude that people who possess them to a high degree are not morally responsible at all.

Conclusions

In summary, we have seen in this chapter that the psychopathic personality is characterised primarily by emotional deficits, which may to some extent be caused by underlying abnormalities in their brain structure, and which manifest in forms of behaviour that can be described as antisocial or amoral.

I believe these factors provide the foundations for a verdict of non-responsibility, but before this verdict can be confidently made, there is considerable ground to cover. In the next chapter, I will interpret the conclusions of this chapter, particularly focusing on their implications for the ability of psychopaths to recognise the value of others. Ultimately, I think the *in*ability of psychopaths to recognise this value provides good grounds for thinking that they are incapable of recognising a broad category of reasons, and therefore that they are not morally responsible for failing to act on reasons which belong to that category.

---

[61] Stiles and Jernigan (2010).

Chapter 3: Psychopathy and moral responsibility

Introduction

I have now outlined my understanding of what moral responsibility is, and made some observations about what I take to be the central psychological deficits which characterise psychopathy. I turn in this chapter to the central question of the thesis, namely whether psychopaths are morally responsible or, more precisely, for what kinds of action, attitude, emotion, etc., psychopaths are capable of being morally responsible. In the second part of the chapter I will identify a category of actions (etc.) for which I believe psychopaths are not capable of being morally responsible. However, before I turn to this part of the project I will assess some prominent arguments both for and against psychopathic responsibility. I will concentrate on those arguments which are rooted in the two fields of thought concerning moral responsibility which I surveyed in Chapter 1: the reactive attitudes view and the reasons-responsiveness view.

3.1 Psychopathy and the reactive attitudes

As set out in Chapter 1, my view of moral responsibility draws heavily on the insights expressed by P.F. Strawson in 'Freedom and Resentment'.[1] There have been some attempts to settle the question of psychopathic responsibility using these insights, and it will be worthwhile discussing these attempts here. Probably the most fully Strawsonian discussions of psychopathic responsibility are those by Piers Benn and Gwen Adshead, and I will concentrate on these discussions in this section.

As we have seen, Strawson's account of 'participant reactive attitudes' in 'Freedom and Resentment' is intended as a way of sidestepping the debate between compatibilists and incompatibilists about determinism and moral

[1] Strawson (2008).

responsibility. His major claim is that the set of attitudes which embody our practices of holding people responsible is not, and should not be, prey to general theoretical convictions such as a belief in the truth of determinism. Thus, Strawson attempts to show, against the incompatibilist position, that we are sometimes justified in holding people morally responsible for their actions, but he attempts to do so without solving the theoretical problem of the supposed incompatibility of moral responsibility and determinism. Instead, he proceeds by illuminating the nature of the normal social practices involved in holding someone morally responsible, so that it becomes clear in his view that they do not require external justification. As part of this project, he sets out a typology of cases in which it would be acceptable and normal to withhold reactive attitudes; for example, towards children or animals, or people who are in the grip of ignorance or compulsion. In these cases, the actions in question are not expressive of either 'ill will or indifferent disregard',[2] and hence we are justified in holding the agents morally responsible. I argued in Chapter 1 that this emphasis on qualities of will provided only an incomplete account of how the morally responsible should be separated from the morally non-responsible, but that the strain on human relations of abandoning our intuitive practice of deciding when to hold someone or something morally responsible provides a sound justification for our maintaining that practice. In all of the cases I described in the previous chapter, this intuitive process delivered a reliable verdict of responsibility or non-responsibility. As I noted right at the beginning of the thesis, however, the case of psychopaths is notable for pulling our intuitions in two directions simultaneously. The question of what kind of reactive attitudes, if any, we should hold towards psychopaths is therefore a difficult one.

---

[2] Ibid., p. 15.

While Strawson does not mention psychopaths specifically, it is worth looking for hints in 'Freedom and Resentment' as to what he might think about them. The most promising clue is in Strawson's list of those excluded from being a target of reactive attitudes, in which he includes both 'the extreme case of the mentally deranged' and the case of the 'moral idiot'.[3] Either of these categories might be thought to include psychopaths. However, this is not in fact suggested by Strawson's discussion. More likely, Strawson intends these terms to pick out individuals whose mental condition more obviously includes them in one of the traditional (Aristotelian) categories of exemption from moral responsibility – lack of control and lack of knowledge. Thus, someone who is 'mentally deranged' in a way that exempts them from moral responsibility would need to be significantly mistaken about the facts bearing on their choices (for example because they suffer from paranoid delusions). A 'moral idiot', on the other hand, might be understood as someone who fails to have a competent grasp of moral concepts and their application. Psychopaths do not fit easily into either of these categories. Most importantly from a Strawsonian point of view, they are typically quite capable of both ill will and indifferent disregard towards others. Indeed, they are somewhat expert at this. It might seem, therefore, that psychopaths are apt targets of Strawsonian reactive attitudes.

Piers Benn has tried to show that this is not the case. In 'Freedom, Resentment and the Psychopath',[4] he sets out an interpretation of the reactive attitudes as essentially 'communicative'. For Benn, Strawsonian participant reactive attitudes should be understood as acts of communication with the person towards whom the attitude is held. If someone has wronged me (for example), and I resent them for it, I thereby have the capacity to create two things in the person who wronged me: firstly, the understanding of what they have done

---

[3] Ibid., p. 13.

[4] Benn (1999).

wrong, and secondly, the motivation to act on this understanding (by making amends, or changing their ways in the future, etc.) This second aspect is achieved through the creation of self-directed reactive attitudes such as guilt, shame or remorse. Now, it is notable of psychopaths, as we saw in the previous chapter, that they appear to be incapable of genuinely holding these self-directed reactive attitudes. Insofar as our own reactive attitudes such as resentment have a supposed role to play[5] in bringing about these attitudes, this role is frustrated when they are directed at psychopaths. In terms of attitudes, it seems, psychopaths simply do not speak our language. This analysis leads Benn to a more general conclusion: 'only creatures able to form participant attitudes are proper objects of such attitudes on the part of others'.[6] Psychopaths are able to hold attitudes, of course, but these attitudes are not *participant* in the way Benn understands this term, and therefore we go wrong when we hold such attitudes towards them.

From this, Benn draws a tentative second conclusion which he describes as 'illiberal'. Endorsing a similar point made by Jeffrie G. Murphy[7] he suggests that, to the extent that our moral treatment of others is motivated by 'Kantian'

---

[5] It is a little unclear how this idea of a role should be cashed out. It is presumably not the case that we are always supposed to be *intending* to communicate anything to the object of our resentment, since (for one thing) resentment can sometimes be unexpressed and is no less justified for that. It is also presumably not the case that the communicative role described by Benn is supposed to be the only role resentment has. Indeed, it is probably not the only communicative role it has: one role of expressing our resentment might be to communicate to others something about what type of a person the object of the resentment is. There might also be some use in simply organising our own thoughts about what attitudes we should have towards the person in future. Nonetheless it is plausible that one of the *uses* that resentment has is to communicate something to its target, and that this might be one way in which resentment can be *valuable*.

[6] Benn (1999), p. 34.

[7] Murphy (1972).

considerations of respect for autonomous persons and a duty to treat them as ends in themselves, we would be justified in withholding this treatment from psychopaths. Of course, we might also be motivated by non-Kantian considerations such as 'sympathy and virtue',[8] but we would not see psychopaths as *persons* or 'as having a full set of rights'.[9] Here we see why Benn describes this view as 'illiberal': if psychopaths do not have 'a full set of rights' then we might be justified in treating them in a way in which it would be unacceptable to treat other human beings. We might conclude, for example, that we are justified in pre-emptively detaining psychopaths on the basis of their high PCL-R scores. We would do well to think very carefully before endorsing a theory with implications of this kind.

Should we accept Benn's two conclusions? One reason to hesitate might be the thought that Benn's discussion seems to have drifted somewhat from the central ideas of Strawson's original paper. What Strawson is at pains to stress above all is the unavoidability, the 'given-ness', of the reactive attitudes. At one point, Strawson compares our attachment to the reactive attitudes to our attachment to the process of inductive reasoning. A theoretical conviction of the truth of determinism, he thinks, could no more force us to suspend in practice our reactive attitudes than a theoretical conviction of the impossibility of logically supporting induction could force us to stop practising induction. For Strawson, 'a sustained objectivity of inter-personal attitude, and the human isolation which that would entail, does not seem to be something of which human beings would be capable, even if some general truth were a theoretical ground for it'.[10] If this is right, simply withholding reactive attitudes towards psychopaths in the

---

[8] Benn (1999), p. 38.

[9] Ibid.

[10] Strawson (2008), p. 12.

way recommended by Benn's first conclusion would seem to be easier said than done.

However, while Strawson may be right that it would be 'practically inconceivable'[11] to suspend reactive attitudes towards everybody at all times, it is surely not the case that we are incapable of suspending reactive attitudes towards particular classes of people. History is full of examples of groups being systematically dehumanised, perhaps most notoriously the Jews (as well as other groups) in Nazi Germany. While this process began with the stirring up against the Jews of reactive attitudes such as anger and resentment, it ended with an attitude that is much closer to what Strawson describes as 'the objective attitude'. The Jews were seen as merely a problem to be dealt with. This comparison brings out very starkly the acknowledged illiberality of Benn's suggestion, though it should also be acknowledged that of course nothing akin to the Nazis' treatment of the Jews is automatically implied by that suggestion. However, the historical example shows that it does seem to be quite possible for people to persuade themselves to treat entire classes of people as less than fully-fledged persons.

There is, however, reason at least to doubt the legitimacy of Benn's move from his first conclusion to his second. This move rests on his endorsing a Kantian view of obligation:

> The guiding thought here is that morality is that set of principles that rational agents could freely agree to observe, on condition that everyone else observe them as well. And commitment to such an agreement entails a commitment to reciprocity. If certain individuals are incapable of understanding the need for reciprocity, or of entertaining the moral feelings that normally

[11] Ibid., p. 3.

> motivate its observance, then for that reason they exclude themselves from the agreement, and may be treated in ways in which normal people may never treat one another.[12]

In their response to Benn,[13] James Harold and Carl Elliott resist Benn's 'illiberal' move, implicitly also rejecting the Kantian view on which it is based. They make their point by drawing a distinction between moral *agents* and moral *patients*. Just because we accept that psychopaths are not moral agents, and they should not be held morally responsible, does not imply that they are not moral patients, to whom we have direct duties, and who might hold a set of rights. For example, we do not think of babies as being fully-fledged morally responsible agents, but we clearly do think of ourselves as capable of having direct duties towards them, and of them as having rights. Indeed, it is also very likely that we have direct duties towards animals. Even if we accept Benn's first conclusion, that psychopaths are not morally responsible agents, the proper way to treat them may be determined by more than 'sympathy and virtue'.

Patricia Greenspan[14] offers an alternative interpretation of Strawson's account, concluding that psychopaths are not excluded from the community of responsible agents in virtue of their social disconnectedness. Greenspan points out that psychopaths are not entirely incapable of reactive attitudes themselves:

> Psychopaths do not lack all varieties of interpersonal attachment… though their relationships are in many ways inconsistent and superficial. They do seem to establish at least deficient interpersonal relationships of the rough sort that Strawson described as based on mutual reactive attitudes. The

---

[12] Benn (1999), p. 38.

[13] Harold and Elliott (1999).

[14] Greenspan (2003).

> problem is that their reactive attitudes (and their awareness of others' reactive attitudes) apparently do not generate motivating attitudes, including guilt and other self-directed forms of blame, that manifest themselves as needed to inhibit impulses to act.[15]

We can see Greenspan's position as implying that reactive attitudes are not essentially, not always, communicative, as Benn insists they are. While it may be inappropriate to hold a reactive attitude towards a psychopath when that attitude is directed at influencing the self-motivating attitudes of its object, not all reactive attitudes need to be like this, and there may therefore be some attitudes that can still be appropriately held towards psychopaths. For Greenspan, such 'non-retributive' attitudes include 'reactive attitudes based on hatred rather than anger (e.g. disgust or contempt)',[16] but they might also include some forms of resentment. Returning to Strawson's original analysis, Greenspan argues that what justifies taking one of these attitudes towards someone is that their actions manifest 'bad qualities of will'. Because psychopaths *intend* their actions, and even intend the harm that those actions cause, we are justified in ascribing bad qualities of will to them, and therefore in holding non-retributive attitudes towards them. On this basis, and if we suppose that what makes psychopaths morally responsible or not is a matter of our being justified in holding reactive attitudes towards them, then psychopaths' responsibility can be said to be diminished, but not completely ruled out.

Thus, the dispute between Benn and Greenspan, and the question of whether psychopaths ought to be held morally responsible under a Strawsonian framework, would appear to hinge on whether we accept Benn's analysis of the

---

[15] Greenspan (2003), pp. 421-2.

[16] Ibid., p. 417.

'communicative' element of the reactive attitudes. Greenspan's view does leave room for non-retributive attitudes to be communicative in a sense, though not in the sense intended by Benn. For Greenspan, 'incorrigible' agents, including psychopaths, are not members of the 'moral community', and so retributive attitudes, which have the function of bringing about change in the object, are not appropriate. However, Greenspan describes attitudes such as contempt and disgust as 'sentiments of personal exclusion or dismissal from the moral community'.[17] Incorrigible agents merit these reactions precisely because they are impervious to the 'communicative' aspects of attitudes such as anger and indignation. We can still see these non-retributive attitudes as essentially communicative: by holding them towards incorrigible agents, we are communicating to them, and perhaps just as importantly to others, that they are to be excluded from the moral community. Because there is no assumption inherent in this act of communication that it will have any effect on the attitudes or behavior of their object, incorrigible agents, including psychopaths, can be apt targets for these reactive attitudes. Thus, in Strawsonian terms, they are at least partly responsible.

This plausible suggestion shows, I think, that the implications of a Strawsonian picture for psychopaths' responsibility are not as simple as Benn suggests. Psychopaths cannot be shown not to be responsible in a Strawsonian sense simply in virtue of their problems with holding reactive attitudes themselves. However, while I agree with Greenspan that Benn's argument is not enough to demonstrate that psychopaths are not morally responsible, I also do not think that Greenspan's argument is enough to show that psychopaths *are* morally responsible. The key point is that it is not clear that Greenspan's description of

---

[17] Ibid., p. 427. Presumably Greenspan has in mind a particularly moral kind of disgust here. People often feel disgust at non-human animals, including rats, snakes etc., but the emotion in these cases is not communicative in the way Greenspan has in mind.

the attitudes which it is intuitively appropriate to hold towards psychopaths is any more accurate than Benn's.

We can see the problem by recalling a point from Chapter 1 which applies to the link between ill will or indifferent regard and the reactive attitudes which Greenspan, drawing on Strawson, assumes. Psychopaths are indeed capable of ill will and indifferent regard towards other people, but what if it could be shown that they are not capable of understanding that other people are *due anything other than* ill will and indifferent regard? As I noted in Chapter 1, it is not clear as a matter of intuition that someone like this is morally responsible, and it is similarly unclear that they are apt targets of the reactive attitudes in which judgments of responsibility are expressed. Contempt and disgust are not obviously appropriate attitudes towards agents in this kind of predicament. We might be more inclined to feel sorrow at the hopelessness of the situation, or simply fear at what someone with this kind of psychology would be capable of doing. These are not Strawsonian reactive attitudes at all, and Greenspan could not use them as the basis for a case for moral responsibility, even of a limited kind.

The trouble is that the question of what it is *natural to do* – what attitudes we naturally hold towards people or groups of people – is essential to the Strawsonian justification for the practices and attitudes involved in holding people morally responsible, at least as I have interpreted it. Strawson's justification is a justification for those attitudes which we do, as a matter of fact, naturally hold towards people. It is because holding any other set of attitudes, if it became a general policy, would cause mental strain and the impoverishment of human relationships, that we should continue holding the attitudes we do. Quite a lot therefore rests on Greenspan's claim that the attitudes of exclusion – disgust, contempt and so on – are the natural attitudes to hold towards psychopaths.

I concluded in Chapter 1 that the Strawsonian emphasis on qualities of will provides only an incomplete and indirect answer to the question of whom, in what circumstances, we should hold morally responsible. I tried to show that there are some unusual cases where the presence of a bad quality of will, or the absence of the minimal good will that we expect as part of ordinary human relationships, does not satisfactorily explain why the person is, or is not, morally responsible. I went on to argue that the concept of responsiveness to reasons does a better job in this respect. Because psychopathy is clearly a very unusual case, I think we have reason to doubt Greenspan's conclusion that the matter of psychopaths' moral responsibility can be settled by the fact of their ill will. An argument based on responsiveness to reasons would put us on a surer footing in this respect. I will move on now to the question of whether such an argument can be developed; to what extent and in what ways psychopaths can be said to be responsive to reasons.

3.2 Psychopaths and reasons

The question that concerns us is whether psychopaths can be said to lack responsiveness to reasons in a way that is relevant to moral responsibility. One way in which an agent can lack (local) responsiveness to the reasons that bear on an act is if they are reasons of which she is not aware, and cannot reasonably be expected to be aware. So, imagine A gives a glass of wine to B, unaware that B is a recovering alcoholic. The two were not previously acquainted, and A had no way of knowing this fact about B. In the event, the temptation is too much for B, who drinks the wine and then several more. We might say that A has helped to set B back on the road to addiction, but we should not think of A as being morally responsible for this act. She was non-culpably ignorant of the circumstances in which she acted, and specifically of the facts about B's history which gave A a reason for refraining from offering her wine.

We can formulate a similar case in which the ignorance is the result of a fact about the psychological state of the person performing the act, rather than to

do with a special fact about the circumstances in which she acts. So, to return to an example which I have used previously, C is a paranoid schizophrenic and, because of her condition, thinks that D is out to get her. She harms D in what she wrongly believes to be self-defence. Now, assuming that C is not morally responsible for her condition or for the delusions that arise from it, it seems clear that, again, C is not morally responsible for the act of harming D. Owing to a confusion which is not her fault, she believes she has a reason to harm D, when in fact she does not, and she has all the usual reasons for refraining from harming D.

Is it possible that a psychopath's condition can cause her to be ignorant, or irrational, in a similarly responsibility-negating way? In his brief discussion of psychopaths in *Responsibility and the Moral Sentiments*, R. Jay Wallace focuses on the possibility that psychopaths might have a diminished capacity to 'engage in intelligent critical reflection'[18] which, if true, may provide support for the claim that they have diminished moral responsibility:

> It has been suggested that psychopaths lack the qualities of imagination and practical understanding required to bring common moral principles to bear in new cases; for instance, they often have great difficulty distinguishing between trivial and important moral concerns, and so lack the capacity to engage in intelligent critical reflection on moral issues. This severe impairment of the capacity for reflective self-control would set the psychopath apart from an 'ordinary' evil person… providing us with a reason for not treating the psychopath as a morally accountable agent.[19]

---

[18] Wallace (1994), p. 158.

[19] Ibid., p. 157.

For Wallace, this impairment in moral reasoning puts psychopaths outside the class of beings that have the kind of rational ability that is required for moral responsibility:

> The understanding required is a kind of participant understanding that goes well beyond the ability to parrot the moral principle in situations in which it has some relevance. What is needed, rather, is the ability to bring the principle to bear in the full variety of situations to which it applies, anticipating the demands it makes of us in those situations, and knowing when its demands might require adjustment in the light of the claims of other moral principles.[20]

It certainly appears to be the case, given the reports of clinicians and scientists, that psychopaths often have trouble engaging with moral principles in anything more than a very superficial way. Take for example the following remarks by three psychopathic inmates, reported by Robert Hare:

> When asked if he had any regrets about stabbing a robbery victim who subsequently spent three months in the hospital as a result of his wounds, one of our subjects replied, 'Get real! He spends a few months in a hospital and I rot here. I cut him up a bit, but if I wanted to kill him I would have slit his throat. That's the kind of guy I am; I gave him a break'....
>
> I was once dumbfounded by the logic of an inmate who described his murder victim as having benefitted from the crime by learning 'a hard lesson about life.'

---

[20] Ibid.

> 'The guy only had himself to blame,' another inmate said of a man he'd murdered in an argument about paying a bar tab. 'Anybody could have seen I was in a rotten mood that night. What did he want to go and bother me for?' He continued, 'Anyway, the guy never suffered. Knife wounds to an artery are the easiest way to go.'[21]

If taken at face value, these three remarks reveal individuals who profoundly misunderstand the moral principles which they are ostensibly employing. They appeal to considerations which any non-psychopathic person would immediately be able to recognise as completely irrelevant, and entirely miss considerations that appear to the reader as glaringly important. The impression is of people who are doing an impersonation of someone engaging in 'moral talk'. The impersonation is so poor, however, that it only reveals a profound and startling lack of understanding of how morality works even on a very basic level. Nor would there apparently be any motivation for these interviewees to exaggerate their misunderstanding. On the contrary, as prison inmates it would presumably be in their interest to convince officials that they understood what they had done and were remorseful. Remarks such as those listed above are only likely to reduce any chance of parole.

If these remarks are taken to be representative of psychopaths generally, it is easy to see how psychopaths might be thought to 'lack the capacity to engage in intelligent critical reflection on moral issues' to the extent that they might be excluded from moral responsibility. If moral responsibility consists in being able to respond consistently to the reasons that bear on one's actions, then it would presumably require some kind of minimal ability to recognise and apply moral principles, an ability apparently lacked by the psychopaths quoted above. Someone who can, apparently sincerely, justify having killed someone by

---

[21] Hare (1995), p. 132.

claiming that it taught them 'a hard lesson about life' appears to have a deeply flawed understanding of what reasons they have and ought to respond to. If it could be shown that this flawed understanding was itself not something for which the person was responsible, then this might count as exempting them from moral responsibility.

On the other hand, it is not clear that such remarks are indeed representative of psychopaths as a whole, as opposed to a small group of psychopaths who have not taken the time to gain a basic understanding of the moral principles that most people take to be important. Elsewhere, Hare quotes another psychopath, a man 'with the highest possible score on the psychopathy checklist':

> I've wasted a lot of my life. You can't get back the time.... I intend to live a much more slowed-down life, and give a lot to people that I never had myself. Put some enjoyment in their lives. I don't mean thrills, I mean some substance into somebody else's life. It will probably be a woman, but it doesn't necessarily have to be a woman. Maybe a woman's kids, or maybe someone in an old folks' home. I think... no, I don't think... I *know*, it would give me a good deal of pleasure, make me feel a whole lot better about my life.[22]

This man, who according to Hare had 'a horrendous criminal record' and 'had brutalized his wife and abandoned his children', clearly had enough understanding of moral principles and concepts to put together a fairly convincing speech on the themes of regret and the pleasure to be derived from being a positive influence on someone else's life. This perhaps bespeaks some understanding of what most people would take to be the reasons that other people's rights, interests and concerns present to them. Furthermore, both this

---

[22] Ibid.

psychopath and the others quoted above, as prison inmates, would fall into the category of 'unsuccessful psychopaths'. They might perhaps be expected to have a lower level of moral intelligence than those psychopaths who have avoided incarceration, people whose success may be partly attributable to them having convincingly assumed the mantle of ordinary, morally concerned agents. It might be that some psychopaths will have a diminished facility with moral principles to the extent that they will be excused to some extent from moral responsibility. However, the evidence that this is really a widespread feature among psychopaths is somewhat mixed. It would of course be preferable to base an argument on robust, quantitative evidence if such evidence were available.

3.3 The 'moral/conventional distinction'

Some evidence that has been taken by many to be of this kind comes from experiments into psychopaths' ability to understand the 'moral/conventional distinction'. In this research, carried out in the 1990s by James Blair,[23] which has been much discussed by philosophers, psychopaths and controls are judged on their ability to distinguish between two different kinds of judgment: 'moral' and 'conventional'. Ordinary subjects judge 'moral' and 'conventional' transgressions to be different from each other on three dimensions: (1) whether or not they are *permissible*, (2) how *serious* they are, and (3) whether or not they depend for their force on the word of some *authority*. 'Conventional' transgressions are typically more often thought to be permissible, are thought to be less serious, and are thought to be authority-dependent. In addition, when asked to explain *why* a given transgression is impermissible, subjects are more likely to adduce reasons relating to a *victim's welfare* if the transgression is 'moral' rather than 'conventional'.

[23] Blair (1995), Blair (1997).

A scenario which is supposed to be an example of a 'moral' transgression might involve a child hitting another child; a scenario which is supposed to involve a 'conventional' transgression might involve a child talking in class. Typically, subjects are more likely to judge that the 'moral' scenarios involve impermissible action (it is impermissible to hit another child). They also tend to believe that it makes a difference in the 'conventional' scenarios whether a relevant authority has given their assent to the act described (e.g. 'it's okay to talk in class if the teacher says you can'). In the 'moral' scenarios, however, they do not judge that the assent of authority makes any difference to the permissibility of the act (e.g. 'it doesn't matter whether the teacher says you can hit a child – it's still wrong'). They also judge the supposedly conventional transgressions to be less serious than the moral ones (hitting a child is a more serious transgression than talking in class). Finally, they tend to believe that the 'moral' scenarios are impermissible not so much for reasons having to do with a victim's welfare (hitting a child harms the child) so much as for other reasons (e.g. it's not fair if one child talks in class when the others can't).

Blair's experiments involved presenting these scenarios to a set of psychopaths (diagnosed using PCL-R) and non-psychopathic controls. Psychopaths, in comparison to controls, were found to be significantly less likely to be able to distinguish between the two types of case on all three dimensions listed above (permissibility, seriousness and authority-dependence). Psychopaths were also less likely than controls, when asked why 'moral' transgressions were impermissible, to produce explanations that appealed to the victim's welfare.

In evaluating these experiments, it is worth noting that the traction they have had with philosophers probably has more to do with the philosophically interesting nature of their conclusions than with the robustness of their empirical base. The original 1995 study involved ten psychopaths and ten non-psychopathic controls. In 1997 there was a follow-up study with children with psychopathic tendencies, which involved 16 such children and 16 controls. Only

in the second study were the results completely as Blair predicted, i.e. the children investigated heard about supposedly moral transgressions and interpreted them similarly to how non-psychopaths judged 'conventional' transgressions. In the 1995 study the results were the other way around: the subjects interpreted all transgressions, both 'moral' and 'conventional', as 'moral'. Blair explained the discrepancy by hypothesising that the imprisoned psychopaths would be motivated to appear 'virtuous' in the hope of securing improved treatment, and would therefore tend to overstate the perceived severity and universality of transgressions in the experiment. He therefore took the latter results to be more trustworthy and concluded that psychopaths think that all transgressions are 'conventional', rather than that they are all 'moral'. This may be correct, but it is interesting to note the slim empirical foundations on which this auxiliary hypothesis was built.

It is also worth noting that, as Vargas and Nichols point out, Blair's experiments did not show that psychopaths consistently 'miss every case of the moral/conventional task.'[24] Rather, the psychopaths in the studies tended to make the relevant distinction less consistently than non-psychopathic controls. It is not clear, therefore, how we ought to apply the conclusions of any arguments built on these empirical foundations. Given that even most of the psychopathic subjects appeared able to make the distinction in some cases, should we be looking for differences among the cases used, and only excuse psychopaths in scenarios that are relevantly similar to those in which they have proved themselves unable to recognise a moral/conventional distinction? In Vargas and Nichols' words, 'experiments on psychopathologies usually produce data that is less ordered than we might hope for', and therefore, 'it is misleading to say that… psychopaths *cannot draw* the moral/conventional distinction'.[25] In

[24] Vargas and Nichols (2007), p. 158.

[25] Ibid., pp. 157-8.

particular, while psychopathic subjects were indeed less likely than controls to appeal to reasons relating to a victim's welfare when asked to explain why moral transgressions were impermissible, several of them did in fact make appeals of this kind (five psychopaths in the original study, compared to nine non-psychopathic controls).[26] It should be noted too that studies showing any difference in actual verdicts given to moral dilemmas are quite difficult to find. As we saw in the 'neuroscience' section in the previous chapter, psychopaths have been shown to give broadly the same type of answers in response to a range of moral dilemmas as non-psychopathic controls.

Moreover, the emphasis on authority-dependence in the 'moral/conventional distinction' as tested in the experiments is problematic in at least two ways. Firstly, authority-dependence cannot be the basis for a distinction between moral and conventional transgressions, at least on any ordinary understanding of the word 'conventional', because many conventions simply have nothing to do with authority. For example, it is a useful convention that people who want to get onto a train wait for all passengers to get off first. This convention is not supported by any authority. Conventions frequently (perhaps even usually) come to exist as a kind of mutual understanding between peers, and are sustained by their usefulness in, say, avoiding inconvenience or social awkwardness, and not by the efforts of an authority.

Secondly, if psychopaths fail to recognise that some transgressions are not authority-dependent, there is evidence to suggest that they may not be unusual in this respect. In Kohlberg's famous experiments into moral development most subjects were found to reside in his 'Stage 4', in which morality is largely authority-based.[27]

---

[26] Blair (1995), p. 18.

[27] Kohlberg (1981).

However, while it is probably true that Blair's own interpretation of his experimental results can be legitimately questioned, it is nonetheless quite likely that something interesting is going on here. There is a genuine distinction between different types of transgression which was somewhat reliably picked up by the normal subjects, but less reliably picked up by the psychopaths. Some transgressions are such that most people would take them to rely on the strictures of an authority figure for their normative force, while other transgressions are generally taken not immediately to depend on a specific authority in the same way, though it may be that, pressed for an 'ultimate' explanation of their provenance, the best most people can offer is an appeal to authority of some kind, which would explain Kohlberg's finding. This distinction, exemplified by the two classroom situations described above, is real enough, though no doubt 'moral/conventional' is not the best label to apply to it. If some psychopaths experience some difficulty with moral judgments in the way described, then it is perhaps reasonable to suppose that these psychopaths' responsiveness to reasons might be compromised to some extent.

Walter Glannon has attempted to use Blair's studies, as well as other clinical studies involving psychopaths, to ground an argument for the view that psychopaths have limited, but not complete, moral responsibility, focusing on their capacity to be motivated by different kinds of reason. Glannon notes that psychopaths have been shown in studies to be adept at using aggression in instrumental, as well as in reactive ways. (Instrumental aggression is defined as the controlled use of aggression as a tool to manipulate others, usually through intimidation, whereas reactive aggression is more impulsive, less focused, and stems from emotions such as anger.) If psychopaths can use aggression in this way, Glannon argues, this shows that they are capable of recognising and reacting to instrumental reasons. Now, assume for the sake of argument that we think that the message we should take from the moral/conventional distinction experiments is that psychopaths think that morality is a form of

convention. Although we have reason to think that psychopaths are not motivated by moral reasons, we know that they can be motivated by instrumental reasons, and we know that they can recognise conventional reasons. Furthermore, the evidence from Blair's experiments does not give us reason to think that they could not be motivated by conventional reasons. If they can be so motivated, they could therefore presumably be motivated by those reasons that they take to be conventional, but which are in fact moral. This, argues Glannon, would be 'normatively equivalent' to being motivated by moral reasons *as* moral reasons, since in either case the subject is following rules which are 'designed to inhibit moral wrongdoing and enable an individual to refrain from performing actions that harm others'.[28] Moreover, in Glannon's phrase, 'the content of the subject's mental states'[29] would be the same in either case. If responsibility is assigned on the basis of motivational states, and of the practical aspects of one's actions, then whether one recognises the rules that one violates as moral or conventional cannot make any difference to whether one is morally responsible. Therefore psychopaths, insofar as the above description applies to them, are morally responsible.

Even if we were to accept the notion of 'conventional' reasons that is implicit in Blair's experiments, and that psychopaths can recognise, and be motivated by, moral reasons that they take to be conventional in this sense, it is difficult to see why Glannon thinks this would be 'normatively equivalent' to recognising and being motivated by moral reasons as moral reasons. To bolster this claim, Glannon emphasises both the motivational capacities of psychopaths and the content of their beliefs, arguing that both of these are similar enough to their equivalents in non-psychopaths to ground moral responsibility. Taking motivational capacities first, one might think that the psychopath (as described

---

[28] Glannon (2008), p. 163.

[29] Ibid.

by Glannon) has motivational capacities that are equivalent to those of non-psychopaths in the sense that both are capable of being motivated by reasons stemming from transgressions, although one takes those transgressions to be moral and the other 'conventional'. But this seems too weak to ground *moral* responsibility. It is not clear why someone who can be motivated by what they take to be *conventional* and not *moral* transgressions is nonetheless *morally* responsible.

Turning to the content of psychopaths' beliefs, Glannon claims that 'their capacity to recognise moral reasons provides them with enough reflective self-control to realise that they *should not* perform actions that are harmful to others and to refrain from performing them'. However, this claim could be questioned. If I believe that an action is outlawed by some authority, in what sense do I believe that I *should not* perform it? Imagine, for example, I am told by a police community support officer not to walk on the grass in the park. In what sense do I believe that I should not walk on the grass? It is quite possible that I believe that I *morally* should not walk on the grass. However, for this to be the case, I would presumably need to take the authority of the police community support officer to have moral legitimacy. That is, I must think that I have moral reasons to follow the instructions of a person with this kind of authority. Without this, we are left with a sense of 'should' that is prudential, rather than moral. I should not walk on the grass because the police community support officer might punish me in some way, or perhaps simply because they would think ill of me, and I do not like people thinking ill of me.

There are also situations in which even this prudential 'should' would not apply. What if the police community support officer sees someone else walking on the grass, and tells me to punch that person? In these circumstances, there is no sense in which I should perform the action. Indeed, it seems clear that I *should not* perform the action, morally.

Now, assuming that the psychopath has no way of recognising the moral legitimacy or otherwise of an authority – and it is hard to see how she would recognise this if we take the conclusions of the moral/conventional experiments at face value, i.e. if we accept that she cannot tell the difference between moral and conventional considerations – then she is left in the position of being unable to distinguish between situations like the one above and situations in which she should, morally, comply with the authority's dictum. It seems, then, that while the psychopath might *believe* that she should, prudentially, comply, we have no reason to think that she can *know* that she should comply, in any sense that is strong enough to ground moral responsibility.

In the latter of his two papers on this subject, Glannon offers an alternative description of the content of the psychopath's mental state which he takes to be enough to ground moral responsibility. The psychopath, claims Glannon, 'could be capable of recognizing that the actions could not be justified by any normative reason, and on this basis he could be capable of recognizing that the actions were wrong'.[30] Again, however, there is reason to doubt Glannon's conclusion, at least if we take 'wrong', as I assume we must, to mean '*morally* wrong'. To recognise that an action is morally wrong, we might think, requires not only the recognition that it is unjustified, but also the recognition that for it not to be wrong would require it to be morally justified. If the psychopath is not capable of taking moral reasons to apply to his own choices, he may not be capable of recognising this. If he is only capable of recognising *conventional* reasons as applying to his choices, then he may be capable of recognising actions which contravene those reasons as being 'frowned upon', or 'not *comme il faut*', or something along those lines. If he is capable of recognising only reasons stemming from the word of some authority, then he is capable of recognising

[30] Ibid.

that they are *outlawed*. Neither of these things amounts to the same as recognising that they are wrong.

Neil Levy is another philosopher who has used the results of the 'moral/conventional' experiments as the basis for an argument that psychopaths have diminished responsibility. This description is in fact compatible with Glannon's description of psychopaths as having 'partial responsibility'. However, Levy's emphasis is on showing what components of moral responsibility are lacked by psychopaths, rather than what components they supposedly retain.

Levy interprets the moral/conventional experiments as showing that psychopaths believe that 'harms to others [are] wrong *only because* such harms are against the rules…'

> For them, stealing from, or hurting, another is no more wrong than, say, double-parking or line-jumping. But the kind and degree of wrongness, and therefore blame, that attaches to infringements of the rules is very different, and usually much less significant, than the kind and degree attaching to moral wrongs. For psychopaths, all offences are merely conventional, and therefore – from their point of view – none of them are all that serious. Hence, their degree of responsibility is smaller, arguably much smaller, than it would be for a comparable harm committed by a normal agent.[31]

If psychopaths think that all transgressions are wrong only because they are against the rules, thinks Levy, then they cannot, when they transgress, be

---

[31] Levy (2008)

thereby expressing the kind of ill will towards others that grounds attributions of moral responsibility.

I argued in Chapter 1 that the presence of ill will gives at best only an incomplete basis for attributions of moral responsibility. Responsiveness to reasons, I argued, is what we should be looking for in deciding whether a given person is morally responsible in a given context. It is plausible to suppose, however, that psychopaths, described in Levy's terms, might lack responsiveness to reasons. The kinds of reason presented by an infraction of 'the rules', while pressing, are much less pressing than the kinds of reason that are salient to normal agents when considering an action that will cause harm to others. If psychopaths can indeed only recognise and respond to these kinds of reason, then they are missing an important piece of the normal agent's psychological repertoire, and their moral responsibility will indeed be diminished.

On the other hand, the moral/conventional experiments do not show that psychopaths *cannot* understand this distinction, only that some of them *do* not, some of the time. The experiments concern a distinction based on the degree of wrongness of certain transgressions, whether they are impermissible, whether they are dependent on authority for their force, and on what it is that makes them wrong. The problem for any argument that would seek to attribute a degree of moral responsibility to psychopaths based on their misunderstanding of these features is that, while they are certainly fundamental to ordinary moral thought, they are all quite capable of being explained to someone who is rational. What, for example, if you explained to a psychopath that a child hitting another child was wrong, not because it was outlawed by some authority, but directly because it caused harm to the victim? Do we have any reason to suppose that the psychopath would be incapable of understanding this? Not on the basis of Blair's experiments, or of any other evidence of which I am aware. There is no evidence that I know of to show that psychopaths do not understand what harm is, for example, or that harm can be a wrong-making

feature of an action. Now, if they are capable of understanding this – and again, the moral/conventional experiments can only show a lack of understanding, not a lack of the capacity for understanding – then why should we think that they lack moral responsibility?

We do not ordinarily think of a lack of understanding as being excusing, if the person concerned is capable of acquiring the relevant understanding. To see this, imagine someone who, somehow, has failed to grasp the concept of property. Perhaps they were brought up in a community that had no notion of property, allowing free use of all objects to any member of the group. If this person, now at large in mainstream society, went about taking other people's things, how would we react to this? I think, while we might excuse them the first few times on the basis of their lack of understanding, we would in the long term expect them to acquire an understanding of the concept of property and to abide by its normative aspects. Confusion over the proper use of moral concepts can only be excusing to some extent, as long as the person concerned is capable of overcoming that confusion. Ultimately, we tend to think that people have a duty to acquaint themselves with the way such concepts are used, and as a result we hold them responsible for their actions which are, or ought to be, informed by such concepts. If all we know about psychopaths is that they are confused about the proper use of concepts such as justification or authority-dependence, and we do not know that they are incapable of overcoming this confusion, then they may be in exactly this situation.

The above discussion of the moral/conventional experiments, and of Glannon and Levy's ideas, then, leaves us with two general conclusions about psychopaths and moral responsibility. Firstly, if some psychopaths can be partially exempted from moral responsibility because they exhibit difficulty in handling moral concepts, then this is unlikely to represent a clean distinction between psychopaths and non-psychopaths. An argument based purely on the moral/conventional experiments could exempt only those psychopaths who

suffer from confusion in using the concepts involved in morality – confusion about their degree of force, authority-dependence, and so on, and it would only exempt them in situations where this confusion actually affected their judgments. Secondly, an argument of this kind would only excuse psychopaths who lack not only understanding, but the capacity for understanding, and the moral/conventional experiments cannot show that any psychopaths lack the capacity for understanding.

Gwen Adshead notes that the 'rational amoralist', which she describes as 'the typical layperson's psychopath', is 'almost nonexistent in clinical samples'.[32] However, she does go on to speculate that the more rationally unimpaired type of psychopath might be more common in the general population than in the prisoners and psychiatric patients who form the basis of clinical samples. Here, perhaps, we see again the distinction between 'successful' and 'unsuccessful' psychopaths. In any case, we must accept that in psychopaths we are dealing with, in Adshead's words, 'a highly heterogeneous group of people'.[33] Even if we accept that difficulty in making the moral/conventional distinction is fairly widespread among psychopaths, it remains the case that it is a distinction that can be explained, and that presumably could be understood intellectually by someone who did not feel the emotional force of moral transgressions as opposed to 'conventional' ones. It seems unlikely – and the experiments themselves do not show – that psychopaths are congenitally unable to understand that most people take transgressions that involve directly harming other people to be wrong *just for that reason*, and not because they are against the rules, or because some authority says they are wrong, or for another extrinsic reason of this kind. If – as I strongly suspect – there is a class of psychopath who is perfectly capable of understanding this, and indeed does

---

[32] Adshead (1999), p. 43.

[33] Ibid.

understand it, then these psychopaths would at least have the ability (to revisit Wallace's formulation) to 'bring the principle to bear in the full variety of situations to which it applies, anticipating the demands it makes of us in those situations, and knowing when its demands might require adjustment in the light of the claims of other moral principles', but might still not see moral principles as particularly important *for them*. Arguments such as Levy's have nothing to say about these people.

It would be much more satisfactory if we could develop an account of the psychopath's moral failings which could explain the results of the moral/conventional experiments, but which itself had clearer, and broader, implications for their moral responsibility, understood in terms of responsiveness to reasons. It is my contention that such an account can be developed by examining psychopaths in terms of how, and to what, they ascribe value. In the next two sections, I will explore this question, and how it relates to psychopaths' responsiveness to reasons.

3.4 Imperviousness to reasons

My overall aim in this thesis is to show that psychopaths lack moral responsibility for certain of their actions (as well as attitudes, emotions and states of affairs). I am working within a framework of moral responsibility as consisting in responsiveness to reasons, having indicated in Chapter 1 why I think this is the most fruitful and plausible way of thinking about moral responsibility. In the last chapter, I indicated what I take the term 'psychopath' to denote, both in terms of their psychological deficits, which I argued are primarily emotional, and in terms of their behavioural profile. Fulfilling my overall aim, then, will involve showing that there are some reasons to which psychopaths, defined in the way I have defined them, are unresponsive. On the other hand, there are clearly a great many reasons to which psychopaths *are* responsive. For example, a psychopath who is hungry will certainly recognise this as a reason to eat (or if they do not, it will not be because they are a

psychopath that they do not). A psychopath who steals an object will typically have done so in response to some reason, or to what they take to be a reason; the simple fact that the object is desirable in some way would constitute a reason of this kind. There are controversies surrounding how we should think of the reasons that any agent has, but we can surely say with confidence that many of the reasons that apply to non-psychopathic agents will also apply to psychopathic agents, and that psychopaths will be equally responsive to many of these. So how are we to pick out the particular class of reasons to which psychopaths are not responsive?

Whatever kinds of reason psychopaths do have trouble with, it is noteworthy that this trouble consists not in their being *oblivious* to such reasons, but in their being in some sense *impervious* to them. Carl Elliott describes the difficulty of characterising this distinction:

> What [the psychopath] does know is what other people think is wrong. He knows what other people feel guilty about, which actions will be punished, which will be rewarded, when to lie and when to tell the truth. In fact, he often knows all these things well enough to be able to manipulate, flatter and bamboozle people with something approaching genius....
>
> On the other hand, the psychopath seems to lack any sort of deep engagement with morality. His knowledge seems limited to morality's most shallow and superficial features. This sort of deficiency can be difficult to describe, a bit like describing a person who is able to say in the most technically correct, clinical terms why Duke Ellington was the greatest jazz composer of the century, yet who is also clearly and unquestionably tone deaf.[34]

---

[34] Elliott (1996), pp. 77-8.

Elliott's description, I think, applies to many psychopaths, though perhaps not to all. The discussion around the moral/conventional experiments appears to show that some psychopaths, to adapt Elliott's simile, might be like someone who has a deficient understanding of music theory, as well as a deficient capacity to appreciate music. Nonetheless, I maintain that the most promising strategy for someone looking for exempting conditions is to see if these can be found in the latter deficiency, not the former.

The type of psychopath with whom I am primarily concerned, then, is one who understands the full range of reasons which normal agents take to apply to their choices, at least in the sense that he understands that *these are the reasons that normal agents take to apply to their choices*. He has no trouble understanding that some reasons are taken to be more forceful than others, or that the fact that an action causes harm to another person is taken to constitute a reason against performing that action, or that the reason against performing an action that is constituted by the fact that the action will cause someone physical harm, for example, cannot typically be nullified by the removal of a diktat from authority outlawing that action. What this putative psychopath lacks, rather, is the understanding that these reasons genuinely are reasons that apply *to them*. They might, and typically would, understand that other people take these reasons to apply to themselves and also that other people take these reasons to apply to the psychopath. But, somehow, these psychopaths do not see these reasons as applying to themselves at all. If they claim to see this, it is only as a means of disguising themselves as ordinary people, and their actions and the way they deliberate about those actions reveal the insincerity of the pretence.

The psychologist Martha Stout presents a series of case studies – some real, some fictional – to illustrate the inability to develop 'real' human relationships that she takes to be a central feature of the condition. One of the real cases concerns 'Luke', the husband of one of Stout's patients, 'Sydney'. Quoting extensively from interviews with Sydney, Stout tells the story of how, over

several years of marriage to Luke, it had gradually become apparent to Sydney that Luke had never felt any genuine love or affection for her, but had married her purely for the easy and comfortable life he was able to lead by taking advantage of her hefty salary and luxurious home. For years he feigned depression and deliberately encouraged Sydney and her friends to pity him in order to avoid difficult questions about why he never worked or helped around the house, and in fact spent most of his time lying by their swimming pool. Even the child they had together was, to Luke, just another means of manipulating Sydney into allowing him to stay:

> For Luke, societal rules and interpersonal expectations existed only to serve his advantage. He told Sydney that he loved her, and then went so far as to marry her, primarily for the opportunity to ensconce himself as a kept man in her honestly earned and comfortable life. He used his wife's dearest and most private dreams to manipulate her, and their son was an aggravation he moodily tolerated only because the baby seemed to seal her acceptance of his presence. Otherwise, he ignored his own child.[35]

Luke's goal was a comfortable and easy life for himself. He viewed other people – even his own child – merely as tools which he could use to attain that goal.

I have included a brief description of this case study here because it illustrates the fact that what psychopaths are motivated by is not always great power or wealth, and nor is it always the thrill of hurting other people. Furthermore, the abilities upon which they can draw are also very varied. Contrary to the way they are often represented in popular fiction, psychopaths are not necessarily particularly intelligent or dynamic, although they do tend to share an ability to

---

[35] Stout (2005), p. 115.

manipulate and charm people, perhaps due to their profound lack of social anxiety. As Stout observes, aside from the core characteristics of psychopathy – the profound emotional absence – psychopaths have the normal range of human motivations and capacities:

> People are not all the same. Even the profoundly unscrupulous are not all the same. Some people – whether they have a conscience or not – favour the ease of inertia, while others are filled with dreams and wild ambitions. Some human beings are brilliant and talented, some are dull-witted, and most, conscience or not, are somewhere in between. There are violent people and nonviolent ones, individuals who are motivated by blood lust and those who have no such appetites.[36]

My interest here is in psychopaths' capacity for practical reason. Specifically, what reasons, or types of reason, do psychopaths take seriously when they are deciding how to act? The things that psychopaths might be seeking to gain through their actions include comfort, wealth, power, pleasure, the cheap thrill to be gained from 'putting one over' on someone, etc. It is clear that psychopaths see the objects of these desires as reason-giving. My task here is to identify that range of objects that psychopaths do not see as reason-giving – not reason-giving *for them* that is, since it may be that they see them as reason-giving for others.

One very obvious and much-remarked fact about psychopaths is their selfishness. Writers emphasise their complete unconcern for the needs and interests of other people. In Hervey Cleckley's words, 'the psychopath is always distinguished by egocentricity… usually of a degree not seen in ordinary people

---

[36] Ibid., p. 2.

and often little short of astonishing'.[37] The impression one gets from reading Cleckley's case studies (as well as those presented by Stout and Hare) is that non-egocentric considerations are not merely outweighed in the psychopath's deliberation. Rather, it appears that such considerations simply do not occur to psychopaths as something of which they should take account or by which they should be motivated.

There are several ways in which non-egocentric considerations enter into the practical reasoning of normal agents. One of these is via obligations. If I believe that I have an obligation, whether this is to another person, to a group of people, to some entity or group of entities other than people, or to no entity in particular, I will take this to present me with a reason for action, a reason that should be taken into account in my practical reasoning. As has been discussed by Joseph Raz, T.M. Scanlon and David Owens among others, reasons stemming from obligations enter into practical reasoning in a distinctive way, and not just as additional reasons to be weighed up alongside whatever other reasons are operative in the situation. According to Owens's analysis, there is a sense in which an obligation 'takes the matter out of your hands:

> ...it is no longer up to you to judge whether doing the required thing would be best, all things considered. An obligation does not shape practical deliberation solely by constituting a point in favour of fulfilling it... it also constrains or limits your practical deliberations.'[38]

Imagine I make a firm promise to you to sell you my car, and then I receive a higher offer from someone else. In this situation, I have an obligation to sell you my car, and because of this I also have a reason to do so. You might think

---

[37] Cleckley (1941), p. 395.

[38] Owens (2008), p. 404.

that I also have a reason to sell my car to the person who has made the higher offer – after all, I would end up better off if I did – and what I should do is to weigh up this reason against the reason stemming from my obligation, as well as any other reasons that are relevant to my decision. I should then act on whichever reasons prove to be more important. However, this would not be true to what Owens calls 'the phenomenology of demand'.[39] There is something wrong about the idea that I should even take into account the larger offer made by the second person. To take this into account would be to fail to take seriously the fact that I have, not just a reason to sell you my car, but an obligation to do so.

Exactly what constraints having an obligation should place on one's practical reasoning (or, alternatively, places on one's practical reasons) is a matter of disagreement. For Scanlon,[40] an obligation excludes certain apparent reasons entirely from the set of applicable reasons. For Owens (building on Raz's analysis) an obligation does not constrain the set of reasons that apply, but it does provide a second-order reason to exclude certain first-order reasons from our deliberations about how to act. Owens's position has the advantage of being able to explain the fact that the reasons in question continue to apply to other things, such as for example my attitude to the decision I have made, or the circumstances surrounding it. For example, the reason stemming from my receipt of a higher offer of the car might cause me to regret having made you a promise in the first place, and it might be perfectly appropriate for me to have this attitude – the importance of the reason is not in this instance nullified by the fact that I have an obligation.

---

[39] Ibid.

[40] Scanlon (1998), pp. 156-7.

Whatever the correct way to analyse the role of reasons stemming from obligations in the practical deliberation of normal agents, it is reasonable to suppose, I think, that the 'phenomenology of demand' which is appealed to in this discussion would be one that would be alien to a psychopath. Psychopaths as described by clinicians and psychologists seem to be impervious to the force of obligations. While this is not a claim for which, to my knowledge, quantitative evidence exists, the evidence from case studies is compelling. For example, many if not all of the psychiatric patients described by Cleckley were admitted to his hospital after squandering the trust of those close to them. Cleckley's book is full of descriptions of patients whose family and friends had repeatedly lent them money, bailed them out of debt, secured employment for them, or vouched for them in situations in which they had got into trouble with the law. Most non-psychopaths, we can surely assume, would have felt the force of deep obligations to these people who had, at considerable expense or risk to their reputation, offered valuable help. The psychopaths in the case study, however, appeared entirely impervious to any sense of obligation, happily squandering money lent without seemingly having any intention of ever paying it back, throwing away jobs in spectacular fashion, and so on. These psychopaths appear to be the polar opposite of normal agents as described by Owens: instead of the reasons generated by obligations causing other reasons to be excluded from their practical deliberations, it appears rather that reasons stemming from obligations are themselves entirely excluded.

I have discussed reasons stemming from obligations as a kind of non-egocentric consideration. However, it is worth noting that it is possible to have an obligation to oneself, and the reasons stemming from this obligation would presumably be egocentric ones in a sense. Would psychopaths be likely to recognise obligations to themselves? This is a difficult question to answer. Perhaps an example will help. Imagine an overworked mother who spends her time juggling competing demands from her grown-up children, while also

holding down a difficult full-time job. One of her children calls and asks if she would be able to give him and him a lift to the airport on Sunday night at midnight, so that he can go on holiday. She has an extremely difficult week ahead, starting with a very important meeting at 8.30 on Monday morning. It is part of her character to want to make sacrifices to help her children, and she really wants to say 'yes', but she recalls some advice from a friend who suggested that she had an obligation to put her own interests first in situations like these, because otherwise she will be acting unfairly towards herself. Now, we can imagine the overworked mother feeling the force of this obligation, and that the reason stemming from it would weigh on her decision regarding whether to agree to give a lift to her son.

To remain true to 'the phenomenology of obligation', this scenario must be distinguished from, firstly, the scenario in which the mother simply ignores the son's wishes and thinks only of herself and, secondly, the scenario in which she weighs her son's potentially stressful trip to the airport against her own busy week and job interview and decides that her own need is greater. The (believed) reason that is relevant to the current discussion is that arising from her (believed) obligation to herself. We might imagine that she begins the process of weighing competing reasons and then, remembering her friend's advice, decides that what she sees as her obligation to herself trumps any competing reasons in the case.

Now, it is hard to imagine a psychopath deliberating in anything resembling this fashion, simply because, as I have been arguing, the non-egocentric reasons represented by the needs and interests of other people, such as the son in the above case, do not weigh on the deliberations of psychopaths in the first place, and therefore there would be no need for them to be trumped by reasons arising from obligations in the way described. This, however, is not evidence that psychopaths do not recognise obligations to themselves. It may be that psychopaths do recognise such obligations, such that, if there were competing

reasons that presented themselves to the psychopath as compelling, they would be trumped by the obligation-derived reasons in the way recognised by Raz, Scanlon and Owens. It is just that, being impervious to non-egocentric reasons, there are never any other reasons to trump.

A more effective test case, therefore, would be one in which a psychopath must weigh reasons derived from an obligation to herself against competing *egocentric* reasons. If the obligation-derived reasons appear to trump the competing reasons in this case, then we have evidence that psychopaths can recognise some forms of obligation at least. Cases of this kind can be difficult to formulate, let alone to test empirically. Perhaps the most easily imagined type of case would be one in which a person has some long-term goal in mind, and must sacrifice immediate gratification to attain that goal. Perhaps we do sometimes regulate our decision-making in this type of case through the formulation of obligations. Imagine I am a student who has resolved to study every Thursday night, because I attend a lecture on Thursdays which I find particularly difficult to understand. I am worried that if I do not spend time reviewing the contents of the lecture on Thursday night I will not, over the course of the module, absorb sufficient knowledge about the subject. I come to see this Thursday night study as an obligation; as something that presents a particularly compelling reason, one which should in the normal run of things trump any competing egocentric reason which might come into conflict with it, for example the reason presented by a particularly good band playing this Thursday in the student union.

It is much more difficult to state, especially based on a mere thought experiment, whether a psychopath would be capable of feeling the force of an obligation of this kind. As I noted in Chapter 2, different experts have different views on the extent to which psychopaths are capable of subordinating their short-term desires in the pursuit of long-term goals. For Robert Hare, the

inability to do this is one of psychopathy's central features. For Kevin Dutton,[41] on the other hand, many psychopaths are attracted to, and can be extremely successful at, careers such as surgery and law which require dedication and self-sacrifice. Even if the latter characterisation is the correct one, it would be impossible to know whether this self-sacrifice is mediated by the formation of obligations; it could just as easily be the case that the reasons stemming from long-term goals are weighed against the reasons stemming from short-term goals, and simply found to be more compelling. The question of whether psychopaths are capable of recognising obligations to themselves, therefore, is one on which we must remain agnostic.

Aside from through obligations, there are of course other ways that reasons stemming from other people – from their rights, interests and concerns – enter into the practical reasoning of ordinary agents. We could designate this broader class of reasons the class of supererogatory altruistic reasons – reasons that are not egocentric but which do not arise because of any obligations the agent has. For example, while walking in the town centre you see someone who is obviously new in town and having trouble finding their way around. You stop and ask if you can help. You clearly need not have believed yourself to have an obligation to help, but nonetheless, the fact that they were lost and in need of help gave you, so you believed, a reason to help them. I think we can say with some confidence that psychopaths do not recognise reasons of this kind. The behaviour of psychopaths as they are described in the scientific literature is always, ultimately, directed at their own egocentric ends.

The word 'ultimately' is important here, however, because it should be acknowledged that psychopaths are apparently capable of recognising other people as presenting a certain kind of reason for them, namely instrumental reasons. Psychopaths are perfectly capable of acting in another person's

---

[41] Dutton (2012).

interests in order to fulfil their own desires. One can certainly imagine situations, for example, in which a psychopath in a work situation might be helpful towards someone in a position of power in order to attain influence with that person and thereby increase their own access to power. In the Martha Stout case study discussed above, the psychopath Luke began his relationship with Sydney by treating her extremely well, and being highly attentive to her needs and interests. In both of these cases, however, and in others in which psychopaths apparently treat others in a way that would appear to imply that they recognise them as presenting reasons for action, the kind of reason that they do recognise is purely instrumental. The other person is valuable to the psychopath as a means of attaining their own goals, and the attentiveness is, ultimately, merely a form of manipulation. If the psychopath could achieve the same outcome more easily by using force or violence, they would be equally willing to take this approach.

As well as reasons derived from other people, the conclusion that psychopaths are impervious to non-egocentric reasons applies equally to other types of reason that are not egocentric in the way I have been discussing, but which also do not derive from any particular person or group of people, or even from people at all. One such type of reason would be the type of reason presented by animals. It should be noted that the idea that animals present us with reasons which guide our behaviour need not be derived from any potentially controversial thesis, such as that animals have rights, or are persons, etc. Presumably any philosopher on any side of the various debates in animal ethics, would at least agree that, presented with a kitten, and in a situation in which my actions would never be known to anyone else and which I could never suffer any negative consequences in respect of my actions towards the kitten, I nonetheless would at least have a reason not to torture the kitten to death. A true psychopath would plausibly not recognise such a reason as applying to them.

There are also, of course, a great number and variety of reasons that are neither egocentric nor arising directly out of consideration for other people or animals. This set of reasons would include reasons stemming from abstract ideas such as justice or from more concrete entities which are still neither human nor animal, such as the environment. Again, it seems unlikely that psychopaths, impervious as they are to reasons stemming directly from the rights, interests and concerns of others, will be any less impervious to these more abstract moral considerations.

A final category of reasons which deserves attention is that of aesthetic reasons. Reasons can be derived from a number of different aesthetic considerations. For example, the fact that something is aesthetically valuable – whether a painting or sculpture, or a musical performance, or something with natural beauty such as a tree or an unusual rock formation – would normally be taken as constituting a reason not to destroy or deface that thing. But this same fact might also constitute a reason to look at, listen to, or otherwise experience it. Often people can take these reasons to be quite powerful, as when someone expends considerable time, money and effort to travel to the city where a particular artwork is kept. Artists, of course, are also driven to create by aesthetic reasons: the fact that the artwork an artist intends to create promises to have aesthetic value presents itself to the artist as a reason for the artist to work at creating it. It might also be taken by those close to the artist to constitute a reason for them to indulge the artist's difficult personality and behaviour, or to provide an environment in which the artist can create without the distractions of everyday life. More prosaically, aesthetic reasons are prominent among the reasons people are responsive to when choosing what clothes to wear, or what house to live in, or where to go on holiday. There are, in short, countless ways in which aesthetic considerations affect our choices and behaviour.

Are psychopaths impervious to aesthetic reasons? This is a fascinating question to which the answer is unfortunately unclear. Several fictional characters who

are either supposed psychopaths or have psychopathic traits spring to mind here, including Thomas Harris's murderous aesthete, Hannibal Lector, and the Beethoven-loving narrator of Anthony Burgess's *A Clockwork Orange*, Alex. However, while the aesthetic appreciation exhibited by these characters serves a literary function, it may not necessarily constitute a realistic depiction of actual psychopaths. The clinical and scientific literature offers no clues here and so, again, we must remain agnostic.

3.5 The role of value

What is it that unites the various categories of reason presented above? It seems to me the best answer to this question can be derived from considering the relation of reasons to value.

It is, I take it, a commonly accepted claim that, if something is of value, then we have reasons to act in certain ways with respect to it, and to have certain attitudes with respect to it. The particular acts that we have reason to perform, and attitudes that we have reason to have, will depend on the nature of the thing in question. They might include, for example, reasons to refrain from defacing a work of art, to respect a person's dignity, or to protect an area of natural beauty that is under threat. It is important to note that subscribing to this fairly innocuous claim does not commit one to the more controversial claim that *all* reasons depend on value in this way – it is at least *prima facie* possible that some reasons depend on value, but that other reasons are generated in other ways.

Now, a further claim which I also take to be true is that if someone truly understands the value of something, they will also understand the reasons that they and others have with respect to that thing. If someone claimed to understand the value of a work of art but then defaced that work of art, the value of a person but then harmed or humiliated that person for no reason, or the value of an area of natural beauty but were happy to see it levelled and built on, then we would have reason to doubt that they can really understand these

things. While it may be possible to question this point about the relation between understanding value and understanding reasons, I am not aware of philosophers who have tried to do so, and it appears to be quite broadly accepted. I will therefore not attempt to defend it here.

Now, it seems to me that what unites the various categories of reason to which I have argued that psychopaths are impervious is that these are all reasons which depend in some way upon the value of entities other than oneself. The pathological selfishness of psychopaths (or at least 'hardcore' psychopaths), I contend, extends to their being unable to see value of this kind, to *ascribe* it to others, and they are therefore unable to understand that they have reasons to act which depend on this value.

It is fairly simple to see how each of the types of reason to which I have described psychopaths as 'impervious' can be understood as depending on the value of entities other than oneself. To begin with the case of people, if someone understands the value of people, they will presumably believe that they have reason at least to respect their rights, their interests, perhaps to some extent their goals and projects (assuming those goals and projects are not immoral). They are also likely to believe that they have reason to support those goals and projects, though facts about the relationship between the valuer and the valued are likely to affect what one has a reason to do in support of these (such reasons are far more demanding, for example, in relationships between a parent and child than they are between work colleagues). All of these are reasons which, if someone did not take themselves to have with respect to a given person or group of people, we would have reason to doubt that they truly understood the value of that person or those people. All of these are reasons which, in fact, psychopaths do not appear to take themselves to have.

The case of animals is also explicable in terms of value and the reasons that depend on value. There are, as noted above, controversies over exactly what

reasons we have with respect to animals, and exactly which animals we have them with respect to, but it seems very plausible that whatever reasons of this kind we do have depend on animals having value.

Reasons stemming from abstract ideas such as justice or the environment can also, I think, be explained as stemming from those ideas having value. In the case of many of these reasons, it may be fairly clear that the value of some more concrete entity or group of entities – either people, animals or some other valued object – lies behind the reasons which we perceive as bearing on our choices and actions. The idea of justice, for example, makes little sense in the absence of some entity – generally people but perhaps also some animals – which has a kind of value which implies that it must be treated justly. In other cases, it may be somewhat more controversial whether such a valuable entity must exist in order to make sense of the reasons in question. In environmental ethics, there is controversy over whether the value of the environment can be reduced to its extrinsic value in serving the needs of humans, whether currently living or yet to be born, or animals or other entities, or whether it also has irreducible intrinsic value of its own.[42] In any case, it seems clear at least that the ordinary way in which we value the environment implies that we value *something*, other than ourselves, whether this is simply the environment itself, or whether it is the people, animals or other entities whose important rights, interests and concerns are served by the environment.

Aesthetic reasons of the kind discussed above, too, presumably depend on aesthetic value. T.M. Scanlon in *What We Owe to Each Other* gives the example of Beethoven's late string quartets – if we understand 'the value of music of this kind' then we will understand that a recording of them should not be 'played in

[42] Routley (1973).

the elevators, hallways, and restrooms of an office building.'[43] Now, it might be thought that aesthetic reasons present a potential counterexample to the claim that the reasons to which psychopaths are impervious are those which depend on the value of entities other than themselves. If psychopaths can be shown to ascribe aesthetic value to objects, then they must be able to ascribe value to entities other than themselves. I have two answers to this. Firstly, as noted above, it remains to be shown that psychopaths actually can ascribe aesthetic value in this way. Secondly, if psychopaths can ascribe value of a kind to aesthetic objects, this may merely be the kind of extrinsic value that those objects have as potential sources of pleasure for the observer. Since the observer is in this case the psychopath, it may be that the ultimate source of value here is the psychopath herself, whose value I am not claiming she is unable to understand.

I am not claiming this because it seems clear to me that psychopaths are able to ascribe value in other ways, where that value is not intrinsic to the entity being directly valued, but is derived from other considerations, ultimately amounting to the value of the psychopath herself. Thus, as noted above, psychopaths may value other people instrumentally, as a means to the satisfaction of their own goals or desires; so too for animals, or other entities such as the environment. Since the ultimate source of value does not reside outside of the psychopath herself, I take it that this type of valuing does not constitute a counterexample to my overall claim here.

I also take it that I need not be committed to the claim that the entities I have been discussing have *intrinsic* value themselves, at least if the alternative is something like utilitarianism, according to which only some ultimate good such as preference-satisfaction has intrinsic value, and any value possessed by

---

[43] Scanlon (1998).

people, animals etc. is derived from the value of this ultimate good. Whatever the ultimate good may be, it must be an entity other than oneself.

The idea that psychopaths are incapable of ascribing value to entities other than themselves (unless that value is derived from the value they ascribe to themselves) is a natural conclusion to draw from reading the various books and studies that have been written about them. It also explains the various types of reason to which they are impervious. But it also explains why they are *impervious* to these reasons rather than *oblivious* to them. If I do not ascribe value to a particular entity, this in no way precludes me from recognising that others may do so, or that in doing so they may take the value of that entity as presenting various compelling reasons for them. Examples to show this can be easily formulated involving any of the types of value explored above, as well as others. In the aesthetic sphere, imagine we have radically different views of a particular artwork: I think it has no aesthetic value at all whereas you think it has tremendous aesthetic value. The fact that I do not see the artwork in the same way that you do has no effect on my ability to understand that you see it in the way you do, and that as a result you take yourself to have various reasons arising from what you perceive as its aesthetic value (reasons to contemplate it, to tell your friends about it, and so on). It is just that I don't see myself as having the same set of reasons. I might further find it difficult to understand *why* you see it as having value in the way you do, but this does not imply any lack of understanding of the fact that you do.

3.6 The implications for responsibility

The next question we must ask is, of course, what this means for moral responsibility. If someone is impervious to a certain set of reasons in the way described, but not oblivious to them, can they be said to be morally responsible for their particular actions and choices that involve those reasons? It is not difficult to think of cases in which someone is oblivious to certain reasons (i.e. they are not aware of the facts constituting those reasons) and, as a result they

are not morally responsible. In Chapter 1, I discussed a case in which, unaware of the fact that B is a recovering alcoholic, A offers B a drink. Let us say A harms B by doing this, because perhaps B is having a difficult time maintaining sobriety and A's offer is the temptation that puts her back on the slippery slope to full addiction. It is surely not the case that A is morally responsible for this harm, however, since she did not know about B's predicament, indeed had no reason to suspect it, and thus was blamelessly oblivious to an important reason that counted against her action of offering B a drink, an action which in the absence of that reason would be perfectly harmless.

However, this type of case does not help us with the present question. In this type of case, the agent is blamelessly unaware of – oblivious to – an important member of the set of facts which provide a reason against performing the act in question. If A knew that B was a recovering alcoholic, she would know that she had a compelling reason not to offer B a drink, and she would therefore be morally responsible for acting against that reason and offering the drink. Now, if A is a psychopath, then she might know that B is a recovering alcoholic, might even know that other people would take this to constitute a compelling reason against offering B a drink, but still she would not, according to my view, be capable of knowing that this fact constituted a reason *for her*, A, not to offer B a drink. If I am to provide support for the position I am defending, that someone in this predicament would not be morally responsible, I need to find a case in which someone is aware of the facts constituting the reason in question, but does not know *that those facts constitute a reason for her*, and as a result is not morally responsible for the act. Specifically, the type of case I am interested in is one in which the protagonist fails to ascribe value to someone or something else, and as a result does not perceive them as providing a reason for action.

I have three cases which appear to me plausibly to fall into his category, and I will discuss each of them in turn. None of these cases is without controversy. However, I believe that for each case, the correct reading is the one I have

outlined above. Together, then, these cases provide support for the idea that to be morally responsible for an action depends not only on knowing the facts that constitute a reason, but also knowing that those facts constitute a reason for them.

The first case is one that has been discussed in a slightly different context by Neil Levy.[44] It is the idea of an anthropologist living amongst an alien civilisation. This anthropologist, in the course of his work, becomes aware of a number of moral practices and beliefs of the civilisation he is studying. In order to further his project, he transgresses moral strictures which the aliens take to be binding. Let us imagine, specifically, that the anthropologist takes a number of plant samples in the course of his visit, despite knowing that the aliens believe that plants have value of the kind that a normal human would take another human to have, in the way that I have been exploring. For the sake of argument, let us suppose that the anthropologist knows enough about the plants to know that they are not essentially different from plants on earth, in a way that would give him reason to refrain from cutting them – they are not sentient, don't have a nervous system etc. Nonetheless, in cutting away parts of the plants, the anthropologist is committing a great moral wrong, in the eyes of the aliens. Now, three questions present themselves. Firstly, does the value supposedly possessed by the plants present reasons which ought to bear on the actions of the anthropologist? Secondly, is the anthropologist responsive to those reasons? Thirdly, is the anthropologist morally responsible for the moral wrong which the aliens take him to be committing?

Note first of all that the putative reasons with which we are concerned here are those which might be presented directly by the plants as valuable entities. These reasons are distinct from reasons arising ultimately from the aliens rather than the plants. It is very likely that the anthropologist would be concerned

---

[44] Levy (2014), p. 358.

about the feelings of the aliens, so that he would want to avoid 'harming' the plants out of respect for the feelings of the aliens. In this situation, the anthropologist would not see the plants as having the kind of value that they have according to the aliens, but would nevertheless treat them as though they did out of respect for another entity which he did take to have the right kind of value, namely the aliens. It is important to discount reasons that are generated in this way because they have no equivalent in the case of psychopaths. Psychopaths, I have argued, see no entities other than themselves as possessing value. Therefore, there would be no situation in which, not seeing one entity as having a particular kind of value, they might nevertheless treat it as though it had that kind of value out of concern for some other entity which they did see as having value in the right kind of way.

Now it seems clear to me, to respond to the second of the three questions above, that the anthropologist cannot be responsive to any reasons directly arising from the value of the plants in this way, as distinct from reasons relating to the aliens and the importance of respecting their beliefs, etc. The anthropologist knows that such reasons exist, at least in the eyes of the aliens, but he cannot possibly take them as applying to him, because he does not share the aliens' views about the value of the plants. To him they are just plants. What is not so clear, however, is how we should answer the first question, in other words whether he actually *has* reasons of this kind – whether in fact such reasons can be taken to apply to him. This really depends on whether the plants in question do have the right kind of value or not. If they do, then the anthropologist is mistaken, and the reasons in question do indeed apply to him as to everyone, but he is not (locally) responsive to them. If they do not, then the aliens are mistaken, the reasons in question do not exist, and the anthropologist cannot be responsive to reasons that do not exist. Since in either case the anthropologist is not responsive to any relevant reasons directly arising from the plants as possessors of value, he cannot be morally responsible for

committing the wrong that the aliens stand to accuse him of, though of course he might be morally responsible for being insensitive towards the feelings of the aliens and for not treating them or their beliefs with adequate respect.

While this first case is based on the fact that the kind of value attached to particular entities can differ between cultures separated spatially, the second and third cases that I have in mind both rely on the possibility of different practices of valuing occurring over time. In the second case, a time-traveller from an ancient culture vastly different from ours arrives in our time with, as one would expect, their set of values and beliefs intact. Now let us say that in this ancient culture people of a certain ethnicity were considered not to have rights or interests that needed to be respected, were routinely owned as slaves and could be used for whatever purposes people of the dominant ethnicity saw fit. Observing our own practices, the ancient traveller comes to understand that we see things differently, but thinks this is no more than a weird quirk on our part and born of a misunderstanding of the proper status of the different races. As a result, he behaves appallingly towards a number of people of the ethnicity in question.

In the final case, a traveller from our own time is transported in a time machine to a point some time in the future, and finds herself in the midst of a civilisation in which vegetarianism has become a universally accepted norm, and in which the idea of eating animals is looked on with universal horror and revulsion. Not a vegetarian herself, she has brought with her a packed lunch which includes some chicken sandwiches. Upon arriving in the future, she soon becomes aware of the different norms surrounding food compared to her own time, and hides the sandwiches, not wanting to incur the wrath of the people of the future. However, later, she finds herself alone and peckish and, not seeing the point of wasting the sandwiches, she eats them, all the time fully aware that she is committing what would be seen by the vast majority of people in the world at the current time as an abomination.

We can ask the same three questions relating to these two cases as we did with the first one, namely: 1) are there any reasons arising from a) people of the relevant ethnicity or b) animals, as possessors of value, which apply to the time-travellers and which ought to count against their a) treating people badly and b) eating the chicken sandwiches, 2) are the time travellers responsive to those reasons and 3) are they morally responsible for the wrong the locals would take them to be committing? The correct way to respond to these questions is also likely to be similar, in that it is going to depend on whether the people or animals do indeed possess the kind of value in question. If so, then reasons against their actions will apply to the time travellers, but they will be blamelessly unresponsive to them, and hence not morally responsible for committing a wrong. If not, then the reasons in question do not apply to them and they clearly cannot be responsive to them. In either case, the time travellers are not morally responsible for committing a wrong.

In each of the above examples, the agent is impervious, though not oblivious, to a set of reasons arising from particular entities as possessors of value, and is not morally responsible for acting or failing to act on those reasons. The general conclusion we can draw from the cases is that responsiveness to reasons depends not only on the ability to understand the reasons in question, but also on the ability to take them as constituting reasons for oneself. The cases, along with the description of psychopaths which I developed earlier in the chapter, also suggest that the ability to take another entity as presenting one with reasons for action depends on having the ability to see that entity as a possessor of value.

According to the hypothesis under discussion, psychopaths are in an equivalent position, but they are incapable of seeing *anything* other than themselves as having value, unless that value is derived from their own value. If this is right, then the range of reasons to which they are responsive, and hence the range of

acts (and attitudes, states of affairs, etc.) for which they are morally responsible will be much more restricted.

Now, there are two objections which might be made to my use of the above examples and the claim that they are analogous to the case of psychopaths, and these objections will point to gaps in the argument which will need to be filled.

The first objection is that it is perhaps somewhat unclear that the protagonists in the cases actually have reasons to be responsive to. That is, they may actually be correct in the assumption that the reasons of which they are aware do not apply to them. This would be true if in each case it was the locals, and not the anthropologist or the time travellers, who were mistaken about the reasons that apply. It would also be true if some form of moral relativism were true, such that the reasons that apply to the travellers, in virtue of their origins, might be different from those that apply to the locals.

It should be noted first of all that, as I have already observed, the outcome of this would not be that the travellers are morally responsible for the actions in question, considered as a harm or a wrong act. Rather, they would be in a situation akin to that of an animal that attacks another animal. While a human performing the same act might have committed a wrong, such considerations simply do not apply to animals. They are not morally responsible for the act, not because there are reasons to which they are unresponsive, but because the relevant reasons relating to the act do not apply to them.

Also, while it might be plausible to suppose that the aliens might be mistaken about the plants' possessing value, or that the people of the future might be mistaken about animals, it is perhaps less plausible, and certainly unpalatable, to suppose this about our attitudes to people of other ethnicities. To make the equivalent supposition about the case of psychopaths would involve believing that people actually do not possess value and that we are wrong to believe that

we have reasons to refrain from harming them, for example. While this is not impossible, it is certainly an extreme view.

Nonetheless, there is a philosophically respectable position, internalism about reasons, which holds that people cannot have a reason unless they are somewhat motivated to comply with that reason, and I have suggested that psychopaths are not moved by considerations arising from other entities as possessors of value. An internalist would therefore be forced to conclude that such reasons do not apply to psychopaths. It is certainly an interesting question whether psychopaths are in this predicament – reasons arising from other entities as possessors of value do not apply to them – or in the predicament of someone who has reasons but is not responsive to them. Not only is this interesting, it has important implications for how we should think about psychopaths. Are they people whose appalling acts are, however counter-intuitive this may be, not contrary to any reasons when they are committed by psychopaths? Or are they people who commit acts contrary to reasons but are not morally responsible for doing so? In Chapter 6, I will address this question and give some considerations in favour of the claim that they are in the latter category.

The second possible objection is that, while it might be plausible in all three cases to claim that the protagonist lacks moral responsibility *at first*, it is far less plausible to claim that they would continue to lack moral responsibility if they remained in the situation and failed to adjust to it. This is particularly true if we suppose that the locals in each case are correct. In this case, we might expect the protagonist in the case to come to see the force of the relevant reasons eventually, at which point they would become responsive to them. If psychopaths are going to be truly and permanently lacking in moral responsibility for their acts, it must be the case not only that they do not see the relevant reasons as applying to them, because they do not see other entities as

valuable, but that they *cannot* do these things. The next two chapters will be dedicated to showing that this is the case.

Before embarking on this project, however, I need to make a final clarification and refinement of the position I have arrived at in this chapter. This is to make clear exactly what I mean when I talk about psychopaths (and others) 'seeing others as valuable', or 'ascribing value' to others. What I am describing here is a belief, but there are several different beliefs which could be described in this way, and I need to make clear exactly what kind of belief I have in mind. The kind of belief which I want to make clear I do *not* have in mind here is a kind of abstract, philosophical belief that entities other than oneself have value. One can imagine a philosophical person (even a philosophical psychopath) becoming convinced through argument that other people have value. This would be a belief held at a theoretical level. It cannot be, however, that a belief of this kind is a necessary condition for moral responsibility, since the vast majority of people are fully morally responsible without having this kind of belief. But can it be a sufficient condition?

To believe that someone or something is valuable, in the sense that I have in mind, is to believe something about certain acts, attitudes or beliefs relating to that person or thing. For example, if I believe someone is valuable, then I must believe that their interests are valuable, and this will have implications for the ways I can act towards them, or the reasons for which I can act in certain ways towards them. Ordinarily, for example, believing someone's interests are valuable would discount acting or deliberating in a way that does not take proper account of those interests. If you are blocking my way on the pavement I cannot simply barge you into the road in order to pass; to do so would be to ignore the fact that your interests would be harmed by my action, and my own interest in getting to my destination more quickly is not enough to outweigh this consideration. Believing that someone's interests are valuable also makes them a candidate to be the object of supererogatory action. If I give money to

a homeless charity, it is presumably because I think the interests of homeless people are of value – it is worthwhile to act in their interests.

Our believing that someone or something has value, then, is a basic condition which must be fulfilled before various evaluative beliefs can be entertained about various acts, attitudes and so on relating to them – for example that it is worthwhile helping them, that harming them without some overriding reason is impermissible, that (in the case of persons at least) they have rights which must be respected, and so on. Now, it is possible to imagine someone coming to believe that someone or something has value in this sense purely through having come to adopt a general, theoretical belief in the value of people and things. Say I am persuaded to the theoretical conviction that all people have value in this sense. Then I must, rationally, believe that you, a person, have value. Would this be enough to make me morally responsible for my actions towards you? Perhaps it would. In such a hypothetical case, I would understand not only that your interests, rights and concerns provide reasons, but also that those reasons apply to me. However, because the belief that someone is valuable is the basis for other beliefs such as that helping these people is worthwhile, that harming them is impermissible, etc., in order to show that I genuinely have this founding belief, I would need also to have the accompanying beliefs. And these accompanying beliefs are intimately connected to how I act, and to how I deliberate about my actions. If I were to find myself seriously considering punching you in the face for my own amusement, for example, this would demonstrate that my theoretical belief in your value did not amount to a genuine belief of the kind that can ground a proper appreciation of the reasons generated by your value.

Applying this to the case of psychopaths, it seems unlikely that a hardcore psychopath who came to believe that others are valuable as a matter of theoretical conviction, would as a result develop the full set of evaluative beliefs and attitudes which I have alluded to here. On the other side of the divide, it

seems highly unlikely that this kind of general theoretical conviction has anything to do with the way non-psychopaths come to see others as valuable. This, in my view, is more likely to be the result of patterns of value-ascription being formed through a developmental process, employing emotions and empathy, in which others are represented as valuable. In the next two chapters I will set out how I think this works.

Conclusions

In the previous chapter, I gathered evidence of the emotional deficiencies characteristic of psychopathy. In this chapter, I have considered various interpretations of these deficiencies in terms of moral responsibility, offering as the best interpretation that psychopaths do not recognise reasons stemming from the rights, interests and concerns of other people, due to their inability to recognise sources of value other than themselves.

So far, my conclusion that psychopaths are indeed incapable of recognising others as sources of value, and thus the reasons based on that value, is based on a plausible reading of the scientific literature describing psychopaths' behaviour and attitudes. The task of the next two chapters is to trace a line to that conclusion from the conclusions of the previous chapter, relating to the peculiar emotional deficiencies which make up the psychopathic personality. I will begin this task in the next chapter by examining the tendency for general emotional deficiencies to interfere with one's ability to make evaluative judgments.

Chapter 4: Emotions and value

Introduction

In the second part of Chapter 3, I presented psychopaths as impervious to certain kinds of reason, namely reasons that depend on seeing entities other than oneself as sources of value. I argued that this leads to psychopaths lacking moral responsibility in cases in which such reasons bear on their choices. However, the claim that psychopaths cannot see anything other than themselves as a source of value was based only on a plausible reading of descriptions of cases in the scientific and clinical literature. It would help to bolster this claim if it could be shown that more firmly established facts about psychopaths' psychological makeup would be likely to result in such a radically unusual outlook at the level of value. The best-established facts, as I have explained, relate to their deficient, 'shallow' emotional experience. In this chapter, I will explore the first of these and examine its implications, arguing that it begins to explain the unusual pattern of value ascription, and hence responsiveness to reasons, that I described psychopaths as exhibiting at the end of Chapter 3.

To get to this point, it will be necessary to engage with the extremely vexed question of what emotions are. If psychopaths have emotional deficits, what exactly is it that they thereby lack, and what implications does this lack have for the ability of psychopaths to ascribe value? In order to answer this question, it will be necessary to digress somewhat from my central argument. I will argue that a general shallowness of emotional experience interferes profoundly with the psychopath's experience of value. To see why this is so requires developing an overall account of what emotional experience *is*, which I will argue is a complex of embodied feelings and evaluative judgments. This takes us into difficult philosophical territory, and disentangling the various conflicting views in order to arrive at a settled position will take the majority of the chapter.

Having done this I will be able, at the end of the chapter, to turn to the implications of the view I favour for psychopaths and the way they ascribe value to the world, connecting this to the conclusions of the previous chapter.

Shallowness of emotional reaction on its own, however, is not enough to explain why psychopaths do not in fact see others as valuable, and is certainly not enough to establish that they *cannot* do so. To get to this point requires building on the general account of emotional experience set out in this chapter with an account of the developmental role played by emotional experience, and specifically empathy, in the ability to ascribe value. I will turn to this task in Chapter 5.

4.1 Theories of the emotions

The debate between competing 'theories of the emotions' is a highly contested area in philosophy. In order to begin to negotiate this territory, it is useful to consider what it is that we know, or apparently know, about emotions - the data which theories of the emotions must try to explain. Peter Goldie[1] has a list of these, which include: diversity in duration, focus, complexity, physical manifestation, degree of development and degree of action-connectedness; the fact that many (if not all) emotions appear to be evolutionarily adaptive; the fact that animals and babies, without language, appear to be capable of at least some emotions, and the fact that emotions stand in rational relation to other psychological states (for example, they can be justified by the same reasons which justify beliefs). I will consider each of these in some depth later in the chapter, so will leave them unexplained for now.

Two further properties of emotions listed by Goldie are particularly important for my purposes in this thesis. Firstly, emotions are about what matters, what is important, what is of value to us. If something makes you sad and you cry, it

[1] Goldie (2007a).

is because whatever has made you sad is something that is important to you. Secondly, emotions seem to provide motivation in some way. So, for example, it would be strange to claim that one is angry about something and yet to have no motivation at all to do anything about it. Whether that motivation will translate into action is of course another question altogether. However, if one professed absolutely no motivation, this would bespeak a kind of indifference that seems to be incompatible with genuine anger.

Another feature of emotions which is generally agreed upon is that they are, or are capable of being, intentional. The word is used here in the phenomenological sense indicating *directedness*, or '*aboutness*'. Emotions are, at least typically, *about* something.

There is a common practice here of drawing a distinction between emotions, which are always about something, and moods, which may be about nothing. If we were to follow this practice, we might also want to try to analyse the connection between moods and emotions. Moods, we might say, involve a disposition to experience emotions in response to certain stimuli. For example, I might go around in a sad mood all day, without that sadness being directed at anything in particular, but one result of my being in that mood will be that when a subject of conversation, say, is presented to me, I will be more likely to feel sad about that particular thing, as a result of my generally sad mood. This is a general claim about the probability of my being sad about any given thing; it does not imply that I will exhibit increased sadness about every single thing that is presented to me. There may well be things about which I will continue to feel happy when I think of them, regardless of how sad a mood I am in. Nonetheless, if there were not a general increase in my propensity to be sad about things, it would seem inaccurate to describe me as being in a sad mood. However, while there is some intuitive appeal in this distinction between emotions and moods, it seems to me to be, to a large degree, stipulative. There seems nothing unnatural, as a matter of ordinary language, in describing the kind of general

mood of sadness described above as being an *emotion*. It might therefore be more helpful to describe emotions as usually, but not always, intentional. At least, they clearly *can* be intentional. Furthermore, their intentional objects can be things external to the subject. For example, if I am afraid of a bear, my fear is *about* the bear. This fact will turn out to be important when we consider what kind of thing an emotion might be.

Theories of the emotions are usually divided into three general types: non-cognitive, 'feeling' theories, cognitive theories and perceptual theories. 'Feeling theories' usually have their roots in the work of William James,[2] who claims that emotions are perceptions of bodily changes brought on by stimuli. Modern feeling theorists, including Jesse Prinz[3] and Jenefer Robinson[4] are indebted to James to varying extents. Cognitive theories, by contrast, hold that thoughts are in some way essential to emotions. Cognitive theories are to be found in the work of Martha Nussbaum[5] and Robert Solomon,[6] both of whom hold that emotions are essentially a species of evaluative judgment. According to perceptual theories, emotions either are, or are closely analogous to (can be modelled on) perceptions, usually of value. Ronald De Sousa[7] is a prominent defender of a perceptual theory of the emotions.

---

[2] James (1884).

[3] Prinz (2004a). Prinz's theory is a sophisticated one which contains elements of perceptualism, but it is a feeling theory in the sense that it identifies emotions with internal perceptions of physiological states and changes.

[4] Robinson (2004).

[5] Nussbaum (2004).

[6] Solomon (2004).

[7] De Sousa (1987), De Sousa (2002).

As noted by Goldie, there is a high degree of variability, in respect of a number of qualities, between the phenomena that are usually called, as a matter of ordinary language, emotions. This variability has led some, including Paul E. Griffiths[8] and Amélie Oksenberg Rorty,[9] to call for the abandonment of the idea that there can be a single unifying 'theory of the emotions'. Griffiths argues that emotions are not a natural kind – that 'the psychological, neuroscientific, and biological theories that best explain any particular subset of human emotions will not adequately explain all human emotions'.[10] The alleged distinction between mood and emotion discussed above is one example where the limits of the vernacular category of the emotions may not coincide with boundaries that can be drawn at a theoretical level. Another such distinction has been suggested based on the work of Ekman and Friesen[11] which identifies six 'basic' human emotions: anger, disgust, fear, joy, sadness and surprise. It may be that some instances of some of these basic emotions do not have enough in common with more complex emotions such as indignation or resentment for it to be plausible that what explains one group at a theoretical level will also (completely) explain the other. In any case, those who do wish to pursue a project of theorising about 'the emotions' must be careful to be clear about which phenomena are to be included in the category and which are not.

The question of what we think emotions are, then, is relevant to the question of how we would expect someone's emotional capacity (or lack of it) to affect the way they experience value. Crudely, if emotions are evaluative judgments, then they are central to our ability to access value through judgment; if they are

---

8 Griffiths (2004).

9 Oksenberg Rorty (2004)

10 Griffiths (2004).

11 Ekman and Friesen (1971), Levenson, et al. (1990).

perceptions of value, then they are central to our ability to access value through perception; if they are feelings, then it is not clear how they could be *central* to our ability to access value, though they would still be likely to affect the extent to which we value some things, and the kind of value we attach to them. This much is true even of other, non-emotional, feelings: for example, my feeling cold and wet when out for a walk is likely to affect my evaluative judgments about the weather.

It is important to understand, however, that even if a cognitive or a perceptual theory is the correct one, in neither case can it be true that emotions are the *only* available means of accessing value.

First, considering cognitive theories, it seems clear that some evaluative judgments, which look equivalent or at least similar to the kind of evaluative judgments that a cognitivist would identify with emotions, can be made in an apparently non-emotional way. Imagine a very experienced judge who is used to making complex evaluative judgments about the accused in criminal cases. She has presided over thousands of such cases, and in each one she has been required to make a number of judgments about the character of the accused and, ultimately, whether and to what degree each person *deserves* to be subjected to criminal sanctions. It certainly does not seem incoherent to suppose that this judge, with the benefit of her great experience, could make such judgments without becoming emotionally involved in each case. In fact, it might even be thought concerning if the judge was bringing her emotions to bear on cases; handing out a more severe sentence, for example, when the case made her angry. Clearly, the operation of emotions is not a necessary condition of evaluative judgment. Furthermore, it appears that – at least in some cases – alternative routes to value might be at least as reliable as emotional routes. It might be that these routes are available even to people with emotional deficits.

Considering perceptual theories, even if emotions are evaluative perceptions, it might be that we have other ways of accessing value than direct perception – for example through judgment. Seen in this way, the case of the psychopath might be somewhat analogous to the case of a driver who suffers red/green colour-blindness. Even though such a person would not be able directly to perceive a traffic light as red, they would nonetheless be able to infer that it is red, from its position on the traffic light, for example. The colour-blind driver would have cognitive access to a fact that others would be able to access directly through perception. Similarly, it might be that a psychopath, though unable to access the value of other people through direct perception, might be able to infer such value cognitively. Of course, it would need to be established that value can be inferred in this way.

In the following sections, I will argue that emotional experience involves both cognitive elements and elements of 'feeling'. There are, I think, strong arguments in favour of each element being part of what we experience when we experience an emotion. Rather than to isolate a single element and call that the emotion, we reach a better explanation by accepting that both are present in and essential to emotional experience. To show why I think this is, I will first examine arguments in favour of, and against, the three broad types of emotion theory, taken in turn. I will then try to show why we should accept a hybrid theory which combines elements of cognitive and feeling theories.

4.2 Feeling theories

Why might someone be attracted to the position that emotions are essentially feelings? To answer this question, we must first understand what is meant by 'feelings'. The Jamesian view equates feelings with internal perceptions of physiological reactions. One simple reason to favour this view, then, is that emotions very often do appear to involve physiological reactions of one kind or another. The hairs on the backs of our necks stand up when we are afraid. We get 'butterflies in our stomachs' when we are nervous. When we are angry, we

become physically agitated and our faces go red. To identify emotions with the perceptions of these physical reactions would have the virtue of explanatory economy: we know they exist, and we know they tend to happen in cases when emotional experiences occur. An explanation which identifies one with the other is at least a simple one.

Another apparent reason to favour this view might come from the thought that it does justice to the phenomenology of emotional experience. This point has been made forcefully by Peter Goldie. As Goldie points out, the presence of such feelings in emotional experience is 'utterly familiar to us',[12] and an explanation that leaves them out, or tries to explain them away as something other than feelings, fails to do justice to this experience. This is an argument from introspection: examining one's own emotional experience, according to Goldie, reveals it to have a phenomenal character closer to that of feelings than to that of beliefs or judgments. This phenomenal character is more readily explained by the idea that emotions actually *are* feelings.

It is worth considering what emotions are being compared to on this view. Other, non-emotional, feelings include the feeling of being cold, or hungry, or in (physical) pain. These feelings are psychological, rather than physiological, yet they are reactive to physiological phenomena in a way that means it is natural to think of them as perceptions of those phenomena, though they would need to be a unique kind of perception, and not the only kind of perception available, at least in many cases. For example, hearing my stomach rumbling is an alternative means of perceiving some of the physiological activity of which hunger would be a feeling-perception.

One thing we know about perceptions is that they can sometimes go awry. The usual hackneyed example is a stick held in water: I perceive it as bent but I know

---

[12] Goldie (2004).

that it is in fact straight. Can feelings go awry in a similar way? It is possible for me to be sitting in a room with the heating turned up to full, so that the vast majority of people in that situation would feel hot, and yet still feel cold, perhaps because I am ill. In these circumstances, it is perhaps natural to say that I only *feel* cold; I am not *really* cold. On the other hand, if I have just eaten a large meal and have no need of further food, and yet still feel hungry, it is less obviously natural to say that I am not *really* hungry. To be hungry just is to feel hungry. The difference between these two cases may simply be one of linguistic convention: when we apply the predicate 'hungry' to a person, we are referring to the feeling they are experiencing, whereas when we apply the predicate 'cold', we are referring to an objective fact about them – their body temperature, perhaps. In any case, it does seem that we can have the feeling without the corresponding physiological process or property being present, which is compatible with these feelings being perceptions of physiological processes or properties, which can go awry in a similar way to other perceptions. Of course, it will remain true in these cases that we are experiencing a feeling of hunger, or of cold, or whatever, but this may simply mean that we are experiencing a feeling *as of* our stomach being empty, or our body being cold, or whatever. In a similar way, the person seeing the stick in water is perceiving it *as if it were* bent.

The idea that emotions too are perceptions of physiological reactions has received some empirical support from scientific studies. One series of studies has been particularly influential in philosophy. In Levenson, Ekman and Friesen's studies,[13] each of six 'basic emotions' – anger, disgust, fear, joy, sadness and surprise – were shown to be accompanied by a unique pattern of physiological response, specifically heart rate, skin conductance, finger temperature and somatic activity. The significance of these results is that they

[13] Ekman and Friesen (1971), Levenson, et al. (1990).

appear to show that the physiological response alone is enough to distinguish between emotions, at least for the six 'basic emotions' that were studied. If this is right, then our experience of each physiological response pattern could plausibly be expected to have a sufficiently different character to be constitutive on its own of the corresponding emotion. There would be no need to bring in other mental phenomena in order to explain how we can tell one emotion from another.

One objection that could be made here is that it is not clear that all of the physiological states listed above are really the kind of thing that we would typically perceive. Even if we can perceive them, we would certainly not need to be aware of perceiving them when we are experiencing one of the six emotions also listed above. When we feel afraid, we need not be aware of an increase in our heart rate or our body temperature. There is a potential difference here between emotions and feelings such as hunger: the feeling of having an empty stomach is at least part of the way we experience hunger, and this can be established just by examining the kind of feeling we have when we are hungry. By contrast, it is not clear that the perception of any of the physiological states listed above is part of the way we typically experience fear, and examining the experience closely does not make this any clearer. On the other hand, the claim under consideration is not that the four physiological states identified in the experiments are specifically central to emotional experience – they are simply states for which relatively simple, accurate measuring techniques exist. If emotional experience really is a matter of perceiving physiological states, then the states involved must presumably be quite complex, and individuating them through introspection might be expected to be very difficult indeed. The claim made on behalf of the data is that distinguishing unique patterns of response across these four factors is enough to establish that the basic emotions can be individuated through physical processes alone. The existence of further processes only strengthens

this conclusion. Perhaps it is plausible that a more complex pattern of physiological response might be processed by the brain in such a way that it is experienced as a single emotion, without the complex underlying structure of the emotion being transparent to the person experiencing it. In an analogous way, the experience of recognising my friend emerging from a shop further up the street is made possible by the processing and interpreting of a huge amount of complex visual data, but my subjective experience is simple: I see my friend, over there.

An objection that has been made to feeling theories is based on the apparent fact that many emotions are extremely long in duration. Robert Solomon uses the example of love:[14] one can be in love for a long time – decades – and it is not plausible to think that one's body is in a continual state of perturbation for the entirety of this time. If not, then the emotion must, for some of the time at least, exist without any physiological correlate. The long duration of some emotions is supposed to be better explained by a cognitive model: we are used to accepting that someone can hold certain beliefs, for example, for a long time, and without those beliefs being present in that person's consciousness at any given moment. It seems extremely strange to suppose that perceptions of physiological reactions can operate in the same way.

Jesse J. Prinz has given an answer to this objection on behalf of feeling theories, which is to draw a distinction between *occurrent* and *dispositional* emotions.[15] No doubt this distinction is a real one: being in love for a long time does not imply that one constantly *feels* in love. Rather, one is disposed to feel in a certain way in certain circumstances – when seeing or thinking about one's beloved, for example. The feeling theorist can say that the long-term lover has

---

[14] Solomon (1976).

[15] Prinz (2004a).

a long-standing disposition to experience the relevant embodied reaction in those circumstances. To ascribe an emotion to someone is therefore sometimes a matter of describing them as having such a disposition, rather than as (currently) experiencing the relevant embodied reaction. Thus, the feeling theorist is able to explain both the dispositional and the occurrent emotion in terms of embodied feelings, without being committed to the implausible position implied in Solomon's objection.

One problem with this answer is that, if emotions are supposed to be feelings, then we might expect a similar distinction to exist between 'occurrent' and 'dispositional' in the case of other feelings, but in fact we do not find this distinction. I can be disposed over a long period of time to feel hungry in certain circumstances, i.e. when I have not eaten – in fact, perhaps barring certain pathological conditions, *everyone* is thus disposed – but this does not imply that I am permanently hungry, in any sense. Why should emotion be different, if emotions are just feelings?

The feeling theorist can reply that a disposition to feel hungry when one has not eaten is not analogous to the lover's disposition to experience loving feelings when thinking about the beloved. Rather, it is analogous to a disposition to experience loving feelings when one falls in love. Anyone might have this disposition without this implying that they are already in love. However, there is a closer analogy available which does seem problematic for the feeling theorist. Imagine I really like donuts. This might imply a disposition to feel hungry when thinking about donuts, seeing a sign advertising donuts, etc. Clearly it would not imply that I am always hungry. Why then should the disposition to experience loving feelings when thinking about (seeing, etc.) the beloved, imply that one is actually in love over the entire time that the disposition exists?

However, this analogy points to a major problem with the use of the example of love to make this point. The problem is that, just as a preference for donuts is not a feeling, it is not clear that love (or at least the kind of love that exists over long periods of time without a corresponding occurrent emotion) should be called an emotion at all. Indeed, the very wide range of unique emotional reactions that can be precipitated by love – including euphoria, anger, jealousy, pride, yearning, self-disgust, resentment and so on, as well as what we might think of as the occurrent emotion of love – suggests that it should rather be thought of as a very complex state of being that manifests in dispositions to experience an array of possible emotional reactions according to circumstance.[16]

Love, then, is perhaps an unusual case and therefore a difficult one with which to illustrate the point. Do we think differently about other phenomena? Take patriotic pride as another example. There is apparently a *feeling* called pride, and yet it does not appear to be a condition of one's being proud that one is actually experiencing that feeling at any given moment. It might be thought that a simple dispositional model applies here: patriotic pride as a dispositional emotion is a disposition to experience pride as an occurrent emotion when thinking about one's country, perhaps. However, as with love, the occurrent

---

[16] Indeed it is far from clear to me that there is an 'occurrent emotion of love'. I was struck recently when reading the precisely expressed autobiographical writing of the Norwegian writer Karl Ove Knausgaard (2013) that he does not use the word 'love' to describe the emotion he feels when seeing and empathising with his daughter, preferring a word which the English translator had rendered as 'tenderness'. Words such as 'tenderness', it seems to me, tell us much more about the phenomenology of emotional experience than the word 'love' could. Love, then, might be a fact about the individual that explains a range of different emotions, and not at all an emotion itself. The song 'F.E.E.L.I.N.G.C.A.L.L.E.D.L.O.V.E.' by the band Pulp also illustrates this very well. In it, the singer describes an array of intense, mostly unpleasant feelings precipitated by being in love – the irony being that none of these is anything like what might naturally be brought to mind by the phrase 'feeling called love'.

emotion which is the offshoot of dispositional patriotic pride may not always be pride itself. It may rather be fear or anger if the country is perceived to be under threat, or even shame if the country is perceived to have disgraced itself.

We might be inclined to ask, however, where this disposition comes from. Why would the patriot have a disposition to feel emotions in a certain pattern in connection to her country? The most obvious answers involve beliefs or evaluative judgments: the patriot believes her country is great, or judges it to be important, or some combination of the two or of other related beliefs and judgments. But these are cognitive phenomena and therefore are supposed to be unavailable to the feeling theorist as constituents of emotions. The feeling theorist would therefore have to accept the existence of these cognitive phenomena as playing a causal role in producing long-term emotions, but deny that the cognitive phenomena are part of the emotion, identifying this instead purely with the disposition which results from them. This seems implausible – surely it is more natural to think that the patriot is disposed to feel (occurrent) pride when thinking of her country *because* she is proud of her country, rather than that she is disposed to feel occurrent pride and this disposition is itself pride. By contrast, a cognitive theorist who believes that emotions are (something like) evaluative beliefs can perhaps more readily explain how a belief about something held over a long period of time can result in a particular set of beliefs held about that thing in different situations. We are used to the idea that believing one thing can dispose us to believe another when faced with a particular set of circumstances.

The feeling theorist, then, is committed to the idea that long-term emotions are noncognitive, but these emotions are intricately, causally related to cognitive phenomena to the point where a more plausible explanation might have the cognitive phenomena as at least partly constitutive of the emotions. It seems to me that this is a problem that arises not only for long-term emotions but for short-term emotions too. Again, there is a potential disanalogy with other kinds

of feeling here. I can feel hungry without having any particular beliefs or making any particular judgments, not only about what I am feeling, but also about any of the circumstances which are producing that feeling, for example that I have not eaten for a while, that my stomach is empty, etc. By contrast, it is not clear that emotions can be experienced in the same cognitively unmediated way. Can they?

It may be useful to return to the supposed distinction between 'simple' and 'complex' emotions here. Leaving aside the 'simple' emotions for the moment, in the case of 'complex' emotions at least, I would suggest it is rather implausible to suppose that they can be experienced without cognitive mediation. For example, how could I experience jealousy without having some kind of belief or judgment that precipitates my jealousy (the belief that the girl I like is flirting with the handsome stranger at the party, for example). How would the jealousy originate if there is no relevant belief or judgment? The same could be said of indignation, or pride, or resentment.

However, even in the case of the simple emotions it is far from clear that cognitively unmediated emotional experience is possible. Firstly, the emotions that are included in the list can have objects which are not accessible to direct perception, either in principle or just as a matter of fact, and hence that require cognitive or imaginative activity to bring them to mind and thereby prompt the emotion. I cannot experience fear of a global environmental catastrophe, or disgust at the cynicism of a nation's foreign policy, or anger at the litterbug who I infer must have dropped that fast food wrapper in the street, without calling those things to mind and, presumably, having some beliefs about them. Secondly, even in the case where I experience a 'basic' emotion towards an object that can be directly perceived, it is still far from clear that cognitive mediation is not required. Imagine I am walking through the woods when a bear emerges from behind a tree. I instantly feel afraid. Is there cognitive mediation involved in this emotional reaction? Not if by cognitive mediation

we mean deliberation. I am unlikely to have time to think about whether I should be frightened of the bear: I simply see it and am frightened. However, cognitive mediation need not involve deliberation; it could consist in the application of a concept to the object I am perceiving, for example the concept of fearfulness or even just of being a bear. As another example, imagine there is a mouldy loaf of bread in my breadbin. Wandering into the kitchen, I smell the bread and am disgusted. I need not know that what I am smelling is mouldy bread in order to be disgusted by it. However, it may not be possible for me to be disgusted without applying a concept such as disgustingness to my experience.[17]

It would appear, then, that at least some emotional experience requires cognitive mediation, and it is quite possible that all emotional experience (or at least all emotional experience on the part of mature humans) requires it. The position the feeling theorist needs to defend, then, is that while cognition may be present in some if not all emotional experience, this cognition is not essential to the emotion itself. The cognition, for the feeling theorist, would not be what comprises the emotion, but what causes it. Thus, it might be that some emotions can be triggered by direct perceptual experience, while others will always require the intervention of thought, but that thought is what causes the emotion, and is not the emotion itself. Perhaps a person cannot experience

---

[17] There may be some room to doubt whether disgust is always really an emotion. The purely visceral reaction that I experience immediately upon smelling the mouldy bread might be something more akin to non-emotional embodied reactions such as physical pain, with truly emotional disgust only entering the picture with some degree of awareness of and reflection on that feeling and its object. For this reason, the fact that infants and animals can apparently experience disgust in some form may not be a counterexample to the position expressed above. One might want to call these reactions only 'proto-emotional' or simply to accept that in these unusual cases emotions like disgust can indeed be unmediated by cognition, while maintaining that nonetheless, in mature human emotion, cognitive mediation is essential.

righteous indignation without believing that they or someone else has been wronged in an important way. Still, this belief is not part of the emotion itself, but merely its cause, by being the cause of the physiological reactions, the experience of which comprises the emotion.

If these physiological reactions are supposed to play this role in emotional experience, one would presumably expect them each to have a unique and distinctive structure. If jealousy and indignation really are different emotions – as they surely are – and there is nothing to them but the experience of different physiological reactions, one would expect the physiological reactions to be distinguishable from each other at the level of physical description. The experiments I have already mentioned suggest that this is possible for the 'basic' emotions. In fact, however, there is also some evidence for the contrary position: that even basic emotions are not distinguishable from each other in this way. Schachter and Singer[18] present evidence to this effect, described here by Jesse Prinz:

> [Schachter and Singer] argue that bodily changes qualify as emotions only when coupled with judgments that attribute those changes to emotionally relevant objects or events. To show this, they injected subjects with adrenaline, which causes autonomic arousal. All subjects were told that they had been given a drug that was designed to improve vision. While waiting for a vision test, some subjects were seated in a room with a stooge who engaged in silly behaviour, such as playing with hula hoops and making paper aeroplanes. Other subjects were given an offensive questionnaire to fill out and seated with a stooge who feigned being irate about the questions contained therein. All subjects

[18] Schachter and Singer (1962).

> were secretly observed as they interacted with the stooges, and all were given a questionnaire about their physical and psychological states after waiting in the room. Schachter and Singer observed that subjects with the silly stooge behaved as if they were happy, and subjects with the irate stooge behaved as if they were angry. There were also control subjects who had been given a placebo and subjects who were forewarned about the effects of the drug. Both showed less response to the stooges. The experimenters concluded that bodily change is indeed necessary for emotion, but cognitive interpretation is needed to determine what emotion a bodily change amounts to.[19]

The supposed implication of Schachter and Singer's experiment is that bodily change alone cannot account for emotion, since the bodily change produced by the drug was identical in each case. Only when coupled with the subjects' judgments about the events around them do these physiological changes count as distinct emotions.

While this is a conclusion I endorse, I do not think Schachter and Singer's experiments offer strong support for it. As Prinz observes, there are alternative interpretations of the experimental data:

> The experiment does not actually establish that the subjects in the two conditions have different emotional states. While their behaviour is different, subjects in both groups report being relatively happy when they filled out the questionnaire about their current emotional state in the final part of the experiment. Schachter and Singer dismiss this, saying the subjects may have been trying not to offend the experimenters, but the same logic

---

[19] Prinz (2004a).

> could be used to explain their behaviour while interacting with the stooges. Perhaps they were just playing along with the stooges to be sociable. On the face of it, this would not explain why the control subjects were less responsive to the stooges, but there is an explanation for this as well. If the adrenaline made the subjects happy, they may have become more sociable, and thus more likely to mimic the stooge. Subjects without the drug were simply less sociable. Subjects who were informed about the effects of the drug may have recognised that their expected states of arousal felt pretty good. They would have concluded that their happiness was caused by the drug, and knowing that it wasn't caused by being in the presence of another person, they may have been reluctant to act in the sociable way that happiness otherwise promotes.[20]

These interpretations are indeed plausible, but there is an even simpler alternative explanation available to the feeling theorist. Even if it is conceded that the subjects in the different groups did indeed experience different emotions, the experiment as described does not sufficiently isolate the physiological from the cognitive elements of the subjects' experience, because it does not guarantee that the physiological reaction of the subjects is identical in each case. This assumption appears to be based on the supposition that the only physiological reaction present is that caused by the adrenaline. However, by introducing the different stooges, the experimenters have introduced an additional potential cause of physiological reactions in the subjects. It may be that, having been made happy or angry by the presence of the stooges, the subjects experience the unique pattern of physiological reactions which the feeling theorist would expect to be present in such cases. To assume that the only effect that the stooges have on the subjects is to provoke judgments on

[20] Ibid.

their part is to beg the question against the feeling theorist, who would expect the stooges to provoke emotional reactions which they (the feeling theorists) would identify as physiological, not cognitive. Nor, again, are the two control groups sufficient to contradict this interpretation of the results. Feeling theorists are committed to nothing that would imply that the adrenaline should not increase the force of the emotion. Two emotions that have unique accompanying patterns of physiological response might nonetheless be such that the addition of a drug such as adrenaline has an intensifying effect on them both. If so, one would naturally expect the subjects who had been given a placebo to feel less anger, or less happiness, than those who had been given adrenaline. Those who had been forewarned about the effects of the drug, on the other hand, might be expected to 'tone down' the behaviour that results from their anger or happiness, even if they felt that anger or happiness to a similar degree. Aware that some of the intensity of their experience was attributable to the drug, they might be wary of acting inappropriately, and modify their behaviour accordingly. Even if the subjects in this last group had an emotional experience, as opposed to merely exhibiting behaviour, that was weaker or less intense as a result of their being forewarned, this does not contradict the feeling theorist's position either. The feeling theorist's claim is not that emotions never *respond* to cognitive processes – this claim is indeed obviously false. The claim is rather that the cognitive processes are not themselves *part* of the experience that is properly described as the emotion. This is compatible with the idea that the subjects who were forewarned about the drug had a less intense emotional experience as a result.

So, at least for the basic emotions, the idea that each emotion is accompanied by a unique pattern of physiological response, so that the inner perception of that response might be expected also to be unique and might therefore account for the experience of the emotion, has some empirical support and is not falsified by the experiments described above. Nonetheless, this kind of account

starts to look less plausible when applied to the more complex emotions. To borrow a useful list, again from Jesse Prinz, the states of mind that are normally described as emotions include such things as "guilt, shame, jealousy, love, indignation, amusement, resentment, nostalgia, schadenfreude, and existential dread".[21] Are we really to believe that each of these complex emotions has a unique pattern of bodily response associated with it? There is (to my knowledge) no experimental data to draw on here, but common sense would seem to suggest that, for example, shame and guilt are very closely associated in terms of what it *feels* like to experience them. Similarly, it seems unlikely that anger and indignation could be distinguished according to their bodily correlates alone. These emotions, it would seem, require cognitive elements to provide the context which is necessary to distinguish one from the other. Indignation is only indignation, and not anger, if accompanied by a judgment that some injustice, or some slight, has taken place. Yet anger and indignation are clearly distinct emotions. If this is right, one might think, then these emotions cannot be 'pure feeling' but must include cognitive elements too.

Prinz's reply to this objection is brief but interesting. He draws on an analogy made by Gordon[22] between emotions and the phenomena of windburn and sunburn. Windburn and sunburn are physically identical reactions of the skin. They cannot be distinguished from each other in respect of their physical manifestations, but only in respect of their cause: one is caused by wind and the other by sun. Yet we have no trouble accepting that windburn and sunburn are distinct conditions, or that they consist in their physical manifestations. The burn on the skin just *is* the windburn, or the sunburn, even though we would need to know how a particular burn was *caused* in order to identify it correctly as one or the other. Similarly, the thought goes, emotions such as indignation

---

[21] Ibid., p. 53.

[22] Gordon (1987).

and anger can only be distinguished from each other by their eliciting conditions. If the emotional reaction is elicited by some perceived injustice, or slight, then it counts as indignation. If not, call it anger. Assuming the physiological correlates of indignation and anger are identical, there must be some cognitive phenomenon – some judgment, perhaps, that an injustice or a slight has occurred – if we are to tell indignation from anger. But this need not imply that this cognitive phenomenon is *part of* the indignation, any more than the wind or sun is part of the windburn or sunburn.

This explanation makes some intuitive sense, partly because it fits our definitions of the emotions concerned. *Anger at an injustice or slight* is a pretty good working definition of what indignation is. This point also applies to several of the other complex emotions mentioned above. Schadenfreude is joy at another's misfortune – that is, in fact, precisely what schadenfreude *means*. So, the feeling theorist might say, one would expect it to be distinguishable from other forms of joy through its eliciting condition: it is schadenfreude because it is joy specifically at someone's misfortune, and not joy at some other object. Still, it might still be that the *feeling* of joy – the perception of an embodied reaction – is what we should identify as the schadenfreude, and not our judgment that someone has suffered a misfortune, or a judgment that this misfortune is enjoyable, or any other cognitive phenomenon.

However, this point reveals an ambiguity in the notion of an 'eliciting condition' that I have been using above, which points to another important feature of emotions. Windburn is windburn because it is *caused* by the wind, but schadenfreude is joy *at* another's misfortune, and the language of causation does not capture this quality of *directedness* that emotions have. The fact of the other person's misfortune – or my awareness of that fact – is not just part of the causal explanation for my experiencing schadenfreude; it is also what the emotion is *about*. In the jargon, it is the *intentional object* of the emotion. This fact about emotions is more readily explained if they are cognitive phenomena

such as beliefs or judgments, or if they are perceptions, than if they are embodied feelings. It is generally accepted that beliefs, including evaluative beliefs, are *about* things: If I believe the bear is frightening, my belief is about the bear – the bear is its intentional object. This is also true of perceptions: when I see the bear, the bear is the intentional object of my visual perception. However, if a feeling – an internal perception of a physiological change – has an intentional object, then that object is surely the physiological change itself, and not some additional thing outside my body.

The intentionality of emotions is one of the most difficult aspects of them for feeling theorists to explain, and their attempts to do so are sometimes elaborate and complex. For example, Jesse Prinz's theory holds that emotions are about external objects in the sense that the embodied feelings that constitute emotions represent external objects as part of an evolved mechanism. I do not intend to evaluate this or any other specific feeling theory in detail here, or indeed any other specific theory of the emotions. I think it is fair to say, however, that the apparent intentionality of emotions sits very uneasily with the claim that they are pure embodied feelings. This can be further brought out by noting a related problem for feeling theorists, which is the difficulty of explaining what separates emotions from other feelings. As Prinz acknowledges:

> If the essence of being an emotion is being a perception of a (relatively global) bodily change, then fatigue and starvation should qualify. This suggests that emotions must have some other essence. The [feeling] theory leaves the most fundamental question unanswered: What is it to be an emotion?[23]

[23] Prinz (2004a), p. 52.

This problem exacerbates the intentionality problem, because intentionality seems as though it must be part of what separates emotions from 'other' feelings. 'Fatigue and starvation' are not about anything but the bodily states they represent. By contrast, fear is about whatever I am afraid of. The feeling in my stomach after I have eaten some bad food is about nothing other than whatever is going on in my stomach. My disgust at the corrupt politician, however, is about the politician. This intentionality of the emotions seems as though it must be part of what demarcates them from embodied feelings, because it is so hard to explain how embodied feelings could be intentional in this way. Now, it may be that the feeling theorist can use intentionality as part of an explanation of what separates emotions from other members of the set of embodied feelings of which they are one sub-type. Indeed, this is Peter Goldie's approach in the development of his idea of emotions as 'feelings-toward'. However, such an explanation would run counter to the common-sense intuition that says that emotions have distinct intentional objects whereas embodied feelings do not.

Another apparent feature of emotions which is difficult for feeling theories to explain is the fact that emotions appear to be open to justification in the same way that judgments and beliefs are open to justification. Imagine I am annoyed with you because you said you would meet me for dinner and then you failed to turn up without providing an explanation. If your failing to turn up was simply a result of your being careless and forgetting the appointment (and let us say this is simply the latest incident in a long line of similar lapses on your part) then my annoyance might be justified. It would be justified, then, in the same way that a number of evaluative and non-evaluative beliefs on my part would be justified: for example, my belief that you are careless, or just my belief that you have forgotten the appointment. If it turns out that your failure to turn up is due to a medical emergency, then this fact would render my anger unjustified, in the same way that it would render the aforementioned beliefs unjustified.

The status of the emotion as justified or unjustified is responsive to a number of facts about you and your behaviour in the same way that beliefs and judgments are responsive to such facts. This is a further bolster to the case for emotions themselves being, or being akin to, beliefs and judgments. If not, and they are simply 'brute' feelings, it is hard to see how they can be the kind of thing that can be subject to justification at all. It would be as though someone who had sunburned skin despite not having been in the sun was described as having an 'unjustified' reaction. Such a reaction would be unusual, to be sure, but the language of justification is simply inappropriate in such cases. How might a feeling theorist respond to this argument?

I think the most plausible response would be similar to the one I sketched above in relation to the question of whether emotions can be experienced without cognitive mediation. That is, the feeling theorist would need to hold that people experiencing emotional reactions do have beliefs and judgments that are part of the cause of those reactions, but that the beliefs and judgments are not the reactions themselves. Othello becomes convinced that Desdemona has been unfaithful to him and experiences jealousy as a result. Someone who was aware of the facts of the situation would be entitled to infer a set of beliefs on Othello's part – that Desdemona is unfaithful, that she is treating him with contempt, and so on – beliefs which would not be justified by the facts. When we say that Othello's jealousy is unjustified, therefore, we might really mean that the beliefs which cause him to be jealous are unjustified. This would make the idea that Othello's emotion of jealousy is unjustified by the facts – or that in the alternative case where Desdemona really has been unfaithful, it is justified – an intelligible description of the case in ordinary language, though it may not be precisely correct at the level of philosophical analysis.

As in the previous discussion, however, it seems as though the feeling theorist has conceded quite a lot with this response. Cognition, it is suggested, is closely bound up with emotional experience, to the point where it may not be possible

to experience an emotion without accompanying beliefs or judgments. If there is indeed pure feeling involved, it must be so intricately connected to those beliefs or judgments, so embedded in the experience of having an emotion, that they are naturally described as a single entity, so that we think of our emotions as being subject to justification, and not just the beliefs and judgments on which they are based. (It is worth noting again, by the way, that this would apply as much to 'simple' as to 'complex' emotions, since we think of these too as being subject to justification.) Still, the feeling theorist must claim, only *that bit there*, the embodied feeling, is the emotion itself. The rest is incidental. But why should we believe this? Why not instead believe that emotions involve both cognitive and feeling elements? In fact, I think this is the most plausible description of emotions.

4.3 Cognitive theories

I have given some reasons to reject the idea that emotions are pure feeling, but there are at least two further possibilities on the table: firstly, that they are pure cognition, and secondly, that they are perceptions, or something akin to perceptions. Let us consider the cognitivist view next.

The considerations that speak in favour of cognitivism are essentially those I have adduced above in arguing against pure feeling theories. The idea that emotions are (or are akin to) beliefs or judgments can make better sense of a range of apparent attributes of emotions than can the idea that they are pure feeling. These attributes are their intentionality, their justifiability, and the fact that cognition appears to be present in most if not all emotional experience, at least as a causal factor, and is probably needed to distinguish between different emotions, again in most if not all cases.

Nonetheless, there are a number of reasons to doubt that emotions can be accounted for purely in cognitivist terms. One of these is that, quite simply, the experience of having an emotion is not like that of having a thought. A major

reason why feeling theories have any traction in the first place is presumably just that the view of emotions as being at least partly embodied accords with the subjective experience of having an emotion. Strong emotions can manifest in powerful bodily reactions, including shaking, paralysis, increased body temperature, restlessness and so on. Here we see the mirror image of an argument against feeling theories I presented earlier: clearly when we experience emotions we do undergo physiological changes, and our experience at the time is partly constituted by our internal perceptions of these changes. Why then deny that this is part of the emotion we are experiencing? The pure cognitivist would need some reason why we should think of only the beliefs or judgments as being the emotion, and not the embodied reactions we are experiencing at the same time.

In arguing against pure feeling theories, I suggested that these theories have difficulty explaining why some feelings are emotions and others are not. Again, there is an equivalent worry for cognitivists. If emotions are evaluative beliefs or judgments, then why are some evaluative beliefs or judgments apparently not emotional? More specifically worrying for cognitivists is the fact that many of the particular evaluative beliefs or judgments that might be thought to constitute emotions in one case, can in another case apparently be experienced in an entirely non-emotional way. I used the example earlier of a judge who makes careful evaluative judgments about her cases without becoming emotionally engaged. In an alternative case, she *might* become emotionally engaged, empathising with the defendant or the victim, feeling angry or disgusted about the crime itself, and so on. In the two cases, her evaluative beliefs and judgments might apparently be exactly the same. How then can the cognitivist explain the difference between the two cases? It seems to me that the difference is more readily explained in terms of an absence of feeling in the latter case than by an absence of beliefs or judgments.

Another aspect of emotional experience which cognitivists find difficult to explain is the fact that emotions can sometimes persist after the judgments or beliefs which caused them, and which the cognitivist would identify with them, have changed. Imagine I am alone in the house late at night, when I suddenly see what appears to be a human face looking in through the kitchen window, which naturally terrifies me. Now imagine upon closer inspection it becomes clear that the 'face' is actually some trick of the light. Nonetheless, it had looked so real at first that I still feel afraid after realising this. Perhaps I have to switch the lights on and sit down to compose myself. Why am I still afraid? If my emotion of fear is purely a set of evaluative judgments about the terrifying face at the window, then one would expect that emotion to disappear as soon as it becomes apparent that no such face exists. If, on the other hand, my fear is partly my perception of the physiological reactions which result from my experience, then it is natural to think that at least that part of the emotion will persist, to a gradually decreasing extent, while my body reverts to its normal equilibrium.

One answer the cognitivist might have to this challenge would be to say that, although I may not be aware of it, some evaluative judgments do persist in cases like this. Indeed, there may be some cases in which we tend to think of the lingering emotion as betraying a certain judgment on the part of the person experiencing it. Imagine David is annoyed with his wife for losing his keys, and then he realises that she did not lose them, but his annoyance persists. Beyond a certain point, an observer might start to suspect that it was not really the keys that David was annoyed about. There is some other judgment he is secretly, or perhaps not fully consciously, making about his wife, some other belief he secretly holds about her, that is the real, deeper cause of his annoyance. In fact, the cognitivist might say, it may be more difficult for feeling theories to account for cases like this than it is for cognitivist theories. After a certain point, it becomes unlikely that David's residual annoyance is attributable to his body

gradually regaining equilibrium, and the observer becomes entitled to infer a second belief, or set of beliefs, that is sustaining his emotional state. However, a cognitivist would need to explain *all* cases of emotional persistence as betraying the presence of some hidden belief or set of beliefs. This seems less plausible. Sometimes, as with the 'face at the window' case above, emotions persist in a manner that apparently puts them in conflict with *all* of the relevant judgments and beliefs that we hold.

4.4 Perceptual theories

The difficulty both cognitivist and feeling theories have in explaining how emotions can persist apparently in the absence of relevant evaluative judgments is frequently cited as a consideration in favour of theories which either identify emotions with perceptions, or 'model' them on perceptions in some way, i.e. suggest that they share some key attributes with perceptions, rather than with thoughts or feelings. The perceptualist can observe that we have no trouble accepting that perceptions can exist in conflict with beliefs or judgments. For example, I can perceive a stick held in water as if it is bent. It *looks* bent. Nonetheless, I do not believe it is bent, because I am aware that my perception is unreliable. In the same way, the example of David and his wife above does not seem problematic if his annoyance is neither an evaluative judgment nor an embodied feeling, but a perception of his wife as having certain evaluative attributes (thus, a perception of value). David does not hold any particular evaluative beliefs about his wife that constitute his emotion, so the explanation goes, but he continues to see his wife as annoying. In the 'face at the window' case I have ceased to believe that there is a face at the window, but nonetheless perhaps whatever it is that looks like a face is so uncannily face-like that it is still frightening to me. In each of these cases, the perception involved is in conflict with the beliefs and judgments entertained, but there is nonetheless no logical contradiction involved in experiencing the perception while entertaining the belief or judgment. This phenomenon of conflict without contradiction is

used by Döring[24] as the basis for an argument in favour of perceptualism and against cognitivism. If emotions were judgments or beliefs, we would expect them to be susceptible to change in the light of alterations in the other beliefs that we hold, but in fact they are recalcitrant to such change. Because perceptions are also recalcitrant to change in the same way, as illustrated by the stick example, we ought to think of emotions as being, or being akin to, perceptions.

However, as Salmela points out in a reply to Döring,[25] this argument mischaracterises the particular kind of conflict that can exist between emotions and beliefs. This relates to the fact about emotions that I explored in a section above – that they are subject to justification, and specifically to rational justification. To put this another way, in the emotion cases, there is a norm of rationality governing our emotional reactions which is violated by the persisting emotion, whereas in the straight perception case, there is no such norm of rationality governing the perceptions. It is not the case that I ought to see the stick as straight. I am not irrational if I continue to perceive it as bent despite believing it to be straight. However, my continuing fear after realising that the 'face' at my window is not really a face is irrational, and I am irrational in continuing to experience it. Similarly, David's continued annoyance at his wife is irrational, at least insofar as it is supposed to be justified by his wife's losing his keys. Unless there is some other good reason for David to be annoyed with his wife, he ought to stop being annoyed with her. If he is concerned about his own rationality, and wants to be reasonable, David will try to stop being annoyed with his wife, in a way that does not work in the stick case – despite

[24] Döring (2004).

[25] Salmela (2011).

knowing that the stick is straight, I am under no normative pressure to try to see it as straight. As Salmela observes:

> The fact that we regard many recalcitrant emotions as well as pathological emotions as irrational rather than *arational*, and try to get rid of them, implies that the problem with recalcitrant emotions is not so much whether they *need* to be revised in the light of better knowledge, but rather whether they *can* be so revised.[26]

This represents a fundamental disanalogy between emotions and perceptions, as it also represents a fundamental disanalogy between emotions and embodied feelings (there is no norm of rationality or 'rational ought' governing these either – if I feel cold despite not being cold I am not thereby irrational and it is not the case that I, rationally, ought not to feel cold).[27] Despite the claims made by Döring and other perceptualists, the rational relations which apparently hold between emotions and beliefs are in fact better explained by the cognitivist view. It is not impossible for someone to feel afraid of something while also knowing it not to be worthy of fear – to experience an emotion and to hold a belief that are in conflict with each other – but they are irrational in doing so. In the same way, it is not impossible for someone to hold two contradictory beliefs, but they are irrational in doing so.

Another issue with perceptualist theories is that, in order to provide a clear alternative to cognitivist accounts of emotional experience, perceptions would

---

[26] Ibid., p. 15.

[27] There are, perhaps, applicable norms here. We might say that someone 'should not be feeling cold' with the implication that there is something either physiologically or psychologically awry which is interfering with the feeling and its relation to external conditions. Whatever kind of norm this is, however, it is not a norm of rationality.

need not to be cognitive themselves. As Salmela notes, however, cognitive activity is bound up intimately with perceptions, or at least with the perceptions of non-infants. Most, if not all, of our perceptual experiences are experiences of 'perceiving as'. In the case where I am afraid of a bear, it seems unlikely that I can see the bear without seeing it *as* something, whether as a bear, or a big hairy brown thing, or a threat, or something else. While I may not be aware of this, I am nonetheless applying a concept to my perceptual experience, which involves thought. As Salmela puts it, 're*cognition*' is 'a kind of cognition'.[28] Certainly, this appears to be the case in those instances where my perception is emotionally 'coloured': in order to be afraid of the bear I must see it as a threat, or as something fearsome, or whatever. But then this means that, again, cognition is in the frame as a possible explanation for the emotional aspect of the experience. As in the discussion of feeling theories, it appears to be impossible to isolate perceptions in an account of a specific emotional experience, so that we can say, 'here is the emotion, and here is the perception, and there is no cognition, therefore the emotion must be a species of perception and not of cognition.'

4.5 Reconciling theories of the emotions

In this chapter, I have set out considerations in favour of, and against, the three main families of theories of the emotions. With only a limited space in which to do so, I have had to confine myself to reasoning at quite a high level, not seeking to engage with the detail of specific theories but rather looking at the general considerations that are either friendly or unfriendly to each family of theories. Of course, it is entirely possible that an answer can be found to each of the criticisms I have raised against each family of theories, but these answers will in each case need to convince us of a somewhat counterintuitive conclusion.

---

[28] Salmela (2011), p. 10.

For clarity's sake, it will probably be a good idea to summarise here the considerations I have looked at so far. I have argued that feeling theories are attractive partly because they do justice to the phenomenology of emotion. There is something that it *feels* like to experience an emotion, and embodied feelings are a plausible model for this. It is also apparently the case that, at least in many cases, emotions are accompanied by physiological changes, and we do experience an internal perception of those changes. Why not then think that this internal perception is itself a part of the emotional experience? I also noted some empirical support for feeling theories which comes from experiments which have apparently shown that at least the 'basic emotions' can be distinguished from each other based on their physiological correlates alone. On the other hand, I argued that the prospects for distinguishing emotions on the basis of their physiological correlates alone were much shakier when considering the 'complex emotions' (i.e. those other than anger, disgust, fear, joy, sadness and surprise) for which it is likely (and this is acknowledged by feeling theorists such as Prinz) that their eliciting conditions, which will probably include cognitions, will be necessary to distinguish between them. Against feeling theories, I noted the difficulty they have in explaining why some feelings are considered emotions while others are not. I also noted their difficulty in explaining how emotions can apparently persist over long periods of time – and that the most plausible available solutions to this problem appear to push the feeling theorist into cognitivist territory.

In favour of cognitivist theories, I argued that judgments or beliefs are present in, and apparently integral to, the vast majority of emotional experiences, to the point that a good reason would be needed to exclude these cognitions from what constitutes emotions (and again, this is particularly well brought out by considering the phenomenon of emotions persisting over a long period of time). I also noted the intentionality of emotions, a feature of emotional experience which feeling theories have difficulty explaining but which is much more readily

explained by cognitivist theories. I also pointed out that emotions are apparently subject to justification in a way that is characteristic of beliefs and judgments but not of embodied feelings. Against cognitivism I noted that cognitivist theories, too, have difficulty explaining what is distinctive about emotions, given that it is apparently possible to entertain the evaluative judgments and beliefs which the cognitivist would hold to be constitutive of emotions, in an entirely non-emotional way.

Finally, I considered the perceptualist alternative, which is attractive because it purports to account for the phenomenological character of emotions while also accounting for the rational relations which hold between emotions and beliefs (as well as other emotions). However, I argued that perceptual theories in fact mischaracterise these relations in a way that makes them poorly suited as a model for emotions. I also argued that, as with feeling theories, cognitive elements appear to be present in cases of emotional experience which are candidates for explanation along perceptualist lines, so that, again, we need a special reason to exclude those elements from the set of what constitutes the emotion.

So where does this leave us? If each of the prevailing families of theories have problems which they have great difficulty addressing, what should we conclude about the nature of emotions? I believe this very difficulty of fitting emotions into an existing category points us towards the most plausible answer, which is that emotions are complex entities with elements of both feeling and cognition. When we experience an emotion we make judgments, or entertain beliefs, with evaluative content, but we also experience the embodied feelings that those judgments or beliefs cause in us. My suggestion is that it is our combined experience of these thoughts and feelings that we call the emotion. When I see the girl I like apparently flirting with the handsome stranger at the party, I believe that she is flirting with him. Perhaps I judge this situation to be threatening to my plans, projects or desires: perhaps I was planning to flirt with

her myself, or would like to. Perhaps I thought she was interested in me, and this flirtation – when she knows I am right here, after all – is an indication that this belief is mistaken. This storm of cognitive activity also sets off a visceral response in me, and I experience this from the inside in combination with these various cognitions. My combined physiological upheaval together with cognitive upheaval, I experience as jealousy. Crucially, the elements of thought and feeling involved in this experience are intimately connected with each other and act on each other in subtle and complex ways. Not only does the complex of beliefs and judgments I am experiencing precipitate my embodied feelings, but the embodied element of the experience sustains and intensifies its cognitive elements. As I watch her flirting with *him*, it is the churning in my guts, and the hot feeling in my skin, that signifies to me how *bad* this situation is, and perhaps also makes me aware in a way that I had not previously been, of what *she means to me*.

It seems to me that this combined view makes better sense of what it is like to experience an emotion than either feeling or cognitivist theories alone. The fact that my experience includes elements of belief and judgment explains its intentionality: it is directed at the scene I am watching because many of the beliefs and judgments involved are beliefs and judgments about that scene, or about the people involved in it. On the other hand, there is something that it *feels* like to witness this scene, and I am viscerally engaged in it in a way that I would not be if I was merely making a set of judgments, or entertaining a set of beliefs. It would be conceivable for me to make all of the same judgments in a completely non-emotional way; the fact that there are embodied feelings involved in my experience explains why it is an emotional experience at all.

There is an immediately apparent difficulty with this kind of combined account, which might be put as follows. While it may be true that the experience of having an emotion has elements which are best explained as cognitive, and others which are best explained as embodied feelings, it is apparently not the

case that we experience these things as separate entities. When I experience jealousy, I just experience *jealousy* – I experience it as a single thing, perhaps one with elements of cognition and feeling, but not as a bunch of separate feelings and cognitions. Why would I mentally combine all of these disparate elements into a single experience and call it 'jealousy'?

There may be a way to explain this while also explaining two apparent facts about emotion which I noted at the beginning of the chapter, and which I stated would turn out to be important for my overall thesis: the fact that emotions are about what is important to us, and the fact that emotions are motivating. The explanation is that emotions serve an evolutionary purpose which would not be served if they were experienced simply as their disparate elements. Fear needs to be motivating, so that we will consistently take steps to avoid the things we are afraid of, which in typical cases will be dangerous. Disgust needs to motivate us to avoid its object, because the things we are naturally disgusted at – at a very basic level – are things that are likely to make us ill. Anger, perhaps, in its basic form, motivates us to take action on behalf of ourselves or our family or community when they are threatened. Each of these emotions has a physical manifestation which also serves a parallel evolutionary purpose. When afraid, we enter a state of high alertness, our muscles tense, we prepare physically for 'fight or flight'. When disgusted, our stomach churns, motivating us to avoid eating or drinking, and to remove ourselves from the vicinity of the object of disgust because of the unpleasant sensation it evinces in us. When angry, blood flows to our muscles, our fists clench, we instinctively make threatening gestures, all of which can be seen as preparations for aggressive and fighting behaviour. What I am suggesting is that we have evolved to experience these physical changes in combination with related beliefs and judgments in a way which motivates and readies us, at a psychological level, to take the form of action appropriate to the stimulus which has prompted them. This experience is what we call an emotion. That we experience these sets of phenomena as

coherent, apparently unified mental phenomena called emotions can perhaps be explained by the fact that the experience needs to motivate us to respond quickly to cues from our environment.

Another problem that the combined view I am proposing faces is similar to one I noted in connection to feeling theories before: if an emotion is a combined experience of physiological reactions with relevant beliefs and judgments, why do only some such combined experiences count as emotions? For example, I can feel that my stomach is empty, and I have a set of evaluative beliefs and judgments directed at the cake in the shop window, concerning how delicious it looks and so on. The feeling of hunger will intensify the judgments I make about the cake in a way which looks quite similar to what I have described in the case of emotions. Why then is the hunger I experience merely a feeling, and not an emotion?

One difference between this and seemingly analogous cases of emotional experience is related to intentionality. It might be said that I am 'hungry for' the cake, and in this sense the hunger has the cake as its intentional object. All the same, the intentionality of this experience is not essential to it in the same way that the intentionality of emotional experiences is essential to them. I would still be hungry if no cake existed, and I can be hungry without thinking about or perceiving any food. While my hunger might motivate me to take action – to find food for example – it does not need to have any object other than my own internal physiological state. It has served its purpose by alerting me to the fact that my stomach is empty and I have not eaten recently. It is not clear that it is possible to be angry, or fearful, in the same way. Even if I am stomping around being angry at nothing in particular, my anger will still find objects, whether trivial or important, and which may include myself.

A related difference is that it is apparently not essential to feelings other than those involved in emotions that they depend on beliefs or judgments, whereas

the preceding discussion suggests that this is the case for emotions. When we experience jealousy or indignation, our jealousy or indignation is precipitated by events and states of affairs in the world, but the link between those events and states of affairs and the emotional reaction is indirect – there is a cognitive link in the chain between these two things. I have to believe that the girl I like is flirting with the handsome stranger before I can feel jealous about this fact. Non-emotional feelings such as hunger and fatigue, on the other hand, are not, or at least not typically, like this. We do not have to believe that we have expended a lot of energy, or spent a long time without sleep, in order to be tired. We just *are* tired, as a direct result of our having expended a lot of energy, or spent a long time without sleep. The relationship between emotional feelings and emotional thoughts is thus more intimate, more interactive, than the relationship between non-emotional feelings and thoughts. This difference might be enough to explain our practice of putting emotions and non-emotional feelings and cognitions into two distinct categories.

Against this, it might be pointed out that it is not the case that non-emotional feelings are never cognitively mediated, that they never depend on beliefs and judgments. It is a well-recognised phenomenon that feelings such as pain, or the feeling of being hot or cold, or fatigue, can be greatly influenced by the judgments and beliefs of the person doing the feeling. Imagine you have been running around trying to get things done all day, having had a bad night's sleep last night. You don't feel tired because you haven't had time to think about being tired. Sometime in the evening, you are telling a friend about the kind of day you have had, when she interjects, 'you must be really tired'. Suddenly, you realise that you *are* tired, and immediately feel an overwhelming sense of fatigue. In this case, it might be that the belief that you are tired has precipitated the feeling of tiredness.[29] This seems like a case of a non-emotional

[29] It could be asserted that it is not the belief that you are tired that precipitates your tiredness in this case. Rather it is the fact that you have allowed

feeling being cognitively mediated in a way that I have suggested is characteristic of emotional feelings. So now we have arrived at a position where both emotional and non-emotional feelings can sometimes be cognitively mediated. How then can this characteristic be the basis of a distinction between the two?

One slightly weak answer to this objection is that, despite the cases I have just described, nevertheless in most typical cases the difference in character which I have ascribed to emotions and non-emotional feelings, holds true. Moreover, this is how we typically think about the two different kinds of feeling. There is something weird, and surprising, about cases like the tiredness case. We expect to feel tired just because we are tired. The fact that we sometimes have to first believe that we are tired, though it is undoubtedly true, seems to undermine our common-sense beliefs about how these things work. In contrast, it is essential to emotional experiences that, while they involve embodied feelings, they also involve evaluative beliefs.

Another difference between emotional cognition/feeling complexes and non-emotional ones is that the beliefs involved in the emotional cases are, essentially, evaluative beliefs about things that are important to us. It is quite possible, if perhaps unusual, for someone to feel hungry without this fact being particularly important to them. In contrast, I cannot imagine being angry but not caring about whatever I am angry about.

The clearest difference, though – and hence the strongest answer to the objection above – points to a truly distinguishing feature of emotions on the account I have been developing. In the case in which my feeling of tiredness is

---

yourself to relax while talking to your friend. But it need not be the case that you have allowed yourself to relax – in fact you might still have things to do and might be anxious to get away from your friend so that you can get them done. Still, her observation, and your recognition of its truth, makes you feel tired.

precipitated by my coming to believe that I am tired, the relation between the belief and feeling in question is causal and contingent. The tiredness is the embodied feeling that I am experiencing. In this case, the tiredness has been caused (partly) by my coming to believe that I am tired; in another case, some other cause might bring about exactly the same feeling and there is no question that this feeling would still constitute tiredness. In the jealousy case, however, not only is the embodied feeling I am experiencing partly caused by a certain kind of belief or judgment having to do with the scene I am witnessing and my relation to it, but the experience I am having would not even count as jealousy unless something like that set of beliefs or judgments were present, bound up with the embodied feeling. The cognitive element of the experience is not just causal but also constitutive of the emotion in a way that the cognitive element of the experience in the tiredness case is not constitutive of the feeling of tiredness in that case. Moreover, it is a constitutive *condition* of the emotion, without which (or without something of the same general kind) it would not count as that specific emotion.

The complexes of thought and feeling that comprise emotions, then, are distinguished by the complex interactivity which exists between their two components. Emotions essentially involve – are partly constituted by – evaluative beliefs and judgments about things we care about, and the feelings involved in emotions depend on these beliefs and judgments. In turn, the feelings involved in emotional experience give that experience its characteristic phenomenal intensity. Although these feelings are separate entities from the beliefs and judgments involved, because we experience them as coherent wholes, the phenomenal character of the feelings affects the character of our evaluative beliefs and judgments. When observing the girl I like at the party, I do not simply draw a set of conclusions about her behaviour and its effect on me, I *feel* the importance of these implications of what I am observing. The intensity of my perceptions is affected too, because I am not simply perceiving

a scene taking place, I am seeing this scene *as* something disastrous for me personally. My powerfully held evaluative beliefs are giving my perceptions emotional colour and force.

I have argued that it is possible to make evaluative judgments, and to hold evaluative beliefs, either emotionally or non-emotionally, with the difference between the two cases explained by the presence or absence of an embodied feeling within the experience. I have also described such feelings as both motivating and concerned with things that we care about. Does this then mean that it is impossible to hold a belief or make a judgment that is a) motivating and b) about something I care about, without this belief or judgment being part of an emotional complex, i.e. without my experiencing it as part of a whole which includes elements of embodied feelings?

When it comes to motivation, presumably this *is* possible; it is not the case that we always need to be emotionally engaged in order to be motivated to act. Furthermore, we can sometimes be motivated to act in a way that is opposed to the motivation provided by our emotional state, as in situations where 'the head rules the heart'. In the case of the emotionally unengaged judge, the judge is certainly motivated to make judgments, and to act on them, but she is apparently not motivated by emotions. Philosophical accounts of motivation reflect this possibility. According to Humean moral psychology, for example, motivation requires the presence of beliefs and desires. Neither of these things is the same as an emotion, or implies the presence of an emotion. The link between emotions and motivation appears to be a contingent and a defeasible one. Desires can arise from various sources, which would include emotional experience: we might foster a desire to right a wrong, for example, as a result of being angry about that wrong. We might foster a desire to avoid something as a result of being afraid of it. But it is quite possible to imagine both of these desires arising in the absence of the relevant emotional reaction. The tendency for desires to arise in this way is perhaps enough to explain our tendency to

think of emotions and motivation as closely linked, but we need not invoke a necessary connection to make sense of this, nor, I think, is it plausible to do so.

Then we have the apparent connection between emotions and value. At the beginning of the chapter I suggested that emotions are about what is important to us. I then went on to argue that emotions are a combination of evaluative beliefs and judgments, together with embodied feelings, where the embodied feelings play the role of intensifying and sustaining the evaluative judgments. This, of course, is highly pertinent to the question of whether (or to what extent) psychopaths are morally responsible for their actions. Psychopaths, as we have seen, experience general attenuation of the emotions, and particular deficits of anxiety and empathy. I have also suggested that they appear to lack the ability to see entities other than themselves as possessing the kind of value that would imply that their rights, interests and concerns provide reasons for action. If emotions are partly a matter of evaluative judgments and beliefs, then psychopaths' emotional deficits promise to provide an explanation for their deficit in seeing value.

However, the link between emotions and value turns out to be far from straightforward. Firstly, in the account I have presented, there is nothing special about the evaluative judgments and beliefs involved in emotion that confines them to the realm of emotion. Any of these judgments and beliefs can, theoretically, be made and held without embodied feeling, and therefore without emotion. In that case, it is at least theoretically possible for someone to judge something to be valuable without emotional engagement. Nor, when we examine cases, do we find that emotional engagement appears to be a necessary condition for something's being seen as valuable. Again, it is perfectly possible that all aspects of the cases with which the judge is concerned are important to her: the decisions she has to make, their outcomes and consequences. It seems clear, then, that at least actual, in-the-moment

emotional engagement is not a necessary condition of seeing something as important.

On the other hand, there would be something unusual about a judge who *never* engaged emotionally with cases, or with the issues raised by them, to the point where we would perhaps suspect her ability to make certain kinds of judgments – genuinely moral judgments, perhaps – as opposed to mechanistically applying laws and legal precedents to cases. Still more worrying would be a judge who had been genuinely unable, since childhood and perhaps for her whole life, to feel empathy, or deep emotional engagement with the world around her.

4.6 Psychopaths' emotions

In the previous chapter, I argued that psychopaths are incapable of seeing entities other than themselves as possessing value. We have also seen that psychopaths have an unusual, 'shallow' emotional experience, apparently lacking emotional engagement with the world around them. In this chapter, I have argued that emotional experience essentially involves both cognitive and feeling elements. A natural question to ask at this point, then, is, how should we characterise the 'shallowness' of psychopaths' emotional reactions in terms of the cognitive and feeling components of emotional experience? Are psychopaths missing the cognitive elements, the feeling elements, or both?

Robert Hare's descriptions of psychopaths are highly suggestive in this regard. Discussing fear, Hare makes the following remarks:

> For most of us, fear and apprehension are associated with a variety of unpleasant bodily sensations, such as sweating of the hands, a 'pounding' heart, dry mouth, muscle tenseness or weakness, trembles, and 'butterflies' in the stomach. Indeed, we often describe fear in terms of the bodily sensations that accompany it: 'I was so terrified my heart leapt into my throat'; 'I tried to speak but my mouth went dry'; and so forth.'

He goes on:

> These bodily sensations do not form part of what psychopaths experience as fear. For them, fear – like most other emotions – lacks the physiological turmoil or 'colouring' that most of us find distinctly unpleasant and wish to avoid or reduce.[30]

Earlier in the same chapter, Hare describes an individual whose reports of his own emotional experience illustrate this point:

> Another psychopath in our research said that he did not really understand what others meant by 'fear.' However, 'When I rob a bank,' he said, 'I notice that the teller shakes or becomes tongue-tied. One barfed all over the money. She must have been pretty messed up inside, but I don't know why. If someone pointed a gun at me, I guess I'd be afraid, but I wouldn't throw up.' When asked to describe how he *would* feel in such a situation, his reply contained no reference to bodily sensations. He said things such as, 'I'd give you the money'; 'I'd think of ways to get the drop on you'; 'I'd try and get my ass out of there.' When asked how he would *feel*, not what he would think or do, he seemed perplexed. Asked if he ever felt his heart pound or his stomach churn, he replied, 'Of course! I'm not a robot. I really get pumped up when I have sex or when I get into a fight.'[31]

Hare's descriptions suggest, not that psychopaths are incapable of bodily feeling, or have reduced capacity to experience bodily feeling in general (they 'really get pumped up' in some situations), but for whom the interplay of

---

[30] Hare (1995), pp. 55-6.

[31] Ibid., pp. 53-4.

cognitive activity and bodily feeling which is characteristic of emotional experience is missing, or greatly reduced. The suggestion is also that the missing element is primarily on the feeling side, not on the cognitive side. Psychopaths, it would seem, are capable of the beliefs and judgments involved in emotional experience, but in them the tendency for these cognitive elements to provoke physiological reactions is greatly reduced. In turn, the feeling element of emotional experience which intensifies and sustains it – in Hare's word the 'colouring' – is also greatly reduced. This explains the tendency – remarked upon consistently not only by Hare but also by other experts including Stout and Cleckley – for psychopaths' emotions to be 'bloodless', short-lived, and not deeply felt.

The feeling component of emotions intensifies and sustains emotional experience, and affects the depth of our commitment to the evaluative judgments which are the other component of emotion. I can appreciate a piece of music which does not move me, but if I experience emotional upheaval while listening to a piece of music, my evaluative judgments about that piece are likely to be more deeply held, and to mean more to me. Psychopaths, lacking the intensity of full-blooded emotional experience, are likely to experience less depth of commitment to evaluative judgments they make. If their interactions with others are lacking in feeling, any evaluative judgments they make about others are likely to be less deeply held.

Conclusions

Psychopaths' emotional deficiencies, then, lead to evaluative judgments which lack depth of commitment. This provides something of an explanation for psychopaths' tendency not to see value in others. Whereas non-psychopaths come to see others as having value in a way which is relatively deeply felt, psychopaths might only come to see this value in a way that is shallow and unmotivating. In turn, this might interfere with their ability to recognise

reasons stemming from that value, including reasons relating to the interests, rights and concerns of other people.

However, my conclusion in Chapter 3 went further than this by suggesting that psychopaths are actually *incapable* of seeing others as sources of value. The mere possession of attenuated emotions, and the effect of this attenuation on psychopaths' evaluative judgments is not in itself enough fully to explain this unusual pattern of valuing. Again, the judge example shows us this. The judge does not engage emotionally with the cases she presides over, and yet she is capable of making a series of complex and incisive evaluative judgments about those cases. She ascribes value to the people involved in the cases perfectly competently. (We can imagine a judge who, as a result of weary repetition and habituation, begins to forget the value of the people in her cases, but this *need not* happen, and the mere absence of direct emotional engagement in a case would not be enough to make it happen in respect of that case.) Therefore, people can make evaluative judgments that imply the possession of value by other people (and, by extension, non-human entities other than themselves) without those evaluative judgments being *emotional* evaluative judgments.

Nonetheless, again, it is less obvious that a judge who had *never* made emotional evaluative judgments about other people would be able to make reliable non-emotional evaluative judgments about other people. That would be a strange kind of person indeed, and one whose capacity to ascribe value to others we might reasonably question.

But why would we question this, and would we be right to do so? This is the topic of the next chapter. In it, I will present an account of the developmental role played by empathy in the practice of ascribing value to entities other than oneself, which builds on the account of emotional experience I have developed in this chapter. I will argue that psychopaths' inability to empathise, either

because of a genetic predisposition or because of a traumatic childhood or both, accounts for their unusual pattern of value ascription.

Chapter 5: Empathy and moral development

Introduction

There is an old controversy over whether or not empathy has a central role to play in morality. The controversy partly stems from the wider dispute around broad ethical and meta-ethical positions. Rationalists about morality are inevitably opposed to allowing a central role for something as apparently *emotional* as empathy. Those following in the Humean tradition are more likely to do so; Hume's own notion of 'sympathy' is similar in many ways to what we might now call empathy. Certainly, it would be odd if empathy turned out to have *nothing* to do with morality. After all, we are surely often motivated to do good, or to refrain from doing harm, by our feelings of empathy for other people. People I know whom I would think of as being particularly moral people – not moral fetishists or those who are very morally punctilious, just *good* people – are invariably people whom I also think of as having a large capacity for empathy.

That empathy can have a motivating role is self-evident. The more interesting question is whether empathy is in any way *necessary* for morality. This is particularly interesting in the context of psychopathy research, because as we have seen, psychopaths suffer from emotional deficits and from a deficit of empathy specifically. In a way, psychopathy constitutes at least weak empirical support for the proposition that empathy is necessary for morality. At least, there is a correlation here: psychopaths are not good at empathy, and they are also not good at morality. One possible explanation for this correlation is that the former capacity is a necessary condition of the latter. (Of course, this is not the only possible explanation.)

If there is going to be a plausible version of the proposition that empathy is necessary for morality, it cannot imply that, for every given instance of acting morally, or of making a moral judgment, there must be a corresponding event

of empathising. This is clearly not plausible. We very often act for the benefit of other people without actively empathising with them. Many acts are broadly 'moral' without being aimed at any particular person or group of people. Recycling one's plastic is (or can be) a moral act, but no-one goes around empathising with future generations while taking the bins out.

In fact, one might think that it would not be desirable for people to be empathising all the time while making moral judgments, because empathy might even interfere with their ability to make those judgments effectively. We know that empathy is subject to a number of biases – we tend to empathise more effectively with people from a social background similar to our own, for instance – so too much reliance on empathy might lead us to make judgments with a partiality that is inconsistent with the demands of morality; even with what we ourselves conceive of as the demands of morality. Another way in which our tendency to empathise can introduce bias is simply that we are more likely to empathise with those directly in front of us than with others who are further away from our immediate attention, but whose rights and interests might be equally or more important.[1]

In the previous chapter, I introduced the case of a judge who considers cases in court. I suggested that, for such a judge, too much direct emotional engagement with each case is likely to be a barrier to effective moral judgment. It seems to me that this point holds for empathy as a specific emotional process. We might imagine the judge lurching from one witness to the next, empathising strongly with each in turn. When it came time for her to make her judgment, the real or imagined emotional condition of those involved in the case would be so powerfully salient that it would be impossible for her to think clearly about the other important aspects of the case listed above. The idea that empathy

[1] Bloom (2016) gives a provocative account of the various ways in which empathy can interfere with morality.

could be a reliable *basis* for judgments in this type of case seems particularly hopeless. Where someone is convicted of a crime with a clear victim, and the maximum sentence is very long, what does empathy require of the judge? Empathising with the victim might lead her to hand out a higher sentence; empathising with the perpetrator might make her want to show restraint. And what about all of the considerations that have nothing to do with anyone in the courtroom? What about the principle of deterrence, for example, or the expectations and interests of the wider public?

In fact, it seems quite likely that the deliberations of effective legal judges are affected by empathy at most only intermittently. On the other hand, would we want a judge whose pronouncements were *never* tempered by empathy for the people they affected? This seems equally undesirable.

We therefore have a puzzle. Sometimes morality requires that we raise our attention from the immediate situation in order to apply principles and values at a much more abstract level. It seems likely that this kind of activity would be rendered much less effective if we were to try to empathise actively with all of the relevant people or groups of people when making moral judgments about them. Doing this would seem to be unnecessary, unrealistic, and also potentially counterproductive. On the other hand, a person who *never* managed to empathise would surely be less effective in making and acting on, or perhaps just less inclined to make and act on, moral judgments. So what exactly is the role that empathy plays in moral judgments?

In this chapter I will argue, based on empirical evidence, that empathy plays a developmental role in furnishing us with a capacity to ascribe value to entities other than ourselves. I will argue that the fact that this capacity is missing in psychopaths is a function of their lack of empathy, either because they lack the neurological hardware to empathise from the beginning, or because any natural capacity for empathy they may have withers away in childhood, and fails to

manifest in normal patterns of value ascription. In this way, I hope to build on the work of the previous chapter by explaining why psychopaths' emotional deficiencies interfere not only with their *tendency* to ascribe value to entities other than themselves, but with their *ability* to do so. In turn, this leads to an inability to recognise reasons which depend on the value of others, including reasons relating to the rights, interests and concerns of other people, as reasons, and hence bolsters my overall conclusion that psychopaths are not morally responsible for failing to act on these reasons.

5.1 What is empathy?

In order to get to the bottom of what role empathy might have to play in morality, we might think a good starting point would be to get clear on what we mean by 'empathy'. It turns out that this is not a simple task, and there have been disagreements both among scientists and among philosophers on this question.[2] In this section I will begin by drawing some central conceptual distinctions which are prominent in the literature on empathy, beginning with the more psychologically 'basic' processes and progressing to the more psychologically sophisticated. While the question of what processes or states should be called 'empathy' is an interesting one, it is for my purposes not so important as the question of which processes have a role in moral development, and which processes, lacking in psychopaths, explain their unusual patterns of valuing others. I will therefore remain agnostic on what exactly empathy is, and instead gloss a number of different processes, all of which I think have some claim to be considered for inclusion in the category of empathy, as 'empathy-

[2] A good starting point for philosophers who would like to engage with the material on empathy is Caplan and Goldie (2011). The introductory chapter provides a thorough and scholarly review of work on empathy from several disciplines: philosophy (including philosophy of mind, ethics and aesthetics), psychology (including clinical, developmental and social psychology), neuroscience and ethology.

like processes'.[3] In the following section I will then consider how these processes might contribute to moral development.

Empathy perhaps consists in a kind of transfer of emotion between people. However, not all forms of emotional transfer are effected in the same way, and it is not clear which should be called empathy. At the simpler end of the spectrum, there is a basic psychological phenomenon usually referred to as 'emotional contagion'.[4] This has been observed in very young children, including some only a few days old[5], and happens when a child observes the outward signs of a particular emotion in another child, and experiences that emotion herself. Child A and Child B are playing happily together. Child A drops a toy and starts crying, upon which Child B also starts crying, not because she has dropped her toy or has any other personal reason to feel upset, but simply because Child A is crying. This phenomenon does not involve the second child in any way 'taking the perspective' of the first. We know this because it occurs in children who are too young to have the cognitive resources to adopt the perspective of another child, or even to recognise that other children have identities separate from their own.

This simple emotional contagion is not unique to children but also occurs in adults. Imagine, for example, that you are sitting on a train working on a laptop while, in the seat behind you, a woman is speaking on the phone. She is audibly upset – perhaps she is talking to her friend about her recent breakup from a

---

[3] For simplicity's sake, and because I want it to be an open question whether any empathy is taking place, I will refer in example cases to the putative empathiser as 'the subject' and the person putatively empathised with as 'the target'.

[4] Eisenberg and Strayer (1987), Hatfield, et al. (1992), Wispé (1987), Hatfield, et al. (1994), Davies (2011).

[5] Field, et al. (1982), Haviland and Lelwica (1987), Fawcett, et al. (2016) and Waters, et al. (2014).

long-term boyfriend. Although focused on your task, and at no point taking the time to engage imaginatively with the woman's situation, it is possible that, over time, the woman's sad tone of voice would cause you to become sad yourself. In contrast, if she was laughing and joking and talking with enthusiasm about an imminent holiday, your own mood might become upbeat, again without actually giving any conscious thought to the woman's situation. These simple transfers of emotion, unmediated by imaginative perspective-taking, are examples of emotional contagion.

Contrasted with this are more sophisticated processes in which the imagination does come into play in adopting the point of view of others, and experiencing an emotional reaction, in some sense, either as if one were another person, or as if one were in that person's situation. An important distinction here is between what Amy Coplan calls 'self-oriented perspective-taking' and 'other-oriented perspective-taking'.[6] Self-oriented perspective-taking (or to use Peter Goldie's phrase, 'in-his-shoes perspective-shifting'[7] describes cases in which the subject imagines herself in the target's situation, but does not imagine herself to *be* the target; she does not imaginatively take on the psychological characteristics, dispositions and preferences of the target, only aspects of her situation such as, for example, the choice with which she is faced. By contrast, in other-oriented perspective-taking, the subject has to imagine actually *being* the target. This is a much greater imaginative feat: in order successfully to achieve other-oriented perspective-taking, the subject must imaginatively take on what she infers to be the target's psychological characteristics as well as her situation. Note in particular that this is distinct from merely logically inferring the other person's emotional state. It is not, as it were, saying to oneself, 'the target has a short fuse, therefore I imagine she would feel angry in this situation.'

---

[6] Coplan (2011).

[7] Goldie (2011), p. 309.

Other-oriented perspective-taking would involve the subject understanding the target to have a short fuse, among other psychological facts about the target, and allowing that understanding to influence her thought processes in a subjective, imaginative engagement with the target's situation in which she imagines herself to be the target. As Peter Goldie has pointed out,[8] however, such an imaginative engagement has at least one important limitation, since the subject would still need to have in mind some conception of the kind of person the target is, and such a conception of her own personality and character is only rarely a feature of the target's internal experience. This, argues Goldie, is a distorting factor in other-oriented perspective-taking.

Other-oriented perspective-taking is, without doubt, a very challenging imaginative process, but it is one which promises to yield results which one might suspect are not available through self-oriented perspective-taking, in cases where the person being empathised with is significantly different from the person doing the empathising. In Amy Coplan's example, she (hypothetically an introvert) attempts to empathise with her extrovert sister Bettie. Bettie has been spending a lot of time alone recently, and Coplan, imagining herself in this situation, feels happiness and contentment. She fails to feel what Betty feels (anxious and upset) because she does not imagine *being* Bettie, she merely imagines herself in Bettie's situation. Self-oriented perspective-taking has led her to fail accurately to model her sister's emotional state when other-oriented perspective-taking would have been a more successful strategy, and the result is confusion and miscommunication between the sisters.

However, there are a number of ways in which the hypothetical Coplan – let's presumptuously call her Amy – might seek to improve matters which stop short of other-oriented perspective-taking. One very simple approach would be to infer Betty's emotional experience logically and then simply to imagine

---

[8] Ibid.

experiencing the relevant emotion. Knowing that Betty is an extrovert, and knowing that she has been spending time alone, she might simply infer that this would be likely to cause Betty upset, and then imagine feeling upset herself. This would be enough to give Amy some kind of simulacrum of Betty's emotional experience. Another slightly more complex approach would be to manipulate features of her own imagined situation in a way which would be likely to reproduce an imagined scenario which is somewhat analogous to Betty's real situation, and which would be likely to produce a similar emotional reaction. Perhaps there is a threshold of time spent alone above which Amy herself would start to feel upset – months rather than weeks, say. Amy simply imagines having been alone for months, and successfully achieves an imagined emotional state similar to Betty's real state.

I think I probably engage in both of the processes described above from time to time, in an effort to empathise with people. Both more common and more intense, however, are episodes of empathy that are much more direct than this, but which do not obviously fall into the category of other-oriented perspective-taking either. In fact it is not clear that there is any perspective-taking, or imaginative reconstruction of another's emotional state, happening at all.

Parents, I suspect, tend to empathise particularly strongly with their children. My own experience of interacting with my three-year-old son is shot through with episodes of what I would think would be properly called empathy, but which do not feel like *imaginative* processes. Elijah is refusing to eat a bowl of pasta. Despite the fact that he regularly eats and enjoys pasta, there is something about *this* bowl of pasta that is very unappealing to him. As I try various methods – encouraging, cajoling, bargaining – to get him to eat it, he just digs his heels in further. He is feeling frustrated, upset and angry with me for trying to make him do something he doesn't want to do. Though I'm obviously annoyed that he's behaving like this and the pasta I cooked is going to go to waste, as I watch his face contort and the tears start to flow, I'm also

feeling upset *for* him. Now, this seems to me to fit neatly into neither the category of self-oriented perspective-taking nor that of other-oriented perspective-taking. I'm clearly not simply imaginatively putting *myself* into his situation and reacting accordingly, because if I was presented with a bowl of pasta I'd eat it happily or, if not, I would politely refuse, not cry and throw my spoon! But neither am I really imagining what it is like to be him. It's possible that I might get somewhere with trying to imagine what it is like to be a three-year-old who doesn't want to eat his dinner. It's true that I was a fussy eater as a child myself, which probably gives me some residual *sense* of what it is like to have that kind of reaction to food. But it has been so long since I experienced the world as a three-year-old does that I think it is a stretch to say that I am capable of imagining accurately what it is like to be one now. I do not think what I am doing here is imagining. Rather, it seems to me that I perceive his emotional state through his behaviour. I know him well enough that I don't need to interpret this behaviour consciously – I just perceive the emotions of which it is a manifestation. And, perceiving them, and being close to him as I am, I feel them, or some version of them, with him. This perceptual process is not imagination, but it is also not the simple emotional contagion I described earlier, since it involves awareness of the existence of another mind, and of its subjective experience, in a way that emotional contagion does not.

So far in setting out distinctions among empathy-like processes I have been varying what we might call the communicative component of these processes – the process by which emotional communication takes place between the subject and the target. As we have seen, the scope of this component can vary from cases in which there is no role at all for imagination (as in emotional contagion, and the process of non-imaginatively perceiving emotion in others) through cases in which one person imagines herself in another person's situation (self-oriented perspective-taking) to cases in which one person imagines herself to *be* another person, imaginatively taking on board what she infers are that

person's character traits and long-standing dispositions, and also more short-term aspects of her psychological state (other-oriented perspective-taking). Empathy-like processes, however, also have an emotional component, which can also be varied.

There is a controversy in the literature over whether in defining empathy one should insist on an 'affective match' between the subject and the target.[9] On the one hand, it might seem obvious that empathy consists in feeling on behalf of another person, and perhaps to a different degree, *just that emotion* that the other person feels. After all, empathy is often recommended as a way of gaining a better understanding of other people. If the emotion one feels when empathising with a person is not the same emotion that the person feels, then how can one claim to have come to a better understanding of that person through empathising? On the other hand, there are other processes which look similar to this in form and effect but which do not include an affective match, and it seems reasonable to include these at least in the gloss of 'empathy-like processes' that I have been using.

The psychologist Martin Hoffman argues for an inclusive definition of empathy as including all 'psychological processes that make a person have feelings that are more congruent with another's situation than with his own situation'.[10] While this definition is intended to include cases where the emotion felt by the observer is the same emotion as that felt by the observed, it is also intended to include many cases in which the emotion is different. Here are three examples of situations which might fall into this latter category:

1. Your friend has tickets for a new play that you would really like to see. You imagine yourself in his situation and feel excitement, despite the fact that he is

---

[9] See Davis (1996), Hoffman (2000), Preston and de Waal (2002).

[10] Hoffman (2000), p. 30.

uninterested in the play and is only going because he has been invited by another friend and doesn't want to cause offense.

2. You are watching someone walking on a tightrope high in the air. You imagine yourself in their situation and feel fear on that person's behalf, even though the tightrope-walker is a seasoned performer and feels no fear herself.

3. You observe a situation in which two people – a bully and his victim – are interacting. The victim, it would seem, is used to being bullied, to the extent that he feels no indignation or anger at his plight. He feels instead a kind of bruised acceptance. However, you, as an observer, imagine yourself in the victim's situation and, doing so, feel yourself becoming angry towards the bully.

All three of these cases are examples of self-oriented perspective-taking. This, presumably, is an inevitable feature of cases of successful, imaginatively mediated empathy-like processes in which the emotion felt by the subject does *not* match the emotion felt by the target. In cases of other-oriented perspective-taking, the subject is imagining herself *being* the target, taking into account relevant features of the target's psychology. In such cases, if the subject's emotion does not match that of the target, then there must have been some failure accurately to model the target's psychology, and so the attempt to empathise has not been successful.

The first case is really just another case of failed self-oriented perspective-taking ('failed' in the sense that it cannot lead to an accurate reading of the target's emotional state) similar to the case of Amy Coplan and her imagined sister. Nonetheless, it is presumably the case that the joy the subject feels at imagining herself about to see the play is 'more congruent' with the target's situation than with her own, since the subject is not actually about to see the play. Therefore, this would seem to count as empathy according to Hoffmann's definition.

In the second case, the emotion felt by the subject is again more congruent with the target's situation than with the subject's, because the subject is safe on *terra firma*. However, it is less clear that the mismatch between the subject's and the target's emotional state represents a failed process. This is because the process promises to tell the subject something useful about the target and her psychological state, as long as the subject has other means of becoming aware of what the target is really feeling. Imagining herself in the tightrope walker's situation, she feels fear. Knowing that the tightrope walker herself does not feel fear, she arrives at a better appreciation of the tightrope walker's courage, or perhaps of the power of her skill and experience to make less frightening a situation which would terrify most people.

The third example is, at first glance at least, perhaps more plausibly an instance of empathy than either the first or second cases. Indeed, it appears to be this kind of case that Hoffmann has in mind as a case of empathy without an affective match.[11] In this case, as in the tightrope case, the fact that there is no affective match between the empathiser and the target does not automatically suggest that the process is a failed one.

What these three cases suggest, it seems to me, is that the usefulness of empathy-like processes may not lie in their propensity to deliver an accurate depiction of the internal emotional state of the target, so that the subject comes to understand better how the target feels. Rather, the usefulness of these processes might lie in their propensity to lead the empathiser to engage evaluatively in a number of ways with the witnessed scene. In the tightrope case, the observer experiences certain emotions on behalf of the tightrope-walker – excitement, pride, concern – and comes to see her as having certain

---

[11] 'The empathy-arousing processes often produce the same feeling in observer and victim but not necessarily, as when one feels empathic anger on seeing someone attacked even when the victim feels sad or disappointed rather than angry.' (ibid.)

evaluative qualities – bravery, skill, grace. Moreover, she is justified in seeing her in this way. The observer's *appreciation* of the tightrope walker and her situation is heightened and sharpened by her engaging empathically with the scene. These are the successful cases. In the unsuccessful 'theatre' case, the subject simply fails truly to appreciate the character of the target's emotional state. It is possible to formulate an unsuccessful version of the tightrope case, however, simply by reversing the subject and target in the case. Now an experienced tightrope walker is watching someone doing their first tightrope walk. This person is terrified, but the observer, taking her perspective in a self-oriented way, and having done hundreds of tightrope walks herself, assumes she will be unconcerned and take it in her stride. Thus, the subject not only fails to appreciate the emotional state of the target – one of fear – but also fails to appreciate certain evaluative facts about the target – the courage it takes for her to be walking on the tightrope, for example.

Hoffman's definition of empathy as including 'psychological processes that make a person have feelings that are more congruent with another's situation than with his own situation' has the disadvantage that it includes some instances of what would appear to be *failed* processes – as in the theatre case. Nonetheless, the diversity of ways in which empathy-like processes can be *successful* – which I have briefly and far from exhaustively sketched – supports a broad conception of what types of process should be the object of our attention. At least this is true if the aim of our enquiry is to illuminate the role of empathy-like processes in moral development, and this is indeed the aim of Hoffmann's enquiry as well as my own. Rather than the idea of congruence, however, perhaps a more useful way to talk about what links empathy-like processes is that they are instances of a subject experiencing an emotion *on behalf of* the target. While it may not be possible to find a precise account of what it means to experience an emotion on someone else's behalf, it does seem to me to capture what allows successful empathy-like processes to be successful.

In the bully case and the tightrope case, the subject feels an emotion or set of emotions – anger and fear respectively – on behalf of the target, even though the emotion they feel is not the same as that felt by the target. In the theatre case, the subject is really feeling the emotion on her own behalf, rather than that of the target.

As well as the two imaginative processes discussed above, thinking of empathy as feeling something on someone else's behalf allows us to include the kind of non-imaginative process that I identified earlier in observing my own interactions with my son, since in that case I was still feeling the emotion on his behalf, even if I was not doing so through imaginative engagement. (This type of process, of course, would not be excluded by Hoffman's definition, since the emotion in question is more congruent with the target's situation that with the subject's.)

It is now hopefully beginning to become apparent how what I have been exploring here links to the argument of the previous chapters: that psychopaths have an inability to see entities other than themselves as possessing value. The examples above show a number of the ways in which empathy-like processes can help us to engage evaluatively with other people's points of view. Because emotions have a component of embodied feeling, which gives them an intensity and motivational force that is not present in mere judgments, aspects of the target's point of view become powerfully salient to the subject through the act of empathising. The sense that the target has value is a central feature of this experience. Through the act of empathising, it becomes part of the subject's worldview.

The ability to engage in empathy-like processes, then, is plausibly an important means of achieving normal patterns of value-ascription, and it is plausible to suppose that someone lacking this ability would be significantly disadvantaged in this respect. Still, this does not establish that we need empathy-like processes

in order to see people – and still less entities other than people – as possessing value themselves. To show how this might be true requires a focus on the developmental role of empathy. Again drawing on Hoffmann's work, the next section will concentrate on this role.

First, a quick recap. There are a number of mental processes that I have glossed collectively as 'empathy-like processes'. Distinctions can be produced by varying both the imaginative (if any) and the affective component of these processes. The imaginative component can be absent entirely (as in emotional contagion, or the phenomenon of perceiving someone as experiencing an emotion), or consist in imaginatively adopting another person's situation (as in self-oriented perspective-taking) or in imaginatively adopting both another person's situation and facts about their psychology (as in other-oriented perspective-taking). The affective component can be characterised by a match between the emotions of the observer and the object, or by no such match.[12] Writers have different opinions about which of these processes ought to be counted as empathy and which ought not. Martin Hoffman's broad definition would count any of these processes as empathy. Amy Coplan, in contrast, insists on 'three essential features of empathy: 'affective matching, other-oriented perspective-taking, and self-other differentiation'.[13] I have opted to ignore this debate, instead focusing on the *value* of empathy-like processes. The processes

---

[12] There are other distinctions to be made here, but the above will suffice as an overview of the territory as background to the discussion of moral development which follows. Coplan (2011) offers a very useful and more detailed taxonomy of empathy-like processes.

[13] Ibid., p. 6. Coplan's third condition is designed to exclude from her definition a process she identifies in some children, whereby the child imaginatively engages with another child and experiences an emotional reaction as a result, but fails fully to identify the other child as the source of this reaction, for example seeking to be comforted herself, rather than seeking to comfort the other child, in cases of empathetic distress.

I am interested in are those in which someone feels an emotion or set of emotions on someone else's behalf, and as a result is able to engage evaluatively with the situation of that other person more effectively than they would otherwise do. This is not only a useful set of processes for fully developed adults to be able to use – I will now turn to the importance of these processes in moral development, and particularly in achieving an ability to value others.

5.2 Empathy and moral development

The most complete account of the role of empathy in moral development is to be found in the work, already referred to in this chapter, of Martin L. Hoffman. Hoffman identifies a key role for empathy in the development of morally motivated behaviour in response to the witnessing of harm befalling another individual ('the bystander model') as well as in the inhibition of harm-causing behaviour in oneself ('the transgressor model').[14] For Hoffman, as for many developmental psychologists, empathy is a key factor in the formation of very many forms of moral or 'prosocial' behaviour, for example avoiding harming others ourselves, alleviating the harm caused to others, preventing harm by perpetrators, or taking action against perpetrators.

Hoffman's description of the developmental role played by empathy is complex. The picture is of a developmental process through which the child's parents (or other adults with significant caring responsibilities – I will use 'parents' as shorthand) employ the child's ability to empathise as raw material with which to encourage the development of a concern for other people. Hoffman's description of how this happens involves what he (in common with other psychologists) calls 'inductions'.[15] Inductions occur when the child either commits a moral transgression herself, or witnesses a moral transgression being

[14] Hoffman (2000), Parts I and II.

[15] I will eschew a discussion of why we should use this particular term and just accept that it is a technical term used by psychologists.

committed by another person. Inductions are discipline encounters between parents and children through which parents attempt to influence their children's behaviour:

> Inductions, like all discipline attempts, communicate parental disapproval of the child's harmful acts. This makes it clear that the child has done something wrong and arouses a certain amount of concern over parental approval. But unlike other types of discipline, inductions do two additional things: First, they call attention to the victims' distress, and by making the victims' distress salient they exploit an ally within the child, the child's empathic proclivity. That is, inductions activate certain empathy-arousing mechanisms.... In this way inductions elicit empathic distress for the victim's pain, hurt feelings, and (if relevant) suffering beyond the situation. Second, inductions are verbal communications that make the child's causal role in the other's distress salient. The child's processing that information under the proper conditions (optimal pressure) results in a self-blame attribution that transforms his or her empathic distress, at least partly, into guilt, that is, transgression guilt, in contrast to bystander guilt over inaction. In short, children's cognitive processing of inductions arouses empathic distress and transforms it into guilt.[16]

Inductions can take a variety of forms, but always contain two key elements: communication of parental disapproval and arousal of empathic emotion. Descriptions of the victim's emotional state, aimed at triggering an empathic

[16] Hoffman (2000), pp. 157-8 (italics author's own).

reaction in the child, vary in subtlety according to the child's ability to understand and internalise them:

> The earliest inductions point up direct, observable physical consequences of the child's action ('If you push him again, he'll fall down and cry'; 'It's uncomfortable when you walk on me, please let me lie here for a few more minutes'; 'If you have to defend yourself that's all right but you may not hit anybody with anything in your hand, you could really hurt them'; 'If you throw snow on their walk they will have to clean it up all over again'). Later, the victim's hurt feelings may be pointed up – at first simple feelings ('He feels bad when you don't share your marbles with him, just as you would feel bad if he didn't share his marbles with you') ... And still later, more subtle feelings ('He feels bad because he was proud of his tower and you knocked it down').[17]

Also available to parents are a number of ways of bringing attention to the moral implications of the victim's emotional reaction and the child's role in bringing it about:

> The harmful effects of the child's action may be mentioned indirectly ('He's afraid of the dark so please turn the light back on'; 'Try to be quiet, if he can sleep a while longer he'll feel better when he wakes up'). The victim's perspective may be implied by stating his intentions or legitimate desires in a way that indicates the child's antisocial behaviour was unjustified ('Don't yell at him. He was only trying to help'; 'Couldn't you let him have it for a few minutes just so he can look inside? He wants so much to look inside and I don't think he'll do any harm'; 'He was only taking his

[17] Ibid., p. 150.

> turn and he has a right to a turn, just as you do'; 'I won't allow you to hit her when she does something by accident. You must understand that it was an accident. She is too young to know what she is doing. She did not mean to hurt you.') And, finally, reparative acts may be suggested ('Would you tell your sister that you are sorry and try to make her feel better about it?'; 'Go over and pat him so he'll feel better'; 'Now I would like you to help him put it [the tower the child knocked down] back together').[18]

The range and variety of descriptions in the quotations above shows how empathy can play a role in making salient the perspective of other people in a wide variety of different types of encounter involving harm and transgression. Through repetition of these various types of encounter, with their accompanying empathic emotions, according to Hoffman, the child gradually begins to see patterns in the behaviour of other children, their own behaviour towards other children, the emotional reaction of those children, their own emotional states brought on by empathic processes, and parental approval or disapproval. When parental disciplinary efforts are consistent, these patterns form 'scripts' or 'generic event memories',[19] including predictive and explanatory links between the various elements of the pattern, and including the consistent message that other people's concerns, rights and interests are important, and that there are reasons to treat them in certain ways, for example to help them or to refrain from harming them. When parental disciplinary efforts are guided by moral principles, the child in turn will begin to form moral principles matching those of the parent. In this way, what starts with the use of empathy-arousal in disciplinary encounters ends with the development of fully-fledged moral principles in which other people are represented as

---

[18] Ibid., pp. 150-1.

[19] Ibid., p. 156.

important and valuable. While Hoffman does not write in terms of value, we can see that the sense that other people have value is a natural product of this form of moral development. When parents point to the upset caused to another child, say, as a reason for refraining from a particular action, this carries the implication that the child's being upset is something that matters. Thus, the value of other people is a feature of the principles that are formed through inductive parenting. The faculties that the parent is trying to encourage the child to develop are therefore a combination of motivational and epistemic. The child acquires the ability to know the effect of actions – primarily their own, but also those of other people – on others, and also the motivation to act in ways that promote some effects and avoid, nullify or mitigate others.

It is notable that the 'inductive' method of parenting described above, in common with other forms of parenting, apparently relies for its effectiveness on the child's wanting to please the parent; if the child does not care what the parent thinks, then the parent's efforts to get the child to see things from others' points of view and ultimately to value them will not have much traction. We might wonder, therefore, whether some hardcore psychopaths might be impervious to inductive parenting because they do not have this desire to please the parent in the first place, so that no number of inductive disciplinary interventions and encouragement of empathy would have any effect. Presumably this desire comes about in most children through parental bonding, and it may be that empathy-like processes play a role in this bonding, in which case, it is quite plausible that children born with low capacity for empathy might miss out on this crucial stage of development. One can also imagine cases in which children who do have a normal capacity for empathy, who then miss out on parental bonding for some other reason, fail to develop fully-fledged empathy or the capacity to value others because any inductive and empathy-based parenting they do receive fails to gain traction because they do not care about pleasing the parent. If this is indeed an accurate description of

development in some children, then we have a further way in which circumstances either at or very shortly after birth can lead to the attenuated value-ascription which is characteristic of the adult psychopath.

Hoffman presents an array of empirical evidence in support of his account of empathy's role in moral development.[20] This research essentially supports two separate claims. Firstly, that inductions are effective motivators of prosocial behaviour in the short-term, compared with other measures. Secondly, in Hoffman's words, 'with a high degree of consistency… the generalisation that mothers… who use induction produce children whose moral orientation is characterised by independence of external sanctions and guilt over harming others.'[21] That is to say, this approach is effective in the long-term at producing people who have internalised a worldview in which other people's interests are made salient by emotional reactions, and in which extrinsic motivations such as the threat of punishment are not necessary to motivate moral behaviour. In the terms in which I have been describing normal moral development, other people come to be seen as having value.

In this chapter, I am concerned with the developmental role of empathy in bringing about the capacity to value others in adulthood. The more important of Hoffman's two claims for my purposes, then, is his claim about empathy's long-term role. I will therefore pass over the evidence for the first claim and concentrate on what I think is the strongest example available of a study which supports the second claim. This study, by Krevans and Gibbs,[22] is broad-ranging

---

[20] E.g. Brody and Shaffer (1982), Crockenberg and Litman (1990), De Veer (1991), Hart, et al. (1992), Krevans and Gibbs (1996), Rollins and Thomas (1979), Kuczynski (1983), Sawin and Parke (1980).

[21] Hoffman (2000), p. 165.

[22] Krevans and Gibbs (1996).

and subtle in its design, and to my mind constitutes powerful evidence that empathy can have the kind of role described by Hoffman.

Krevans and Gibbs took data through questionnaires from 78 children aged between 11 and 14 years, their parents, and their teachers, relating to 1) the children's tendency to engage in prosocial behaviour, 2) the dominant disciplinary styles in the family home, and 3) their empathic responsiveness and maturity. Several measurement systems were used for each factor in order to avoid the limitations of any one system. Recognising the need for a clear and consistent understanding of prosocial behaviour, Krevans and Gibbs used five separate measures of this, all of which identify prosociality with altruism, and not merely with compliance with a parent or authority figure's wishes. The fifth of these measures consisted of data drawn from an experiment which they carried out themselves:

> Each child was promised a bonus of $1 and received ten dimes [while they were filling out questionnaires]. At the end... the child listened to a story about a child from a disadvantaged country.... The child was then given an opportunity to donate money to UNICEF, a charity which helps children who, like the one in the story, live in disadvantaged countries. In order to reduce extrinsic motivations for helping, an illusion of anonymity was created. Children were left alone to make their decision and were asked to put a sealed donation envelope in a collection bag whether or not they actually made a contribution.... The size of the child's donation served as an index of prosocial behaviour.[23]

By combining this with data from the questionnaires, the experimenters were able to build up a rich and detailed data set through which to measure each

[23] ibid., p. 3268.

child's tendency towards altruistic behaviour. The data on disciplinary styles distinguished between 'other-oriented inductions, that is, discipline which directs the child to attend to his or her victims' perspectives… power assertions, that is, discipline which attempts to change the child's behaviour through use of the parent's power over the child… and love withdrawals, that is, discipline which withholds parental approval or attention from the child'.[24] Data on this variable were gathered from both the children and their parents, so the data reflected both the parents' and the children's perspective on the dominant disciplinary styles in the home. Finally, the data on empathy measured both the child's level of empathic responsiveness, i.e. the strength of the affective reaction felt in response to another's plight, and the sophistication of that response. Sophistication was measured through the 'Empathy Continuum System'[25] which used film clips and questionnaires to gauge the subject's ability to engage in complex acts of empathy with fictional on-screen characters.[26]

The results of the study showed a correlation between empathy scores (all measures) and prosocial behaviour scores, suggesting, perhaps unsurprisingly, that children who are highly and sophisticatedly empathic are more likely to behave morally. The study also showed a correlation between children's empathy scores and the use of inductions (i.e. disciplinary interventions exploiting and seeking to encourage empathy) in the home. There are two ways of interpreting this result: either the use of induction is effective at encouraging a general capacity for, and sophistication in the exercise of, empathy in the child, or parents of relatively highly empathic children are more likely to use induction compared to other disciplinary methods because it is more likely – or

---

[24] Ibid., p. 3266.

[25] Strayer (1989).

[26] Interestingly, this measure also includes a 'match score' which measures the degree of affective match between subject and object.

they believe it is more likely – to be effective as a disciplinary method. It seems plausible to suppose that both of these things are true to some extent.

Finally,[27] the study showed that not only were a) highly empathic children more likely to engage in prosocial behaviour, and not only were b) children whose parents used induction more likely to engage in prosocial behaviour, but also c) highly empathic children were more likely to engage in prosocial behaviour if their parents used induction, and d) children whose parents used induction were more likely to engage in prosocial behaviour if they were highly empathic. The findings therefore support the proposition that empathy effectively *mediates* inductive parental discipline interventions.

Krevans and Gibbs' study is a powerful example of the empirical research supporting the hypothesis that good parenting, mediated by empathy, creates a pattern of behaviour in children motivated by an outlook in which other people are seen as valuable. The fact that this pattern of behaviour survives outside the context of parental discipline encounters, and therefore operates independently of threatened punishment or promised reward, suggests that the outlook encouraged through induction is internalised by children. However, this still does not show that this outlook is carried through to adulthood. It also leaves open the possibility that alternative routes to the same pattern of valuing are available to those whose childhood is *not* characterised by inductive discipline and/or who are less capable of empathy.

One way empirically to support this claim would be to show that those who lack the neurological resources to empathise effectively, and/or an upbringing characterised by the encouragement of empathy and its enlistment in discipline encounters do not, as a matter of fact, develop into adults who value other

[27] The study contains a number of other interesting results, but I am focusing here on the most relevant.

people. Given that the alternative (non-inductive) disciplinary strategies identified by Hoffman are 'love withdrawal' and 'power assertion', we might expect to find that children whose parents favour these approaches develop a worldview in which the provision and withdrawal of affection, and the exercise of power, are more salient to them as behavioural motivators than the value of other people. There does not seem to be much evidence either way for the former possibility, but there are some studies[28] supporting the proposition that unqualified power assertion as a parental strategy gives 'children a power-assertive model of how to behave when one wants to change another's behaviour'.[29]

Notably, this finding also fits with what many psychologists say about the moral outlook of psychopaths. Psychopaths tend to see the world in terms of power relationships, and in those cases where they do succeed in developing a rudimentary moral framework, it tends to be one in which moral authority is identified with the possession and exercise of power. In 2008, the philosopher Jonathan Glover carried out interviews with 20 people diagnosed with anti-social personality disorder in Broadmoor, attempting to piece together their moral outlook. One conclusion he drew was that psychopaths tend to have

> rather retributive, rather harsh moral views, which seem to be rooted not in sympathy for anyone else, [but] … often a command morality. 'Why do you think this?' 'It's because my parents told me' or, 'I was brought up to believe it.'[30]

To illustrate this, Glover gives the example of a psychopath who thought capital punishment should be brought back, but specifically for the crime of 'setting

---

[28] Bandura and Walters (1959), Hoffman (1960).

[29] Hoffman (2000), p. 147.

[30] Glover (2008).

fire to the Queen's property'. This description suggests that, in the absence of a moral outlook which includes the value of others, a moral (or perhaps pseudo-moral) outlook based on power assertion and authority might take hold. In this way, the existence of psychopaths can itself be taken as empirical support for Hoffman's theory of moral development.

The psychopathic personality has a complex aetiology almost certainly involving both 'nature' (i.e. genetic) and 'nurture' (i.e. upbringing) components.[31] Among environmental factors, physical abuse and neglect in childhood have been found consistently to contribute to the development of psychopathic traits. The existence of both a genetic factor and of a factor relating to physical abuse and neglect is consistent with Hoffman's theory. Children who are born with a low capacity for empathy might nonetheless increase their capacity for empathy, and build on it to form a mature moral outlook, if they have parents who are highly empathic themselves, encourage empathy in their children and use disciplinary interventions involving empathy. Conversely, children who are born with a capacity for empathy that falls within the normal spectrum, if they then suffer abuse or neglect in childhood, might have that capacity stunted by an experience that is not conducive to the development of fully-fledged empathy. We would also expect to find some individuals, at the extreme low end of the spectrum of capacity for empathy, for whom no amount of constructive, empathy-based parenting would be likely to have much effect. These individuals, having extremely little or even no capacity for empathy, would in effect be predestined to become psychopathic adults. In fact, all of these expectations are indeed consistent with what we do find in psychopaths. The two factors: genetic predisposition and abusive or neglectful

[31] Waldman (2007), Farrington (2007), Farrington, et al. (2010), Viding and Larsson (2010).

parenting have been shown to correlate with the development of fully-fledged psychopathy, though neither has been shown to be a sufficient condition.[32]

The key aspect of Hoffman's theory, at least as it is applied to psychopaths, is that what it describes is a participative, fully engaged developmental process. Empathy-like processes help to form children's moral outlook because they enable them to feel something of other people's emotional condition, or more broadly to feel emotions on other people's behalf. By doing this they come to see other people as valuable. The wrongness of harming others, and the goodness of helping them, flows from this first realisation of them as valuable beings. Thus empathy acts, negatively, as a check on potentially harmful behaviour, and, positively, as a motivator of behaviour that is likely to benefit another person or increase their wellbeing in some way. The 'scripts' that one develops through the exercise of empathy and through constructive parental interventions are therefore always scripts in which one appears oneself, not as an actor but as a character. By the time full-fledged moral principles are formed, these too are understood participatively, as principles which apply to one's own conduct. To deny that they apply to one's own conduct would be to deny that other people, whose rights and interests form part of the content of moral principles, have value.

We should not therefore be surprised that psychopaths are often able to piece together an understanding, however imperfect, of the moral principles by which most of us live. By carefully observing the behaviour and reported motivations of others, we would expect someone with a reasonable level of intelligence to

---

[32] Waldman (2007). It should be noted that the evidence for environmental factors is not conclusive because other factors, including genetic factors but also socioeconomic and other factors, have a confounding effect. For example, parents who themselves have a genetic predisposition towards psychopathic traits are more likely to be neglectful or abusive towards their children.

be able to develop this kind of understanding. However, without the intervention of empathy, this will never be a truly participative understanding, because it does not contain the value of other people. At most, a psychopath can come to understand that other people think of each other as having value, in the same way that the space travelling anthropologist we met in Chapter 3 understands that the aliens think of the plants on their planet as having value. This is far from the same as seeing – or feeling – that value oneself.

It also seems plausible to me that this ability to see value in other people is the basis of a broader ability to see value in entities other than oneself, which includes things other than people, such as animals, or other entities or ideas such as the environment or justice. Once we have fully-fledged moral principles we are then able to reason using them, to refine them and to form new ones. That animals have value of the relevant kind is a reasonable conclusion from the observation that they share many features with humans, including perhaps subjectivity. As it is emotional engagement with another person's subjective experience that enables us to see other humans as having value, the fact that (some) non-human animals, too, have subjectivity means that their value can be inferred, even if it is not made part of our experience through directly empathising with animals (although this is of course also possible). The conviction that other non-human entities such as the environment have value is plausibly the result of further abstraction from principles derived in the way described.

Clearly, an important question here is whether, having failed to develop a truly participative understanding of moral principles in this way in childhood, it would be possible for someone to develop one in adulthood. While it cannot be established beyond doubt that this is impossible, it does appear that, for the vast majority of individuals at least, such patterns set in childhood are not subsequently reversed. As has already been noted, psychopaths are notoriously recalcitrant to treatment, including treatment which has the explicit aim of

encouraging empathy in adult psychopaths.[33] Given this fact, it seems likely that the opportunity to develop a non-psychopathic outlook is lost at some point before adulthood, and that adult psychopaths are therefore incorrigible.

If empathy-like processes are the origin of our ability to see value in entities other than ourselves, then it can hopefully be seen why alternative routes to this value, such as through reason alone, would not be available to psychopaths. It is only possible to arrive through reason at the conclusion that some particular thing has value if one is in the habit of seeing things in general as having value in the first place. If one can only see things in general as being valued by other people, say, then reason can only supply the conclusion that the thing in question is valued by other people. The way real psychopaths think and act suggests that this line of thinking is correct: they have no deficit of reason, and yet they are fundamentally incapable of seeing value in entities other than themselves. If it were possible to 'reason oneself to' value then one would expect psychopaths to do so. The fact that they do not suggests that it is not possible.

5.3 Other disorders of low empathy

At this point, I have hopefully given some good reasons for thinking that psychopaths' deficits in the capacity to empathise, and/or their traumatic childhood experiences, lead to an inability to ascribe value to entities other than themselves, which in turn implies a lack of moral responsibility. However, there is an important objection that could be raised to the account I have been giving, and it is worth pausing to consider this objection before moving on. The objection goes like this: psychopathy is not the only unusual personality type that is characterised by low empathy. If empathy truly plays a key role in moral development, we would expect to find a parallel truncation of moral development in people who are not psychopaths but who nonetheless

[33] Harris and Rice (2007).

experience deficits of empathy, perhaps due to having a different form of mental illness. Indeed, there *are* other forms of mental illness which are characterised by difficulties with empathy, and it is not clear that we do find a truncation of moral development in people with these conditions. So how can it be that the lack of empathy in psychopaths is the cause of their own truncation of moral development?

Two other personality types which are characterised by low empathy are borderline personality disorder (BPD) and autism spectrum disorder. In order to respond to the objection, let us examine each of these in turn.

Here is a brief description of some of the key elements of borderline personality disorder, focusing on empathy:

> Individuals diagnosed with borderline personality disorder (BPD) are highly sensitive to other people's feeling states, but only as those states affect them. They possess an anxious egocentricity, which means that any capacity to empathize is severely reduced. There is no wish to understand the other person's mind, only an anxiety about the impact that the other's feelings and behaviour might have on them.... Individuals with BPD present with a complex array of symptoms, such as unstable moods, volatile social relationships and low levels of trust. The lives of people with BPD – around 1 or 2 per cent of the population – seem to be ones of perpetual crisis.... Episodes of depression are common among them.... Those with BPD suffer a pervasive fear of abandonment by *idealized others*. Therefore, although their need for others is high, trust in those others' emotional availability is low. Anxiety reigns and they are particularly sensitive to any hint of rejection. People diagnosed with BPD feel needy, unloved and vulnerable. They generally see themselves as victims and hard

> done by. Their relationships are characterized by intense feelings, chaos, confrontation and instability. Their behaviour is impulsive, unpredictable and self-destructive.[34]

The type of empathic deficit experienced by people with BPD, then, differs greatly from that experienced by psychopaths. Whereas psychopaths might be able accurately to represent another person's point of view in their imagination, but are unmoved by it, people with BPD care desperately about other people's points of view – albeit only so far as they bear on themselves – but their disorder causes them consistently to misrepresent this point of view in their imagination, creating fantasies of rejection and abandonment in place of whatever attitude the other person actually holds towards them.

Like psychopathy, borderline personality disorder appears to have a combination of genetic and social origins.[35] BPD shows a moderate degree of heritability, suggesting a genetic predisposition in some individuals.[36] However, instances of sexual abuse, neglect and traumatic experiences in childhood are much higher in people diagnosed with borderline personality disorder than in the general population.[37] It is plausible to suppose, therefore, that the problems with empathy associated with BPD may be caused either by an abnormal empathic pattern being present from birth, or by a relatively normal capacity for empathy not reaching full maturity due to traumatic experiences in childhood, or by some combination of these two factors. This would mirror the supposition I made about psychopaths' empathic development, except that whereas psychopaths' attempts to empathise can be

---

[34] Howe (2013).

[35] Leichsenring, et al. (2011), Gabbard (2005).

[36] Skodol, et al. (2002), Torgerson, et al. (2008).

[37] Zanarini, et al. (2002), Yen, et al. (2002).

accurate (in the sense of accurately representing others' mental states) but lack emotional colouring, similar attempts by people with BPD would be likely to have extreme emotional colouring but would be unlikely to be accurate. The person with BPD's attempt to empathise results only in an imaginative construction of the other person's point of view that is a product of the subject's disordered psychological state, rather than accurately representing the target's actual point of view. The affective component of empathy also then goes astray: imagining that the other person holds them in contempt, for example, the person with BPD feels fear, or resentment, or sadness in response. In reality, the other person might be feeling concern for them, perhaps combined with exasperation over their behaviour. Given that BPD (in stark contrast to psychopathy) is characterised by high levels of anxiety, this too will further derail empathic processes, crowding out other emotions and making the complex mental states associated with mature empathy all but impossible to achieve.

Working with Hoffman's account of the role of empathy in moral development, what effect on moral development would we expect BPD to have? If inductive parenting techniques are used, we would presumably expect them to be marred by the exaggerated and unrealistic attitudes to oneself imagined in others by the child with tendencies towards BPD. These imagined attitudes, and hence the other's imagined affective state, would also not be responsive to the child's actions in the same way that others' actual affective states are. For example, a child who steals another child's toy and who is asked to imagine how that child feels, only succeeds in imagining an exaggerated version of the other child's attitude to themselves, either positive or negative, which has nothing to do with the stolen toy. They would not, therefore, make a stable connection between actions like toy-stealing and the real affective states of other people. However, they might well form false associations between their actions and the reactions of others, which they have imaginatively misrepresented. We would therefore

perhaps expect people with BPD to develop a set of 'scripts' and for these to form into principles of a kind in adulthood. However, we would not expect these scripts and principles to match up with the kinds of scripts and principles developed by people without BPD. We would instead expect to find a moral outlook very much centred on others' supposed (in fact exaggerated or inaccurately represented) attitudes to oneself. We would perhaps expect to find the possibility of moral condemnation of others for negative attitudes towards oneself. We might also expect to find a sense that one ought to behave in ways towards other people that are likely to sustain positive attitudes and reverse negative ones. However, because the imagined attitudes are not responsive to one's actual behaviour, we would expect to find an erratic and inconsistent view of what these ways of behaving should actually be.

Most importantly, we would not expect to find that people with BPD fail to ascribe value to other people. As described, they do feel emotions on behalf of other people and, doing so, they should be able to construct a value-laden sense of the other's point of view. This will be less effective than the other person's in matching the affective and evaluative character of the other person's actual point of view, but it can still convey the value of other people in the way implied by Hoffman's theory. In terms of moral responsibility, it may be that people with BPD will not be morally responsible for some things because of their tendency to be mistaken about other people's motives, beliefs and so on. However, their lack of moral responsibility would not have the same explanation as that of psychopaths – an inability to ascribe value to others – and it is unlikely that they would lack responsibility entirely.

Another personality-type characterised at least partly by low or abnormal empathy is autism spectrum disorder. People with this disorder find social interaction very difficult. They tend to like predictability and are easily made to feel uncomfortable by anything which deviates from their accustomed routine. They often develop narrow obsessions about certain subjects, and they are

attracted by order and given to systematising behaviour. They also find empathising with other people very difficult:

> People who interact with autistic children sometimes feel as if they are being treated as no more than objects. Autistic children's empathy and communication skills seem poor. For example, most toddlers react with upset if an experimenter appears to have hit her thumb with a hammer, hurting herself and yelping with pain. In contrast, autistic children generally show little reaction to the experimenter's apparent distress. Children with autism fail to point to objects to achieve joint attention. They also remain uninterested in other people's emotional attitudes towards objects and events in the world. This can lead to social withdrawal.[38]

Unlike psychopaths, people with autism spectrum disorder do not lack emotional experience, although they lack insight into their own and others' emotions. Whereas psychopaths can imagine the world from another's point of view but feel no emotional engagement as a result, people with autism spectrum disorder have trouble imagining the world from another's point of view. They cannot therefore infer others' motivations and find people's behaviour unpredictable and bewildering as a result.

As the name implies, people with autism spectrum disorder exist on a spectrum from relatively mild to relatively severe symptoms. For example, about half of children with the disorder do not learn to speak.[39] People at the very severe end of the spectrum may even be permanently catatonic. However, among those who do have some degree of interaction with other people, people with autistic

---

[38] Howe (2013), p. 79.

[39] Ibid.

spectrum disorder can appear to be highly morally motivated. It is very rare for someone with autistic spectrum disorder ever to hurt another person except through involuntary actions, e.g. lashing out when upset. They 'rarely lie or attempt to deceive'[40] and many will become highly indignant at perceived moral infractions by others. If empathy indeed has a central role in moral development, why would this be so?

David Howe suggests that there may be another route to moral motivation of a particular kind that is available to autistic people that does not involve empathy:

> Their law-abiding behaviour is not solely based on the restraining powers of empathy but on the high value that they give to rules. Laws and rules make the world predictable. Breaching them destabilises conduct and behaviour and is not to be condoned. So although autistic individuals have low empathy, find relationships difficult, sometimes treat others as if they were objects and as often as not ignore those around them, they never behave intentionally cruelly or exploit others.[41]

This idea goes a long way towards explaining why autistic people can seem intensely morally motivated. Their strong negative reaction to moral transgressions by others derives not from an empathically derived concern for anyone who is being harmed, or whose rights are being violated, but from the fact that breaking moral rules is a way (one way among many) of becoming unpredictable, which is highly upsetting to autistic people. Plausibly, then, autistic people are able to understand morality at least as a system of rules – they identify certain forms of behaviour as 'against the rules' based on the

---

[40] Ibid., p. 82.

[41] Ibid., pp. 82-3.

reactions and behaviour of others, and internalise these rules as a way of making the world predictable.

An additional reason might derive from the nature of dishonesty, deception or manipulation, behaviours towards which people with autism spectrum disorder are notably not disposed. Such behaviours require the very capacities which autistic people lack: not the affective component of empathy, but the imaginative component. In order to try to manipulate someone, I need to have a sophisticated idea of how they will behave if subjected to certain interventions on my part, which in turn involves understanding their motivations. I also need to understand how to avoid detection, which involves understanding which considerations are salient to them and which will escape their attention. Without being able to put myself imaginatively in their situation, attempts to manipulate them will not get off the ground. It may be, then, that this type of behaviour is not so much inhibited in autistic people through moral disapproval, as simply outside the scope of what would be possible for them. In contrast, there is nothing about the condition of psychopathy that prevents a psychopath from understanding how others think, what is salient to them, what they take to be important considerations supplying reasons and so on. Therefore, there is nothing to prevent them from successfully manipulating people, as well as nothing to prevent them from *wanting* to manipulate people.

Perhaps the key difference between people with either borderline personality disorder or autistic spectrum disorder, and people with psychopathy, in terms of moral development, is that people with psychopathy have attenuated affective reactions across the board. Therefore, the salient motivational factors which act on them in childhood, and through which they develop a motivational pattern in adulthood, are more to do with satisfaction of their own desires and appetites. By contrast, people with the other conditions I have discussed are capable of intense affective reactions, with anxiety a central feature of both the disorders I have discussed. As a result, people with these

disorders develop a moral outlook (or perhaps we would want to say 'pseudo-moral', since it is an open question whether people with either disorder are genuinely motivated by moral considerations such as the rights and interests of other people) that is skewed towards disapproval of forms of behaviour that they would find anxiety-causing. In the case of borderline personality disorder, this is directed at the imagined and idealised other. In the case of autism spectrum disorder, it is directed at rule-breaking, and hence unpredictable, behaviour by others. Psychopaths, in contrast, have relatively little or no moral outlook.

There are other conditions associated to a greater or lesser extent with distorted, attenuated or absent empathy, including, 'Attention Deficit Hyperactivity Disorder (ADHD), Schizophrenia, eating disorders, conduct disorder… and Obsessive Compulsive Disorder (OCD)'.[42] There is not room here to investigate the effect on moral development of each of these conditions. By briefly discussing two of them, I have aimed to show at least that what we know about each of these conditions is consistent with Hoffman's theory of the role of empathy in moral development, and that this theory can tell us something about why adults with these conditions behave the way they do. It is reasonable to suppose, however, that any condition characterised by unusual empathy is likely also to be characterised by unusual patterns of responsiveness to reasons. This in turn may have implications for moral responsibility. Further philosophical work would be needed to map this out.

5.4 Alternative routes to reasons

I have given, based on Hoffman's work, what I think is the most plausible available account of how people, as a matter of fact, come to recognise the value of other people. In turn, this recognition allows people to be responsive to the

[42] Gillberg (2007) referenced in Howe (2013), p. 74.

various kinds of reason that depend on that value. Because of their depleted emotional resources, I have argued, this route to becoming responsive to reasons is unavailable to psychopaths, or at least to those occupying the low end of the scale when it comes to capacity for empathy. However, to show that this route to reasons is not available to psychopaths is not to show that no route is available, and this is what is required if I am to show that psychopaths genuinely lack the capacity to respond to reasons and therefore are not morally responsible.

There is of course a very popular philosophical position (or set of positions) which emphasises the role of rationality in morality. If some species of rationalism is true, then moral requirements are accessible through reason alone. If psychopaths are rational, then their emotional deficits should presumably not be thought to stand in the way of their coming to have access to the kinds of reason to which I am arguing they are impervious, and so we would presumably have to conclude that they are morally responsible after all.

Short of refuting rationalism – which is clearly beyond the scope of my project here – it is not possible to offer considerations that should conclusively close off this possibility. Nonetheless, there are things that can be said to support my case.

Firstly, let us note that there are several species of view that come under the overall banner of rationalism. To say that rational considerations – some form of the categorical imperative, let's say – form the overall criterion of rightness for morality is not automatically to say that a process of pure reasoning is the way that people do, as a matter of fact, come to understand what is right and wrong. It could be that what is right and what is wrong is ultimately determined by rational considerations, but that the way people come to see what is right and wrong is more akin to the process I have been describing in this chapter, involving emotions and empathy.

However, it can only be true that psychopaths are incapable of responding to reasons in a way that renders them not morally responsible if they *cannot be expected* to become responsive to reasons. That is, any routes to responsiveness which they could reasonably be expected to take must be closed off to them. It does not therefore matter, in determining the responsibility of psychopaths, if they cannot avail themselves of the route to responsiveness that most people do as a matter of fact take. If the faculties they do have – including rational faculties – offer a route that they could reasonably be expected to take, then they are responsible nonetheless.

The description I have set out in this chapter of how people come to see the value of others, and therefore of how they come to take them as presenting reasons for action stemming from that value, if accepted, gives a strong indication that at least the first claim above – that people as a matter of fact tend to come to be responsive to these reasons via a set of processes that essentially involve emotion and empathy – is true. That this set of processes is exhaustive of at least the routes to responsiveness that are typically available to most people is supported by the simple empirical point that psychopaths, lacking these routes, do not as a matter of fact come to see others as valuable and therefore to be responsive to the reasons in question. If there is an easily available route to responsiveness to these reasons that does not involve the faculties that psychopaths lack, then why do psychopaths not avail themselves of them? As we have seen, psychopaths are not (qua psychopaths) lacking rationality. If rationality were enough to see others as valuable then we would expect psychopaths to see others as valuable, but they do not.

Now, it is possible, I suppose, that moral rationalism offers a route to responsiveness that is obscure to most but, nonetheless, which it would be possible for psychopaths to come to see. Perhaps Kantianism is the correct way to think about morality and, if a psychopath were given a thorough introduction to the theory of the categorical imperative, she would come to see that, for

example, she has a good reason to refrain from harming others. Firstly, I am not optimistic that she would in fact come to see this. The categorical imperative is difficult to understand and controversial. Many people who spend time studying Kant do not understand it and many who do understand it do not accept it. To make moral responsibility rest on these foundations – or others like them – seems like a risky enterprise. Secondly, at most this would make responsible only those psychopaths to whom such an education was in fact available. It is not reasonable to expect people to seek out an education in the categorical imperative – or whatever other rationalist theory might be the correct one – by themselves, especially without already knowing that they have a good reason to do so.

If rationalism is an alternative route to reasons-responsiveness, it seems to me, it is too narrow and obscure a route to be available to the vast majority, if not the totality, of psychopaths. Without a comprehensive refutation of rationalism as a means of coming to understand morality, it will never be possible to say with certainty that such a route is unavailable, but the very fact that psychopaths do not avail themselves of it suggests that prospects for its availability are limited.

Conclusions

In the previous chapter, I drew a picture of emotional experience as richly combining cognitive and affective elements. In this chapter, we have seen that empathy, or empathy-like processes, fit into this picture. When empathising, through whatever mechanism, we come to feel emotions on other people's behalf. Hoffman's account of empathy's role in moral development gives us a convincing portrayal of how sophisticated principle-based thinking in adulthood, and the recognition of value in others, has at its root the exercise of empathy in childhood.

As I have already noted, it is key to understanding this developmental process that we see it as something that is experienced 'from the inside', so to speak. From our earliest experiences of social interaction, we have the capacity – and when exposed to 'inductive' parenting techniques, we are encouraged – to empathise with other people. The affective reaction we experience as a result of that empathising is recruited by our parents or other parental figures as a means of setting limits on our self-oriented behaviour, and of motivating other-oriented behaviour. In combination, discipline encounters and the experience of empathy show us that others have value, and thus that the rights, interests and concerns of other people present reasons that bear on our choices. Our experiences of interacting empathically with others outside of a discipline context give us further data on others' experiences, and further opportunities to build a library of situation types with accompanying motivational patterns. Gradually we develop 'scripts', which turn into principles, which generate reasons for action. As adults, we can refine our principles through moral reasoning, but also very often through the adult exercise of empathy. But because this process makes use of *our* motivations, *our* affective reactions based on others' reactions often to *our* behaviour, it is a process that is only fully experienced from the inside. It is as if we are building a house of morality with many rooms, but we are living in it as we build it.

Without empathy, hardcore psychopaths are denied the opportunity to see others as valuable, and therefore to see the rights, interests and concerns of other people as providing reasons for them. We can see that they *do not* in fact see the rights and interests of other people as providing reasons for them simply by observing case histories of psychopaths, but Hoffman's theory of moral development shows why this is so. And because the process begins in childhood and has reached an advanced stage by adulthood, we would not expect the truncated moral development of psychopaths to be something that can be easily ameliorated in adulthood, a prediction which is consistent with the poor

responsiveness of psychopaths to any treatment method that has so far been tried. This means that hardcore psychopaths are effectively doomed to be unable to recognise reasons arising from the rights, interests and concerns of other people from a young age, from causes (genetic inheritance and/or disastrous parenting) for which they cannot be held morally responsible. As a result, by adulthood, when the rest of us have constructed and are living inside a complex moral edifice, the best psychopaths can hope to achieve is to build an imperfect copy of such an edifice, which they will never be able to inhabit.

Chapter 6: What reasons do psychopaths have?

Introduction

To recap: in Chapter 1, I argued that to fulfil the conditions of moral responsibility, a person must be responsive to the reasons that bear on her choices, where this involves being aware of and recognising the relevant reasons (but also including cases of culpable ignorance, i.e. where one is not aware but can be expected to be aware) and being able to control her behaviour in the light of those reasons. In Chapter 2, I introduced the idea of the psychopathic personality-type, which I argued is defined by emotional deficits in general and by a deficit of empathy in particular. In Chapter 3 I put forward the view that psychopaths are not responsive to a certain class of reasons, namely reasons whose recognition depends on the ability to ascribe value to entities other than oneself. In Chapters 4 and 5 I tried to show that psychopaths, because of their emotional deficits and specifically their deficit of empathy, indeed lack this ability.

There remains, however, an important unanswered question. Does the category of reasons in question actually apply to psychopaths in the first place; that is, do they *have* such reasons? The fact that a mouse will suffer gives me a reason not to attack it, but a cat, presumably, has no such reason. What if the value of others turned out not to supply reasons for psychopaths in the same way that the value of a mouse supplies no reason for a cat to refrain from harming it? In that case, psychopaths would, *ceteris paribus*, be responsive to the full set of reasons which bear on their choices, a set which would not include any reasons of the type in question. They would not be morally responsible for these acts because they would have no reason not to perform them, in the same way that a cat has no reason not to harm a mouse. On the other hand, if reasons depending on the value of others *do* apply to psychopaths, then they have a set of reasons to which they are not responsive. They would then be more akin to

a person who is non-culpably ignorant of the reasons they have. In either case, they are not morally responsible, but the metaphysical underpinning of their non-responsibility would be interestingly different.

6.1 Internal and external reasons

It may seem that the answer to the question above is obvious. Of course, considerations related to others as bearers of value – their rights and interests, for example – supply us all with reasons for action. Why should psychopaths be any different? The fact that psychopaths do not recognise such reasons as applying to them, and are therefore not motivated to comply with them, should not imply that they do not in fact apply. However, there is a philosophically respectable position – internalism about reasons – which appears to imply just this.[1]

There are a number of different forms of internalism about reasons, but the basic idea is that for someone to have a given reason R, there must be something about the person – some desire, motivation or project, say – that would be satisfied by the agent's performing the action (or adopting the attitude, etc.) recommended by R. Thus, if one has a reason for action (and certain other conditions apply, to which we will turn shortly), one will always be motivated somewhat to act in the way recommended by that reason. If psychopaths are not motivated by considerations which depend on the value of entities other than themselves, then, on this internalist view, those considerations cannot comprise reasons for them.

The appeal of internalism perhaps lies in the 'democratic' view it gives of reasons – by making reasons depend on our own desires, motivations, commitments, plans or projects, the internalist gives some power to individuals in determining the moral requirements which apply to them. By contrast, the idea

[1] See e.g. Williams (1981), Korsgaard (1986), Smith (1995).

('externalism') that our reasons do not depend on our desires, projects, and so on appears somewhat austere. Nonetheless, I think a form of externalism is the right view. In the rest of this chapter, I will argue for this view, and consider how it applies to the case of psychopaths.

First of all, there is an important preliminary distinction to be made here. The kind of reason that is at issue in this debate is a *normative* reason, and not an *explanatory* reason. To borrow Thomas Nagel's explanation of this distinction:

> We may explain what a man does by referring to his [explanatory] reasons. On the other hand we may assert that circumstances *provide* someone with a reason to act in certain ways, without implying that he will be accordingly motivated (if only because of the possibility of his ignorance).[2]

Normative reasons are considerations that – at least if we are well-informed about them – apply as rational constraints on our choices and actions; they are considerations that bear on whether it is rational that we should act in certain ways.[3] They are thus distinct from explanatory reasons, which are the considerations that can be adduced to explain our behaviour and choices; they

---

[2] Nagel (1970), p. 15.

[3] There is a controversy over exactly how we should think about the relationship between normative reasons and rationality. Either rationality is a matter of conforming with our normative reasons, or with what *we take to be* our normative reasons. According to the latter view, rationality is a matter of good internal processing of reasons, and we may still be rational if we are significantly mistaken about our reasons, as long as we take proper account of the reasons we *think* we have. This controversy, which is played out in e.g. Kolodny (2005), Broome (2007), does not affect my argument here, but it is worth acknowledging that it exists. The same applies to the challenge to the very idea of a distinction between explanatory and normative reasons which has been made by Alvarez (2009). I am taking this distinction as read here and am drawing attention to it for clarity's sake, but I do not think anything I will be arguing for depends on it.

are the considerations that form (some of) the proper answers to the question, '*why* did person A perform action ϕ?'

Nonetheless, we will find that the same considerations in many cases act as both normative and explanatory reasons. Should I take a drink of this orange juice? The fact that I am thirsty gives me a normative reason to drink, so I take a drink. Why did I take a drink? Because I was thirsty; that was my explanatory reason for taking a drink.

On the other hand, there will also be many circumstances in which normative and explanatory reasons come apart. Not all of the normative reasons which speak in favour of my acting in a certain way will enter into the picture as explanatory reasons in subsequently explaining my acting in that way. The fact that the orange juice contains vitamin C gives me, a person with incipient scurvy, a reason to drink the orange juice. But that is not why I drank it; I drank it because I was thirsty. Finally, not all explanatory reasons are also normative reasons. I thought this was just plain orange juice, which explains my drinking it. But the fact that I thought it was just plain orange juice did not give me a normative reason to drink it, because I was wrong about that; it is in fact laced with poison.

So, the question that concerns us is about the *normative* reasons which apply to psychopaths. Let us use an example to illustrate what is at stake here. Suppose a psychopath, in the course of robbing a bank, has bound and gagged the bank's manager and left her in the corner of the room while he fills his bag with money from the tills. When he is finished, he is on his way out of the bank when he happens to glance over at the bank manager and notices that, because of the way he has applied the gag, she is having difficulty breathing. If he stops to loosen the gag, while he does not believe that the bank manager will be able to prevent him from getting away, it may allow her to call for help after he has gone, which may result in him having less time to make his escape before the

police arrive. The psychopath, let us say, does not care what happens to the bank manager. The relevant question, from the point of view of the argument that I will develop in this chapter, is: does the psychopath have any reason to stop to loosen the gag?

Note that it may well be the case that the psychopath recognises no such reason (and thus there is no prospect of its operating as an *explanatory* reason when we try to explain the psychopath's behaviour). The only considerations in the psychopath's mind may be those relating to his chances of getting away successfully. The relevant question is whether, regardless of not recognising a reason, the psychopath *has* a *normative* reason.

There is apparently nothing in the psychopath's existing set of motivations that might be thought to offer a foundation for such a reason. The psychopath does not desire to help the bank manager. The psychopath does not see the bank manager as having any value. The psychopath's aims and intentions would not be served at all by helping the bank manager. Nonetheless, we would perhaps like to be able to say that *anyone* in circumstances in which they can easily help someone who is in trouble, especially when that person is in trouble because of one's own deliberate actions, has a reason to help that person. To take the position that the mere existence of such a reason depends on whether one cares about other people is to deny that one can be subject to normative constraints of this kind – to moral constraints – whether one cares about other people or not.

The tension we feel when considering this question is at the heart of the debate between internalists and externalists about reasons. Answering the specific question of what reasons apply to psychopaths will require us to engage in this debate.

While internalism comes in many different guises, its basic structure is as follows. Internal reasons (if they exist) are reasons that one has in virtue of one's

existing set of motivational states; in Bernard Williams's terminology, *S*.[4] One has an internal reason to ϕ, where ϕ-ing is some action, if and only if ϕ-ing would serve some element of *S*, and *because* ϕ-ing would serve some element of *S*. If no element of *S* would be served by ϕ-ing, then one does not have an internal reason to ϕ. And, furthermore, it is the fact that ϕ-ing serves some element of *S* that explains the fact that one has a reason to ϕ. External reasons, by contrast, if they exist, are reasons one has that do not depend on the contents of *S*. Internalism is just the position that all reasons are internal and no reason is external. Externalism, minimally, is simply the denial of internalism, in other words it is the position that *some* reasons are external. However, perhaps wanting to avoid committing themselves to a mixed metaphysics of reasons, most externalists take the bolder position that *all* reasons are external.

Internalism is a thesis about the link between reasons and motivational states. If it is the case that one's having a reason depends on that reason's serving some element of one's existing motivational set, then it follows that one will always be motivated by one's reasons, providing some conditions are fulfilled. Internalists disagree about the content of these conditions, but they must at least include awareness of the facts constituting one's reasons. It is evident that one can have a reason to ϕ despite not being aware that one has a reason to ϕ, and that in such situations one cannot be motivated by one's reason to ϕ. Imagine I am the world's biggest fan of Dolly Parton and that, unknown to me, Dolly Parton is performing in my local park tonight. I have a reason to go to my local park which, on an internalist reading, is dependent on that element of my set of motivational states *S* which orients me towards enjoyment of Dolly Parton concerts. However, unaware of the impending concert and therefore of the reason for action that it supplies, I will of course be completely unmotivated by these concerns. Any plausible form of internalism, therefore, must make

---

[4] Williams (1981).

allowances for cases like this by making awareness of the relevant facts a condition of the truth of the general claim about reasons and motivation. Someone will always be motivated by her reasons as long as (at least) she is aware of the facts constituting those reasons, including facts about the means to her ends.

There may, however, be other conditions which must be fulfilled before the general claim is true. For example, it might be that someone must be fully rational in order to be motivated by her reasons, on the internalist view.[5] This condition is intended to account for cases in which irrationality of one kind or another gets in the way of motivation. Imagine, for example, that I am so depressed that even the prospect of a Dolly Parton concert cannot motivate me to get out of bed. Nonetheless, if I did manage to summon up the energy to drag myself to the park, the concert would give me a lot of pleasure (though my depression prevents me from realising this right now). It seems that I currently, in my depressed state, have a reason to head to the park, and it would still be the case that doing so would serve an element of *S*, but my depression, because it makes me irrational, blocks my motivation by the relevant reasons.

It is easy to see how other forms of irrationality, such as delusion, might block one's motivation to ϕ in similar ways, despite the fact that ϕ-ing would serve an element of *S*, and despite one's plausibly having a reason to ϕ. I might suffer from a deluded belief that the local park is subsumed in a force field that renders country music inaudible, so that there would be no point in my attending the concert, and I might not, as a result, have any motivation to go. Because there is in fact no such force field, I would still have a reason to go.

---

[5] E.g. Smith (1994).

The relationship between one's reasons for action and one's motivational set *S*, then, on the internalist reading, is conditional, at least on knowledge of the relevant reasons, and most plausibly also on some notion of rationality.

Williams's idea of 'motivational states' – the contents of *S* – can be cashed out in a number of ways, including 'desires…dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent.'[6] The particular manifestation of internalism with which I am concerned here, however, has to do with the recognition of value in others. I am concerned with whether a reason for action of the type that depends for its recognition on the recognition of others as having value (let us say for simplicity's sake, 'a reason based on the value of others'), can exist if the person for whom it is putatively a reason does not in fact recognise others as having value.

Can we show that psychopaths have such reasons, despite apparently not recognising them?

One way to show this would be to show that, even if internalism were true, then psychopaths could still have *internal* reasons based on the value of others. This strategy might appear to derive some plausibility from the fact – already alluded to above – that additional conditions are usually built by internalists into the claim that having a reason implies being motivated by that reason. If internalism says, for example, that someone will be motivated by their reasons unless they are not fully rational, then perhaps we might make *not being a psychopath* a condition of full rationality. After all, other psychological conditions such as depression and delusion already occupy this space. Why not psychopathy too? Indeed, as we have seen, psychopathy does interfere with its possessor's motivation by her reasons, albeit reasons of only a particular kind.

---

[6] Williams (1981), p. 105.

One problem with this move is that the other conditions that have been discussed in this context by Smith and others have been conditions which interfere with the agent's motivation by reasons in general, and not to their motivation by any particular kind of reason. It is easy to see why depression, for example, can be characterised as resulting in a form of irrationality in this sense; it is a pathology which includes a general deficit in the ability to be motivated by, and perhaps also to recognise, practical reasons. Psychopaths, by contrast, have no problem recognising and being motivated by many other forms of reason, including reasons stemming from their own self-interest.[7] Thus it is less clear that psychopathy should be called a form of irrationality. As an analogy, consider someone who happened to have no aesthetic response to music. As such, they would lack access to a broad set of reasons that would be available to most people. Would we be inclined to think of such a person as irrational? I would suggest not. They would be unusual, but their lack of aesthetic response would be better thought of as an element of their personality – an extreme instance of the preferences and patterns of response that make us who we are – rather than as something blocking their personality from reaching its full expression. Most importantly, it seems clear to me that such a person would have no reason, for example, to listen to a great piece of music.[8] The greatness of the music, which would present a reason for the rest of us, would present no reason for them. Whereas depression blocks access to some reasons,

---

[7] As we saw in Chapter 2, there is controversy over the question of whether a deficit of motivation by long-term self-interest might be constitutive of psychopathy, thus this might be another specific source of reasons by which psychopaths are relatively unmotivated. Nonetheless, the point holds that psychopathy is not a condition characterised by a general lack of motivation by reasons.

[8] This assumes, of course, that the act of listening to the music would not itself help the person to develop an aesthetic response to music. Such a total lack of aesthetic response – the result of a hardwired condition, perhaps – would be highly unusual but still, I submit, not a form of irrationality.

and therefore is plausibly thought of as a form of irrationality, a lack of aesthetic response only nullifies them, and therefore is not. Whether psychopathy nullifies or blocks reasons is precisely the question I am attempting to answer in this chapter – I cannot assume that it nullifies them without begging the question.

The strategy suggested above misunderstands the purpose of building factors like rationality into a formulation of internalism. The idea is to show that reasons can still depend on elements of one's subjective motivational set *S*, despite one's not actually being motivated by them. People suffering from depression or delusions can have elements of *S* that would be served by their ϕ-ing, but not actually have any motivation to ϕ because their condition gets in the way. To refer again to our previous example, the fact of my being a huge Dolly Parton fan still means that I have a subjective motivation that would be served by my going to see Dolly in the local park, despite the fact that my severe depression means that I am not at all motivated to get out of my armchair and do so. If I were not depressed but suffered from delusions, I might think that the local park was subsumed within a country-music-neutralising force field, and fail to be motivated to go there for this reason. But it would still be the case, by hypothesis, that I have a reason to go there, because there is no such force field.

Psychopaths, on the other hand, do not have a motivational attitude disposing them towards action that recognises the value of others. It is not that psychopathy 'gets in the way' of its possessor's being motivated by considerations that bear on contents of her subjective motivational set *S*. As we have seen, there is just nothing in a psychopath's *S* that would be served by her taking account of, say, the interests of others. Psychopaths apparently therefore fail to have internal reasons based on these concerns according to Williams's account of internal reasons.

6.2 Smith and convergence of desires

Another possible avenue of attack for someone who wished to show that psychopaths have internal reasons based on the value of others would be to show that everyone has such reasons, and that psychopaths have them *a fortiori*. This might seem to be implied by the idea of *convergence* employed by Michael Smith.[9]

Discussing Smith's ideas introduces a terminological complication which we should deal with before proceeding. Although Smith's ideas are a development of Williams's, talk of 'the agent's subjective motivational set *S*' is replaced by talk simply of 'desires'. This is because Smith has a broad, neo-Humean conception of desires which identifies them with mental states with a world-mind direction of fit, involving dispositions to take whatever actions the agent believes are likely to bring about the object of the desire.[10] Thus all motivating attitudes, for Smith, inherently involve desires, and it is the agent's desires which, ultimately, supply her with both motivation and reasons for action. I am concerned with the recognition of others as possessors of value, and whether reasons based on the value of others depend on this recognition. For Smith, it is the desires accompanying this recognition and implied by it – desires to help people, for example, or not to harm them – which are candidates to supply the agent with the relevant reasons. Because she does not recognise others as valuable, the psychopath lacks the accompanying desires, and thus a simplistic version of internalism would suggest that she cannot have the reasons. However, for Smith, reasons are based, not on the actual desires that we have, but on the hypothetical desires that we would have if we were fully rational.[11] If

---

[9] Smith (1994), Section 5.9.

[10] Ibid., Chapter 4.

[11] Smith's position is presented as a development of Williams's original argument for internalism.

psychopathy can be shown to be a form of irrationality, then they still may have the reasons in question, despite not having the corresponding desires.

Smith points out (and Williams would agree)[12] that desires are subject to rational assessment. As rationally deliberating agents, we may start with one set of desires, but we will submit those desires to a process of *systematic justification*, verifying that they are compatible with each other, as well as with our view of what is important, our ethical commitments, our beliefs about the world, and so on. The aim is to arrive at a position of reflective equilibrium amongst our desires and the other aspects of our psychology, our circumstances, and so on. Now since, also according to Williams, there needs to be a 'sound deliberative route' from our desires to what we take to be our reasons, for these to really count as reasons, Smith draws the conclusion that our reasons are functions not of our actual desires, but of the desires that we would have, hypothetically, if we were fully rational. Smith then makes the optimistic claim that, given the process described above, there would be a convergence amongst the desires of fully rational agents. There is therefore also a convergence amongst our normative reasons: although they are internal in the sense that they are ultimately a function of our desires, albeit our hypothetical desires, they are *objective* in the sense that we all share the same basic set of reasons.

Although Smith does not present anything that purports to be a knock-down argument in favour of this position, he does offer a number of points which are intended to boost the plausibility of what might at first appear to be a rather counter-intuitive claim. The key point here is that we can only expect a

[12] Williams (1981), pp. 101-2.

convergence amongst our reasons, according to Smith, once we 'abstract away from some complications'.[13]

One of these 'complications' is the set of circumstances in which the agent acts. While it is implausible to suppose that all agents have a reason to ϕ, *simpliciter*, it is more plausible to suppose that all agents may have a reason to ϕ in C, where C stands for a particular set of circumstances. So, for example, if my cat is stuck up a tree, then I have a reason to climb the tree. Clearly it is implausible to suppose that this reason, just expressed like that, is universally shared. However, it might be more plausible to suppose that, in circumstances where one's cat is stuck up a tree (and perhaps where other circumstances are in place as well, for example the tree is fairly easy to climb, one is able-bodied, not carrying an injury, fairly good at climbing trees etc.) then one does have a normative reason to climb that particular tree.

The fact that one's cat is stuck up a tree is (in Derek Parfit's terms) an *agent-relative* reason, as contrasted with an *agent-neutral* reason.[14] It is a reason which contains an embedded reference to the agent to whom it applies. It matters, from the point of view of the reason's being a reason for me, for example, that it is *my* cat that is stuck up the tree. The fact that my cat is stuck up a tree might also give *you* a reason to climb the tree, but only an agent-neutral reason, the kind of reason that might apply to anyone who happens to be passing by. The fact that it is my cat, and not your cat, and not some other cat, gives me a special reason, an agent-relative reason, to climb the tree.

Here is the example that Smith uses to illustrate this distinction and its importance:

---

[13] Smith (1994), p. 166.

[14] Parfit (1984).

> Suppose you are standing on a beach. Two people are drowning to your left and one is drowning to your right. You can either swim left and save two, in which case the one on the right will drown, or you can swim right and save one, in which case the two on the left will drown. You decide to swim right and save the one and you justify your choice by saying, 'The one on the right is my child, whereas the two on the left are perfect strangers to me'.[15]

There are both agent-neutral and agent-relative reasons operating in this example. The fact that two people are drowning on your left gives you a reason to swim to the left. This is an agent-neutral reason that would be available to anyone else who found themselves in the same situation. It would, for example, be a reason for me to swim left if it was me instead of you on the beach, and the same people were drowning. The fact that someone is drowning on your right gives you a reason to swim to the right, and would also give me a reason to swim to the right, again an agent-neutral reason. However, the fact that it is *your child* who is drowning on the right gives you a special, agent-relative reason to swim to the right, and this is a reason which would apparently not be available to me if I were in the same situation. (Recall that Smith is concerned with explaining an agent's reasons as being based on the desires that she would have if she were fully rational. It would seem that if I were in your situation on the beach, regardless of whether or not I were fully rational, the identity of the person to the right would not confer on me any special desire to save that person, beyond the desire that I would already have just to save *a* person. Hence it would seem that I would therefore have no special reason based on such a desire.)

However, there is, as Smith points out, another sense in which the agent-relative reason would also be available to me if I were in the same situation,

---

[15] Smith (1994), p. 169.

because *you* are in a situation in which it is *your child* who is drowning to the right. If I were in your situation, in this sense, the child who was drowning to the right would be my child. Now, Smith's claim is obviously not that (hypothetical, rational) desires would converge if we do not hold fixed the agent-relative features of a case (e.g. the fact that the child who is drowning is the child of the person making the decision). Clearly, neither our desires nor our reasons would be the same if this were the requirement. The claim, rather, is that the desires that we would hypothetically have if we were fully rational, and therefore our normative reasons, will converge in cases where the agent-relative features of the case are held fixed. If I were in your situation *and that were my child* then I would desire to rescue that child, and I would have a special reason to do so.

Smith goes beyond this by also building into the 'circumstances' of the agent, from which we should 'abstract away', some aspects of her own psychology. In Smith's example, 'Preferring wine, as you do, you may tell me that there is a reason to go to the local wine bar after work for a drink, for they sell very good wine. But then, preferring beer, as I do, I may quite rightly reply, "That may be a reason for you to go to the wine bar, but it is not a reason for me."'[16] If a normative reason is a reason to ϕ in circumstances C, then, claims Smith, we should build aspects of the agent's psychology into a specification of her circumstances, where this can include such things as a preference for wine over beer or *vice versa*. It is much more plausible to suppose that, for any rational agent who has a preference for wine over beer, and other relevant aspects of the agent's circumstances being equal, the fact that the local wine bar sells very good wine would operate as a reason for that agent to go to the local wine bar.

Smith's overall strategy is to show that our tendency to think that our desires – even the desires that we would have if we were fully rational – would not tend

---

[16] Ibid., p. 170.

to converge, is based on a crudely 'Humean' view of desires as unquestionable and extra-rational. Once we see that our desires are themselves subject to rational assessment, and understand the implications of this, and also see that they depend on facts about our psychology that are properly thought of as aspects of our circumstances, then we also see that the hypothetical desires that we would have 1) in a given set of circumstances C and 2) if we were fully rational, will tend to converge amongst different agents.

How does this overall strategy look if we apply it to the case of psychopaths? As we have already seen, the strategy of simply designating psychopaths as 'irrational' and therefore giving them a set of hypothetical desires that are the same as those of non-psychopaths because of their nonfulfillment of condition 2 above, is not promising. But what about the other condition above? Psychopaths' inability to see other people as valuable is, after all, an 'aspect of their psychology'. Should we then say that it is part of the circumstances from which we are entitled to 'abstract away' based on condition 1? Is it, in short, like a preference for wine over beer?

It is not clear that we can reasonably say this. Even if we accept that things like a preference for wine over beer are part of an agent's circumstances, we might question whether the same is true of traits as apparently fundamental to an agent's psychology as psychopathy. Should a pathological inability to see others as valuable really be seen as 'part of the agent's circumstances' in this way?

A preference for wine over beer might well be liable to change over the course of one's life. One might start off with such a preference, and then have it reversed following a bad experience with some wine, for example. By contrast, as we saw in Chapters 2 and 5, there seems little prospect for fundamental change in the psychology of hardcore psychopaths once they have reached adulthood. Moreover, in a hypothetical case in which a psychopath was 'cured' of her psychopathy, it seems clear that a fundamental part of her personality

would have changed. Indeed, this is why psychopathy is called a 'personality disorder'. A preference for wine over beer is not part of someone's personality (although it may be reflective of her personality in some way). And it is surely counter-intuitive to think of someone's personality as being part of their circumstances. Though it is just possible to imagine saying to a psychopath, for example, 'If you were in my circumstances, and saw the value of other people, you would not desire to hurt them,' this is surely not the right way to describe the hypothetical situation. Person X, a psychopath, suddenly able to see the value of other people, would in a sense no longer *be* Person X.

More importantly, however, doing this would not actually offer us any help in terms of our overall project, because it would not guarantee that psychopaths would have reasons based on the value of others, in the same way that I, not liking wine, would not have a reason to visit the local wine bar. It might very well be that rational psychopaths would have the same basic set of desires as each other, and that rational non-psychopaths would have the same basic set of desires as each other, in the way optimistically supposed by Smith. But it seems clear that rational psychopaths would have a different set of desires from rational non-psychopaths, because those desires would be unaffected by considerations based on the value of others. Therefore, if we are to think of psychopathy as part of one's circumstances, even if Smith's overall claim is correct and reasons are universal in his sense, then we are still left with one set of reasons for psychopaths, and another for non-psychopaths.

In attempting to use Smith's argument to show that psychopaths have internal reasons based on the value of others, we are left with a dilemma that cannot, as far as I can see, be resolved. Either psychopathy is part of an agent's circumstances or it is not. If so, then we only have internal reasons based on the value of others in circumstances in which we are not psychopaths, which is the same as saying that psychopaths do not have internal reasons based on the value of others. If not, then Smith is wrong in his overall claim that, once we

abstract from the circumstances of the agent, and from irrationality on the part of the agent, then we will end up with a universal set of desires, and therefore reasons. And he is wrong specifically because psychopaths, though apparently rational, do not have the same desires based on the value of others, as do non-psychopaths. Therefore, they do not have internal reasons based on those desires.

6.3 Reasons and wrongness

The idea that the desires of rational beings will 'converge' is, I think, dubious in general. But in the case of psychopaths it seems to me it cannot be made to do the job of providing psychopaths with internal reasons based on the value of others. It seems, therefore, that if we are to make room for the possibility that psychopaths have reasons for action based on the value of others, then we will need to meet internalism head on.

Unfortunately, a fact about psychopathy that we have already encountered will cause us some problems here. Some of the more powerful arguments against internalism rest on intuitions, which seem to be fairly widespread, and which hold across a wide range of cases. However, as I have previously noted, some intuitions which are otherwise quite robust, tend to founder when applied to the case of psychopaths. This difficulty with intuitions, which centres around the concepts involved in moral responsibility, blame and so on, causes problems for arguments against internalism.

One particularly powerful objection to internalism rests on claims about the connection between moral wrongness and reasons, and a rejection of the idea that moral wrongness should depend on the desires of the agent. The objection can be put into the form of an argument, thus:

1. Some actions are morally wrong for an agent no matter what motivations and desires they have (moral absolutism).

2. An action is morally wrong for an agent only if there is a reason for him not to do it (moral rationalism).
3. Therefore, there must be some reasons for action that an agent has that do not depend on what motivations and desires he has (external reasons).[17]

This argument is powerful because it appears to show that rejecting the existence of external reasons entails rejecting one or other of two theses which many would not want to reject: moral absolutism and moral rationalism.

To deny moral absolutism is to commit oneself to the position that the conditions determining the moral rightness or wrongness of an action are always partly constituted by what the agent wants. While this is not an obviously unacceptable view, and several philosophers (including Bernard Williams) have espoused versions of it, it is certainly true that it contradicts assumptions that appear to be deeply embedded in the way we think and speak about morality. Take the following dialogue for example:

Debbie: Look, I'm not happy with the way you spoke to Kim this morning. I really thought it was very cruel.

Joni: Oh no, it's okay. You see I wanted to be cruel to Kim.

Debbie: But I just thought you were treating her with a basic lack of respect.

Joni: But I don't think there's anything wrong with that because I have no desire to treat Kim with respect.

Etc.

[17] I have taken this formulation from Finlay and Schroeder (2012).

Joni's replies to Debbie's criticisms appear to be, not just inadequate, but entirely on the wrong track. The point behind what Debbie is saying is that it doesn't matter what Joni *wants*; she has certain duties towards Kim which she has failed to recognise and live up to. There is no point at which appealing to her own desires is ever going to convince Debbie that Joni has not acted immorally. Even trying to do this strikes us as absurd. So someone who would deny moral absolutism will have to provide an explanation of why either 1) Joni's argumentative strategy is, despite appearances, sound, or 2) their position does not in fact entail that Joni's argumentative strategy is sound. If 1, they would also need to provide some kind of explanation of why Joni's approach *looks* so wrong-headed, compatible with its not being wrong-headed, which would need to be at least as plausible as the explanation that it is wrong-headed. 2 could perhaps be achieved by presenting an argument akin to Michael Smith's argument discussed above. I have already said why I find this approach problematic.

This is not to say that a suitable explanation of the kind described above cannot be found. But it is a considerable task, and in the absence of overriding reasons for preferring an account which has moral rightness and wrongness depend on the desires of the agent, an account which justifies deeply entrenched intuitions about the basis of moral rightness and wrongness would seem to be preferable to an account which renders them baseless.

Someone who rejected the view labelled 'moral rationalism', on the other hand, would be committed to the claim that an action can be morally wrong for an agent to perform, without that agent having any reason to refrain from performing that action. This too is a somewhat counter-intuitive claim, and one which again appears to contradict the way we tend to think morality works. It is strange to imagine someone accepting that an action is morally wrong, and yet rejecting the idea that there is a reason against performing that action. There may be countervailing reasons on the other side of the equation, and the

reason stemming from moral wrongness may not be the overriding reason, but the idea that no such reason might exist is bizarre. To adapt the dialogue above, imagine Debbie manages to convince Joni that the way she treated Kim was morally wrong. It would be very strange if Joni then said, 'Okay, but you still haven't given me a reason why I shouldn't have treated Kim like that.' The fact that treating Kim like that was morally wrong just *is* the reason why Joni shouldn't have done it. Asking for a reason on top of this accepted fact appears to reveal that one has not really understood the nature of the fact in the first place. It appears to be entailed by an action's being morally wrong for an agent that this fact constitutes a reason for that agent not to perform it.

But perhaps it is bizarre to imagine someone asking for a reason why they shouldn't have performed an action that they accept was morally wrong, because we tend to make a natural assumption about people, including those we imagine as part of thought experiments: we assume that they care whether something is morally wrong or not. Imagine Debbie convinces Joni that the way she treated Kim was morally wrong, and Joni's reply is, 'Okay, but given that I don't *care* whether the way I treat people is morally wrong or not, can you give me a reason why I should have treated Kim differently?' Debbie would no doubt see Joni in a new light at this point. Something that she had naturally assumed about Joni would have turned out not to be true. We could express what this thing is in a number of ways. We might say that Joni turns out not to have the normal range of motivations. We might, to adopt P.F. Strawson's terminology, say that she turns out not to be a member of the 'moral community'.[18] It is not clear, however, that Joni is mistaken in her assertion that, for her, not caring about the moral status of her actions, there is no reason not to treat Kim in a way that is morally wrong. Further than this, imagine Joni claims not to care about any of the facts on which the judgment of moral wrongness might

[18] Strawson (2008), p. 18.

naturally be thought to be based, for example the harm caused to Kim, or the kind of treatment that Kim has a right to expect from Joni. These things, if true of Joni, take her further outside the normal range of human personality, but it seems to me that they also undermine the intuition that she has a reason not to treat Kim the way she has. This, after all, is the question we have been exploring in this chapter; a question that would be begged by an argument which includes premise 2 above.

There is another argument which has been made in support of externalism which is worth exploring briefly here because it is closely related to our wider concerns. This is really a version of the argument explored above, but differs in that it makes use of the idea of blame; an idea which is of course a central concern in this thesis. The argument can be sketched as follows:[19]

1. Blame is inappropriate in the absence of a reason against the action for which the person is blamed.
2. Moral wrongness is sufficient to warrant blame.
3. Therefore (from 1 and 2) moral obligations must entail reasons.
4. Moral wrongness does not depend on the agent's desires or inclinations.
5. Therefore (from 3 and 4) there must be some reasons that do not depend on the agent's desires or inclinations: external reasons.

As we can see, steps 3 to 5 are essentially the same as the three steps from the first argument we encountered. Premises 1 and 2 are being pressed into service to bolster 3, the claim that a moral obligation to ϕ entails a reason in favour of ϕ-ing.

The second premise above states that moral wrongness is sufficient to warrant blame. There are, I think, good reasons to doubt the truth of this premise in general. It may be that the agent is non-culpably ignorant of the wrong-making

[19] Again, this formulation is from Finlay and Schroeder (2012).

features of the action in question, and thus locally unresponsive to the relevant reasons that speak against the action. To defend premise 2, then, would require asserting that in cases like this the agent does not in fact act in a way that is morally wrong (and that the features of the action which would ordinarily be wrong-making are not wrong-making in cases in which the agent is non-culpably ignorant of them). This is a respectable claim which could be defended. However, as with the first argument above, it might be difficult to defend when applied to the specific case of psychopaths. I have argued that psychopaths are unresponsive to reasons of this kind, but it is another thing entirely to say that they do not act immorally when they fail to act on these reasons. This is a counter-intuitive claim which would require considerable support.

Alternatively, it might be claimed that psychopaths *do* act immorally when they, for example, harm someone, and that this is enough to show that they are blameworthy. But this is also not a claim which is self-evidently true. As I argued at the beginning of the thesis, whether or not psychopaths are blameworthy is not a question whose answer can be assumed, or which can be answered by an appeal to intuition.

6.4 Becoming aware of reasons

The two arguments above are intended to establish the existence of external moral reasons. However, as long as there exist agents who are exceptions to the premises of the arguments, they will be unable to establish the existence of external moral reasons for action which apply to these exceptional agents. Unless we want to accept a mixed metaphysics of reasons, then, we must accept that they do not establish the existence of external reasons at all. I believe psychopaths are a type of agent who are exceptional in sufficiently relevant ways at least to cast doubt on the question of whether the premises of the arguments above hold true of them.

Nonetheless, some of the arguments' power is retained. None of the objections I have mentioned above is a knock-down argument against the existence of external reasons, or even an argument which purports to prove conclusively that internal reasons exist; they merely cast doubt over the universality of certain claims which are supposed to support the existence of external reasons. There is still considerable intuitive support for the truth of these claims in the majority of cases, and it is still an open question whether they hold true for psychopaths.

It may be possible, therefore, to find additional support for the externalist position which, combined with the arguments above, presents a compelling case for this position overall. One way of doing this would be to find some cases, or a particular type of case, for which internalism has trouble accounting, and which can be better explained by the existence of external reasons. I believe such a type of case does exist, and the rest of this chapter will be devoted to exploring this type of case and its implications for practical reasons.

The type of case that I have in mind is one in which an agent who previously had not perceived themselves to have a particular kind of practical reason – due to certain facts about them which on the internalist account would bar them from indeed having that kind of reason – later comes to see themselves as having had such a reason, and having been mistaken about their previous situation. In such cases I believe the best available explanation is that they were indeed mistaken, and that they did have the practical reason or reasons in question, a fact which is incompatible with the internalist position.

Recall that the central question which we are trying to address is this one: can someone have reasons based on the value of others despite not seeing those other people as having value? One tactic we could employ in trying to answer this question is to construct an analogous case in the hope that it provokes firmer intuitions than the psychopath case does. If the analogous case is dissimilar to the case of psychopaths only in ways that are irrelevant to the

question under discussion, then this would give us a reason to favour whatever conclusion in the psychopath case is compatible with our intuitions in the analogous case.

Ordinarily, people can be assumed to have reasons based on their *own* value – prudential reasons. But what if you were incapable of seeing yourself as having value? Would this put pressure on the idea that such reasons applied, in the same way that the psychopath case puts pressure on the idea that reasons based on the value of others apply? In fact, we can probably imagine someone in this predicament. Imagine you had a friend who had extremely low self-esteem, to the point where they could not see themselves or their lives as having value, and to the point at which they genuinely had no desire to, say, look after their own basic needs. In attempting to bring them round from this position, you might appeal to various reasons which, in your opinion but not in theirs, ought to have an influence on their choices. You might, for example, ask them, 'Can't you see that you have a reason to live?', or 'Can't you see that you have a reason to eat?', or just, 'Can't you see that you have a reason to look after yourself?' They might sincerely reply that no, in fact they cannot see that they have any of these reasons. At bottom, this attitude would stem from the fact that they see themselves as valueless. And it is easy to understand how someone in this predicament would not perceive a reason to look after themselves. The important question, however, is whether we should take this as evidence that no such reason exists.[20]

---

[20] There might be many reasons that apply in this case that do not depend on the person in question's directly perceiving themselves as having value, and these might include reasons that stem from the perceptions of others that the person in question has value, such as their friends and family for instance. You can imagine appealing to these reasons in remonstrating with your friend: 'Your friends and family love you and will be upset if you don't start looking after yourself'. Assuming your friend valued the feelings of her friends and family, these appeals to reasons might have some traction with her. But they are not relevant to this discussion, because they take the case

It is worth noting first of all that you, if you were any kind of friend, would certainly not *take* this as evidence that no such reason exists. You would insist that, regardless of whether your friend could see that they had a reason to look after themselves, they certainly *did* have a reason, and you would do everything in your power to try to *make* them see that reason. Of course, one reason why you would see the situation in this way is because you, unlike your friend, would perceive your friend as valuable. This would suggest that to *see* one's own interests as providing reasons for oneself requires seeing oneself as valuable. But this is not the same as saying that seeing oneself as valuable is a necessary condition of having such reasons.

If this were the case then we would need to interpret your attitude in the case as inaccurately representing the facts about your friend. Perhaps you would be mistaken in thinking that your friend had a reason to look after herself. But why then would you think she did have such a reason? Perhaps because you did not believe that she was sincere in her assertion that she saw herself as valueless. But this does not describe the case correctly. It is not that there is some argument that your friend could make to finally convince you that she sees herself as valueless, at which point you would be won over to the idea that she therefore had no reason to look after herself. Becoming convinced that your friend really saw herself as valueless would make no difference to your assertion that she had such a reason.

---

out of alignment with the psychopath case, in which we are not able to refer to things that the psychopath does find valuable in order to appeal to value indirectly in this way. The question in the psychopath case must be whether the rights and interests of others supply reasons for psychopaths directly, despite their not seeing others as valuable. The question in the 'friend with low self-esteem' case must be whether the rights and interests of the friend with low self-esteem supply reasons for her directly, despite her not seeing herself as valuable.

Alternatively, perhaps we should interpret your attitude as insincere. The thought might be that you see, really, that she has no reason to look after herself, but by trying to convince her that she has a reason, you are attempting to *bring it about* that she has a reason, by bringing it about that she joins you in seeing her as valuable. But again, this does not describe the case correctly. The success of such a strategy would depend on your ability to convince your friend that she had a real reason to look after herself. If you did not believe this yourself, it is hard to see how you could hope to be successful in convincing your friend. Your friend would also need at least to see reasons as available to people who do not have the relevant desires to base them on, or again, it is hard to see how this strategy could be expected to bear any fruit. This shows, at least, that we do not typically think of reasons as depending on desires, or at least that interactions such as the one under discussion make sense only on the assumption that they do not always so depend. It is perhaps possible that we are in error about this, but we would need an additional good reason to accept this interpretation.

It seems to me that the internalist view has difficulty making sense of a case like this, where a lack of the relevant desire seems to the agent herself to result in a lack of the relevant reason, but it does not seem to another party to do so. I think it also has difficulty making sense of similar cases where the difference of opinion is not between the agent and another party, but between the agent and herself at another point in time. Imagine your attempt to persuade your friend in the first example is successful, and after time she comes to develop some self-esteem. You might ask her, 'Can you now see that you had a reason to look after yourself all along?' It is easy to imagine her sincerely replying that yes, she can see that she had a reason of this kind, and that she is grateful to you for helping her to come to see this reason; to come to see the situation *aright*. How are we to account for this person's strong intuition that, despite not desiring to look

after herself, or having any other desires which would be fulfilled by looking after herself, she nonetheless had a reason to do so all along?

There is at least one possible explanation available, which might be thought to be compatible with a version of the internalist position, although it clearly departs from what Williams had in mind. This explanation is that the reason that your friend had at T1 (when she had low self-esteem) depends not on a desire that she has at that point, but on a desire that she later comes to have at T2 (after developing self-esteem). She comes to have a desire at T2, the satisfaction of which requires that she earlier acted in a certain way at T1, while still having low self-esteem. Even though she had no such desire at T1, the existence of the desire at T2 is enough to ground the reason at T1. Having come to have the relevant desire at T2, she can now recognise that she had a reason at T1.[21]

This explanation is intended to make sense of the fact that your friend in the example comes to see at T2 that she had a reason at T1, even though at T1 she could see no such reason, and of the intuition that what she is doing is coming to see the situation aright. She had the relevant reason at the time, though her low self-esteem made this reason obscure to her. The explanation modifies the original internalist project of making sense of reasons from a first-person, present-tense perspective. Instead, it allows the link between desires and reasons, while still remaining first-personal, to stretch across tenses. But in doing so, it introduces some metaphysical complications that are hard to explain away.

To see this, imagine you are remonstrating – at T1 – with your friend in the example. She sees herself as valueless, therefore has no desire to look after herself, and therefore recognises no reason to look after herself. You, trying to

[21] See Nagel (1970), especially Chapter VIII.

convince her of the existence of such a reason, might point to the possibility of her recovering from the low self-esteem which leads her to see herself as valueless. "You might have no desire to look after yourself now," you might say, "but if you recover, you'll want to be in good shape, and you'll be glad that I convinced you to look after yourself so that you could be in good shape in the future. You'll see that you had a reason to look after yourself all along."

The problem for the internalist is that there would presumably be no way in which either your friend or you could know that T2 would actually come about while having your discussion at T1. From that vantage point, given your limited knowledge of the future, you would have to talk about her recovery only as a possibility. You would therefore, on the internalist picture under discussion, be in a position to tell her, not that she *has* a reason to look after herself, but that she *might* have, depending on whether or not she is destined to recover from her low self-esteem. The thought would be that looking after herself is a kind of gamble. It would be rational for her to look after herself on the assumption that she would later develop self-esteem. As she cannot know whether she has a genuine reason or not, she might as well look after herself just in case she turns out to have a reason to do so. While this might be enough to convince her – after all, it at least offers *hope* – it is manifestly not how you would think about her situation, and not the kind of argument any reasonable person would offer. The point is that she *has* a reason, not that she may or may not have, depending on what happens in her future.

It seems to me that the internalist project of making sense of practical reasons from a first-person perspective runs into problems, because it cannot account either for the way we think about the reasons which apply either to others, or to ourselves at a different point in time. Our own desires – real or hypothetical – cannot be the only considerations which supply us with reasons. By contrast, the externalist can appeal to considerations other than desires as suppliers of reasons. The correct way to think about practical reasons, I would suggest, and

the way the examples above suggest that we actually do judge these matters, is as external in the sense that I have been discussing. That is, the reasons which apply to a person do not depend on the set of desires that the person has.

Seeing reasons from an external perspective allows us to make sense of the first type of case above, because it allows us to point to a plethora of considerations in support of our claims about the practical reasons a person has, and a desire belonging to that person does not need to be one of these considerations. When arguing with your friend in the example we were considering earlier, you might point to her value as a person as presenting a reason for her to look after herself. That she does not recognise that value, that she has no desire to look after herself nor any other desire which could be indirectly served by looking after herself, does not make any difference to the claim that she has a reason to look after herself. The natural way of describing this case – that she has a reason but lacks a desire and because of this does not recognise that reason – accords perfectly well with the externalist worldview.

Furthermore, the version of the case in which your friend later comes to recognise her own worth and the reasons that it formerly presented to her, can also be understood more naturally if we apply an externalist understanding of those reasons. Following her development of self-esteem, your friend comes to recognise the reasons that were there all along. This is the way the truth of her situation presents itself to her, and this, on the externalist reading, is how it really is. There is no need to introduce metaphysical complexity in the form of desires which may or may not come into being at a time later than one at which the original reason apparently already exists. Because of the simplicity and naturalness of the externalist reading of these cases, we have a powerful reason to prefer an externalist account of practical reasons.

The examples we have been considering have to do with the perception of *oneself* as possessing value. I have argued that these give us a reason to prefer

the externalist account overall. But in the context of the overall argument of the thesis, we are concerned rather with the perception of others as possessing value. I will close this chapter by showing that the same basic set of intuitions apply to cases which deal with this type of perception as apply to the former cases. When we consider cases in which someone comes to see the situation aright, we can see that an externalist reading is better able to make sense of this type of case too.

As well as full-blown psychopathy, there is a type of case much discussed in psychology and neuroscience, known as 'acquired sociopathy'.[22] In cases of this type, previously non-psychopathic people acquire psychopathic traits, usually after a brain injury. This process appears to happen in one direction only: non-psychopaths become psychopaths following injury. However, one could imagine, perhaps in a more medically advanced society of the future, it might become possible to reverse the brain injury, so that someone who for a time had been a psychopath might be reverted to a non-psychopathic state. Assuming they had caused harm to people while in their psychopathic state, how might someone in this predicament view their former actions? Probably with something on a spectrum from regret to horror, depending on how much harm they had caused. Certainly, as with the 'friend with low self-esteem' example, it seems almost inevitable that they would judge themselves as having acted contrary to practical reasons that properly applied to their actions. As with that type of case, they would see their former condition as having obscured a real set of reasons from their view, reasons which were nonetheless real and present.

Now, without running through the same arguments over again, it seems to me that any attempt to explain this type of case using an internalist account of practical reasons is likely to run into the same problems that we saw in our discussion of the earlier case. The internalist account requires the reformed

---

[22] See e.g. Blair and Cipolotti (2000).

psychopath to have had some desire to act in accordance with the belief that others have value while in their psychopathic state, but it is manifest in the case that no such desire exists. We are left with a strong intuition that there is a reason (or set of reasons) that needs to be explained, and no way to explain it on the internalist account. The externalist reading of the case, meanwhile, has no trouble explaining the existence of this set of reasons. We should therefore prefer the externalist account.

Conclusions

In conclusion, there are good reasons for preferring an externalist to an internalist account of practical reasons, as providing a plausible explanation of a range of cases including cases involving psychopaths. We are therefore entitled to conclude that psychopaths, despite not seeing others as valuable, and despite not having desires that are based on the value of others, do have reasons based on the value of others, including reasons stemming from the rights, interests and concerns of other people. While the arguments of this chapter do not constitute conclusive proof that such reasons exist, they have seen off the challenge from internalism, which was the strongest argument available against their existence.

Conclusions

I have argued that psychopaths are not responsive to certain reasons, and are therefore not morally responsible for failing to act on those reasons. Being responsive to the reasons in question, I have argued, depends on the ability to value others. More accurately, it depends on the ability to recognise something other than oneself as an ultimate source of value, since 1) psychopaths might be able to see others as valuable instrumentally, insofar as they can serve the psychopath's end, and 2) they are blind to the value not just of other people, but also of such things as animals, the environment or justice. I think psychopaths are unresponsive to these reasons because they have a general emotional deficiency which stunts their ability to engage evaluatively with the world, and a specific deficiency of empathy which prevents them from achieving an ability to value others. I think because these deficiencies are already well-established in childhood, and appear to be irreversible, psychopaths are not morally responsible for being in this state of unresponsiveness to the reasons in question – this is beyond their control.

The reasons to which I think psychopaths are unresponsive are therefore all those reasons that depend on the value of entities other than oneself. I have focused primarily on reasons stemming from the rights, interests and concerns of other people, for example reasons to refrain from harming people which are due to their having a right not to be harmed, or an interest in not being harmed. However, as I argued in Chapter 3, psychopaths are also unresponsive to any reasons which may stem from the rights, interests and concerns of animals, and from considerations such as fairness and justice, which also depend on the value of entities who must be treated fairly or justly.

In the first chapter, I argued that one can be morally responsible for morally good, bad and neutral acts. I think psychopaths are not morally responsible for morally bad acts, insofar as morally bad acts depend on the ability to value

others. I also think that psychopaths are not morally responsible for morally *good* acts, insofar as morally good acts depend on the ability to value entities other than oneself. Imagine a psychopath gives money to a homeless person in the street. Unless she has some ulterior motive for doing so, and given the premise that she does not think that person valuable (so for example she could just as happily kill the person if she wanted to) then she acts without reason, and does not understand the reasons which would ordinarily make this a morally good act. Therefore, the verdict that the psychopath is not morally responsible for the act (and does not deserve praise for it, for example) seems to me to be the right one. Morally neutral acts are unaffected by reasons based on the value of others, and therefore psychopaths (*qua* psychopaths) are morally responsible for these acts.

Acts are not the only things we can be morally responsible for. We can also be morally responsible for choices, for states of affairs which are the result of our acts or of our negligent inaction, for attitudes and for emotions. Insofar as reasons based on the value of others bear on these things, psychopaths are not morally responsible for them either.

As I noted in Chapter 2, psychopaths exist on a continuum, or rather on several continua. The features that make up a diagnosis of psychopathy according to the Hare checklist exist in a great many people to a greater or lesser extent. However, what the review of clinical and scientific literature in Chapter 2, together with the developmental picture set out in Chapter 5, hopefully make clear is that there is a group of people whose genetic inheritance, upbringing or both renders them truly incapable of seeing others as valuable. This set of people may only be a subset of people who would be assessed as psychopathic using Hare's scale, including only those who score at the high end for emotional deficiencies and lack of empathy specifically. It is these people to whom my verdict of moral non-responsibility applies. There are, of course, likely to be difficult borderline cases. In such cases, the criteria for ascribing moral

responsibility implied by my account would be the ability to value others. Secondary evidence for this ability, or its lack, could in principle be sought by considering the individual's neurological resources and the extent to which they had an upbringing characterised by the encouragement and exploitation of empathy by a caring parent. In practice, of course, making confident judgments on the basis of this evidence is likely to be a difficult endeavour. An opportunity for further research would be to consider in depth how we should think about borderline cases. Common sense would suggest that responsibility is not 'all or nothing', but admits of degrees (as in the parallel legal concept of 'diminished responsibility'). But what exactly is the relationship between a diminished capacity for empathy, a diminished capacity for valuing others, and diminished moral responsibility? How does each lead to the next? Finally, what should be our attitudes and practices towards someone who has diminished, but not absent, responsibility? These are interesting and non-trivial questions which would require further work.

It also remains to be stated what the implications of a verdict of non-responsibility ought to be for our practices and attitudes in respect of the clearer cases of psychopathy, and of the actions of these psychopaths. It is implied by the conclusions of Chapter 1 that lacking moral responsibility for an act (say) invalidates a whole range of practices and attitudes towards that act, including blame or praise, and the reactive emotions such as resentment. Answering the question of whether it invalidates punishment of the agent by the state would require an analysis of the purposes of punishment which is beyond the scope of this thesis. However, insofar as the purpose of punishment is to bring the perpetrator face-to-face with the significance of their actions, perhaps with the motive of encouraging them to better behaviour in future, I would suggest that punishment is misplaced in the case of psychopaths, who are pathologically impervious to the significance of their actions, and unlikely to change their behaviour in future. On the other hand, one purpose of punishment by

incarceration may simply be to incapacitate the perpetrator in order to protect others, and another purpose may be to discourage others from similar behaviour. Neither of these purposes is automatically excluded by my conclusions here.

Commentators[1] have sometimes worried that verdicts of non-responsibility may have the effect of excluding people from the moral community, and therefore of validating forms of treatment towards them which would ordinarily be considered unjust or illiberal. I have stated some actions and attitudes which I believe are inappropriate when directed at psychopaths, but which would be appropriate when directed at non-psychopaths. I have not said whether there are actions which would be justified when directed at psychopaths which would not be justified when directed at non-psychopaths, such as pre-emptive incarceration. Again, disentangling these issues would require significant additional argument, but suffice it to say that I do not think there is anything obvious in what I have said that should lead one to the conclusion that psychopaths should not be accorded something like the normal set of rights.

The waters around criminal justice are further muddied by the use of diagnostic categories different from, but supposedly related to, psychopathy. Significantly, my conclusions in this thesis say nothing about how we should treat people who have been diagnosed using the DSM-V classification of Antisocial Personality Disorder. Indeed, due to the issues with this diagnosis that I explored in Chapter 2, it is difficult to say anything very useful about how we should treat people in this category, who are unlikely to be a homogenous group at the level of personality. Given that APD is a very widely used diagnostic category, this obviously raises difficulties for anyone who would want to draw firm

[1] E.g. Benn (1999).

conclusions about how such people should be treated as a matter of judicial policy.

In practical terms, the conclusions I have presented in this thesis tell us something about how we should think about psychopaths, and point the way towards how we should interact with them, both as individuals and from a societal standpoint. In theoretical terms, they add to our understanding of what moral responsibility, understood on the reasons-responsiveness model, requires. It turns out that psychopaths, though not irrational in the sense that their condition does not render them factually mistaken about anything, are nonetheless unable to grasp a significant set of reasons which are available to non-psychopaths, and indeed which do apply to psychopaths as well. Moral responsibility, it turns out, requires not just the ability to grasp and apply moral concepts, but also the ability to value others.

Bibliography

Adolphs, R. 2010. What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences.* **1191**(1), pp.42-61.

Adshead, G. 1999. Psychopaths and other-regarding beliefs. *Philosophy, Psychiatry and Psychology.* **6**(1), pp.41-4.

Alvarez, M. 2009. How many kinds of reasons? *Philosophical Explorations.* **12**, pp.181-93.

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5.* Washington, D.C.

Appelbaum, P.S. 1999. Ought we to require emotional capacity as part of decisional competence? *Kennedy Institute of Ethics Journal.* **8**(4), pp.377-87.

Aristotle. 1985. *Nicomachean ethics.* Indianapolis, Indiana: Hackett.

Arrington, R.L. 1979. Practical reason, responsibility and the psychopath. *Journal for the Theory of Social Behaviour.* **9**(1), pp.71-89.

Austin, J.L. 1956. A plea for excuses. *Proceedings of the Aristotelian Society.* **57**, pp.1-30.

Ayer, A.J. 1980. Free-will and rationality. In: Van Straaten, Z. ed. *Philosophical subjects: Essays presented to P.F. Strawson.* Oxford: Clarendon Press, pp.1-13.

Bandura, A. and Walters, R.H. 1959. *Adolescent aggression.* New York: Ronald Press.

Barry, P.B. 2010. Saving Strawson: Evil and Strawsonian accounts of moral responsibility. *Ethical Theory and Moral Practice.* **14**(1), pp.5-21.

Bartlett, P. 2010. Stabbing in the dark: English law relating to psychopathy. In: Malatesti, L. and McMillan, J. eds. *Responsibility and psychopathy: Interfacing law, psychiatry and philosophy.* Oxford: Oxford University Press.

Ben-Ze'Ev, A. 2004. Emotions are not mere judgments. *Philosophy and Phenomenological Research.* **68**(2), pp.450-7.

Benn, P. 1999. Freedom, resentment and the psychopath. *Philosophy, Psychiatry and Psychology.* **6**(1), pp.29-39.

Blair, J., Mitchell, D. and Blair, K. 2005. *The psychopath: Emotion and the brain.* Oxford: Blackwell.

Blair, R.J.R, Jones, L., Clark, F. and Smith, M. 1997. The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology.* **34**(2), pp.192-8.

Blair, R.J.R. 1995. A cognitive developmental approach to morality: Investigating the psychopath. *Cognition.* **57**, pp.1-29.

Blair, R.J.R. 1997. Moral reasoning and the child with psychopathic tendencies. *Personality and Individual Differences.* **22**(5), pp.731-9.

Blair, R.J.R. 2001. Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *Journal of Neurology, Neurosurgery & Psychiatry.* pp.727-31.

Blair, R.J.R. 2003. Neurobiological basis of psychopathy. *The British Journal of Psychiatry.* **182**, pp.5-7.

Blair, R.J.R. 2008. The cognitive neuroscience of psychopathy and implications for judgments of responsibility. *Neuroethics.* **1**(3), pp.149-57.

Blair, R.J.R. 2010. Neuroimaging of psychopathy and antisocial behavior: A targeted review. *Current Psychiatry Reports.* **12**(1), pp.76-82.

Blair, R.J.R. 2011. Moral judgment and psychopathy. *Emotion Review.* **3**(3), pp.296-8.

Blair, R.J.R. and Cipolotti, L. 2000. Impaired social response reversal: A case of acquired sociopathy. *Brain.* pp.1122-41.

Blair, R.J.R., Colledge, E. and Mitchell, D.G.V. 2001a. Somatic markers and response reversal: Is there orbitofrontal cortex dysfunction in boys with psychopathic tendencies? *Journal of Abnormal Child Psychology.* **29**(6), pp.499-511.

Blair, R.J.R., Colledge, E., Murray, L. and Mitchell, D.G. 2001b. A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal Child Psychology.* **29**(6), pp.491-8.

Blair, R.J.R., Jones, L., Clark, F. and Smith, M. 1995. Is the psychopath 'morally insane'? *Personality and Individual Differences.* **19**(5), pp.741-52.

Blair, R.J.R., Mitchell, D.G.V., Peschardt, K.S., Colledge, E., Leonard, R.A., Shine, J.H., Murray, L.K. and Perrett, D.I. 2004. Reduced sensitivity to others' fearful expressions in psychopathic individuals. *Personality and Individual Differences.* **37**, pp.1111-22.

Blair, R.J.R., Mitchell, D.G.V., Richell, R.A., Kelly, S., Leonard, A. and Newman, C. 2002. Turning a deaf ear to fear: Impaired recognition ofvocal affect in psychopathic individuals. *Journal of Abnormal Psychology.* **111**, pp.682-6.

Bloom, P. 2016. *Against empathy: The case for rational compassion.* New York: Ecco Press.

Bok, H. 2002. Wallace's 'normative approach' to moral responsibility. *Philosophy and Phenomenological Research.* **64**(3), pp.682-6.

Botterell, A. 2009. A primer on the distinction between justification and excuse. *Philosophy Compass.* **4**(1), pp.172-96.

Brandt, R.B. 1969. A utilitarian theory of excuses. *The Philosophical Review.* **78**(3), pp.337-61.

Brody, G.H. and Shaffer, D.R. 1982. Contributions of parents and peers to children's moral socialization. *Development Review.* **2**, pp.31-75.

Broome, J. 2007. Wide or narrow scope? *Mind.* **116**, pp.359-70.

Calder, A.J., Young, A.W., Rowland, D., Perrett, D.I., Hodges, J.R. and Etcoff, N.L. 1996. Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology.* **13**(5), pp.699-745.

Calhoun, C. 1989. Responsibility and reproach. *Ethics.* **99**(2), pp.389-406.

Ciaramelli, E. and di Pellegrino, G. 2011. Ventromedial prefrontal cortex and the future of morality. *Emotion Review.* **3**(3), pp.308-9.

Ciocchetti, C. 2003. The responsibility of the psychopathic offender. *Philosophy, Psychiatry and Psychology* **10**(2), pp.175-83.

Cleckley, H.M. 1941. *The mask of sanity: An attempt to clarify some issues about the so-called psychopathic personality.* St Louis: Mosby.

Cooke, D.J. and Michie, C. 2001. Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment.* **13**(2), pp.171-88.

Coplan, A. 2011. Understanding empathy: Its features and effects. *Empathy: Philosophical and psychological perspectives.* Oxford: Oxford University Press, pp.3-18.

Coplan, A. and Goldie, P. eds. 2011. *Empathy: Philosophical and psychological perspectives.* Oxford: Oxford University Press.

Crockenberg, S.B. and Litman, C. 1990. Autonomy as competence in 2-year-olds: Maternal correlates of defiance, compliance and self-assertion. *Developmental Psychology.* **26**, pp.961-71.

Damasio, A. 2006. *Descartes' error: Emotion, reason and the human brain.* London: Vintage.

Davies, S. 2011. Infectious music: Music-listener emotional contagion. In: Coplan, A. and Goldie, P. eds. *Empathy: Philosophical and psychological perspectives.* Oxford: Oxford University Press.

Davis, M.H. 1996. *Empathy: A social psychological approach.* Boulder, CO: Westview Press.

de Oliveira-Souza, R., Hare, R.D., Bramati, I.E., Garrido, G.J., Azevedo, I.F., Tovar-Moll, F. and Moll, J. 2008. Psychopathy as a disorder of the moral brain: Fronto-temporo-limbic grey matter reductions demonstrated by voxel-based morphometry. *Neuroimage.* **40**(3), pp.1202-13.

De Sousa, R. 1987. *The rationality of emotion.* Cambridge: MIT Press.

De Sousa, R. 2002. Emotional truth. *Proceedings of the Aristotelian Society.* **supp. vol. 76**, pp.247-63.

De Veer, A. 1991. *Parental disciplinary strategies and the child's moral internalization*. Unpublished doctoral dissertation, University of Nijmegen.

Debes, R. 2009. Which empathy? Limitations in the mirrored "understanding" of emotion. *Synthese.* **175**(2), pp.219-39.

Decety, J., Chen, C., Harenski, C. and Kiehl, K.A. 2013a. An fmri study of affective perspective taking in individuals with psychopathy: Imagining another in pain does not evoke empathy. *Frontiers in Human Neuroscience.* **7**, p489.

Decety, J., Skelly, L., Yoder, K.J. and Kiehl, K.A. 2014. Neural processing of dynamic emotional facial expressions in psychopaths. *Social Neuroscience.* **9**(1), pp.36-49.

Decety, J., Skelly, L.R. and Kiehl, K.A. 2013b. Brain response to empathy-eliciting scenarios involving pain in incarcerated individuals with psychopathy. *JAMA Psychiatry.* **70**(6), pp.638-45.

Deeley, Q., Daly, E., Surguladze, S., Tunstall, N., Mezey, G., Beer, D., Ambikapathy, A., Robertson, D., Giampietro, V., Brammer, M.J., Clarke, A., Dowsett, J., Fahy, T., Phillips, M.L. and Murphy, D.G. 2006. Facial emotion processing in criminal psychopathy. Preliminary functional magnetic resonance imaging study. *British Journal of Psychiatry.* **189**, pp.533-9.

Deigh, J. 1995. Empathy and universalizability. *Ethics.* **105**(4), pp.743-63.

Deonna, J.A. 2006. Emotion, perception and perspective. *Dialectica.* **60**(1), pp.29-46.

Ditto, P.H. and Koleva, S.P. 2011. Moral empathy gaps and the American culture war. *Emotion Review.* **3**(3), pp.331-2.

Döring, S.A. 2004. *Gründe und gefühle. Rationale Motivation durch emotionale Vernunft.* Habilitationschrift, Universität Essen-Duisburg.

Duff, A. 1977. Psychopathy and moral understanding. *American Philosophical Quarterly.* **14**(3), pp.189-200.

Dutton, K. 2012. *The wisdom of psychopaths: Lessons in life from saints, spies and serial killers.* London: Random House.

Eisenberg, N. and Strayer, J. 1987. *Empathy and its development.* Cambridge: Cambridge University Press.

Ekman, P. and Friesen, W.V. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychiatry.* **17**(2), pp.124-9.
Elliott, C. 1992. Diagnosing blame: Responsibility and the psychopath. *Journal of Medicine and Philosophy.* **17**(2), pp.199-214.
Elliott, C. 1994. Puppetmasters and personality disorders: Wittgenstein, mechanism and moral responsibility. *Philosophy, Psychiatry and Psychology.* **1**(2), pp.91-100.
Elliott, C. 1996. *The rules of insanity: Moral responsibility and the mentally ill offender.* Albany: State University of New York Press.
Ellis, H. 1890. *The criminal.* London: Walter Scott.
Eshleman, A. 2009. *Moral responsibility.* [Online]. Available from: http://plato.stanford.edu/entries/moral-responsibility/
Eslinger, P.J. and Damasio, A.R. 1985. Severe disturbance of higher cognition after bilateral frontal lobe ablation: Patient evr. *Neurology.* **35**(12), pp.1731-41.
Farrington, D.P. 2007. Family background and psychopathy. In: Patrick, C.J. ed. *Handbook of psychopathy.* New York: The Guilford Press, pp.229-50.
Farrington, D.P.Ullrich, S.Salekin, R.T. and Lynam, D.R. 2010. Environmental influences on child and adolescent psychopathy. In: Salekin, R.T. and Lynam, D.R. eds. *Handbook of child and adolescent psychopathy.* New York: The Guilford Press, pp.202-32.
Fawcett, C., Wesevich, V. and Gredeback, G. 2016. Pupillary contagion in infancy: Evidence for spontaneous transfer of arousal. *Psychological Science.* **27**(7), pp.997-1003.
Feinberg, J. 1986. *Harm to self: The moral limits of the criminal law.* New York: Oxford University Press.
Field, T.M., Woodson, R., Greenberg, R. and Cohen, D. 1982. Discrimination and imitation of facial expressions by neonates. *Science.* **218**, pp.179-81.
Fine, C. and Kennett, J. 2004. Mental impairment, moral understanding and criminal responsibility: Psychopathy and the purposes of punishment. *International Journal of Law and Psychiatry.* **27**(5), pp.425-43.
Finlay, S. and Schroeder, M. 2012. Reasons for action: Internal vs. external. *Stanford Encyclopedia of Philosophy.* [Online]. Available from: http://plato.stanford.edu/entries/reasons-internal-external/
Fischer, J.M. 1999. Recent work on moral responsibility. *Ethics.* **110**(1), pp.93-139.
Fischer, J.M. and Ravizza, M. 1998. *Responsibility and control: A theory of moral responsibility.* Cambridge: Cambridge University Press.
Flor, H., Birbaumer, N., Hermann, C., Ziegler, S. and Patrick, C.J. 2002. Aversive pavlovian conditioning in psychopaths: Peripheral and central correlates. *Psychophysiology.* **39**(4), pp.505-18.
Gabbard, G.O. 2005. Mind, brain and personality disorders. *American Journal of Psychiatry.* **162**, pp.648-55.
Gao, Y., Glenn, A.L., Schug, R.A., Yang, Y. and Raine, A. 2009. The neurobiology of psychopathy: A neurodevelopmental perspective. *Canadian Journal of Psychiatry.* **54**(12), pp.813-23.
Gao, Y. and Raine, A. 2010. Successful and unsuccessful psychopaths: A neurobiological model. *Behavioral Sciences and the Law.* **28**(2), pp.194-210.
Gardner, J. 2007. In defence of defences. *Offences and defences: Selected essays in the philosophy of criminal law.* Oxford: Oxford University Press, pp.77-89.

Gibbons, J. 2010. Things that make things reasonable. *Philosophy and Phenomenological Research.* **81**, pp.335-61.
Gillberg, C. 2007. Non-autism childhood empathy disorders. In: Farrow, T. and Woodruff, P. eds. *Empathy in mental illness.* Cambridge: Cambridge University Press, pp.111-25.
Glannon, W. 1997. Psychopathy and responsibility. *Journal of Applied Philosophy.* **14**(3), pp.263-75.
Glannon, W. 2008. Moral responsibility and the psychopath. *Neuroethics.* **1**(3), pp.158-66.
Glenn, A.L., Raine, A. and Laufer, W.S. 2011. Is it wrong to criminalize and punish psychopaths? *Emotion Review.* **3**(3), pp.302-4.
Glenn, A.L., Raine, A. and Schug, R.A. 2009. The neural correlates of moral decision-making in psychopathy. *Molecular Psychiatry.* **14**(1), pp.5-6.
Glover, J. 1970. *Responsibility.* London: Routledge and Keegan Paul.
Glover, J. 2008. *LSE meeting with Alan Ryan, Part II. "Antisocial personality disorder" interviews in Broadmoor.* Available from: http://www.jonathanglover.co.uk/philosophy-beliefs-and-conflicts/lse-meeting-with-alan-ryan
Goldie, P. 2004. Emotion, feeling, and knowledge of the world. In: Solomon, R.C. ed. *Thinking about feeling: Contemporary philosophers on emotions.* Oxford: Oxford University Press, pp.91-106.
Goldie, P. 2005. Imagination and the distorting power of emotion. *Journal of Consciousness Studies.* (8), pp.127-39.
Goldie, P. 2007a. Emotion. *Philosophy Compass.* **2**(6), pp.928-38.
Goldie, P. 2007b. Seeing what is the kind thing to do: Perception and emotion in morality. *Dialectica.* **61**(3), pp.347-61.
Goldie, P. 2011. Anti-empathy. In: Coplan, A. and Goldie, P. eds. *Empathy: Philosophical and psychological perspectives.* Oxford: Oxford University Press, pp.304-31.
Gordon, R.M. 1987. *The structure of emotions: Investigations in cognitive philosophy.* Cambridge: Cambridge University Press.
Greenspan, P.S. 2003. Responsible psychopaths. *Philosophical Psychology.* **16**(3), pp.417-29.
Griffiths, P.E. 2004. Is emotion a natural kind? In: Solomon, R.C. ed. *Thinking about feeling.* Oxford: Oxford University Press, pp.233-49.
Habel, U., Kuhn, E., Salloum, J.B., Devos, H. and Schneider, F. 2002. Emotional processing in psychopathic personality. *Aggressive Behavior.* **28**(5), pp.394-400.
Haji, I. 1998. On psychopaths and culpability. *Law and Philosophy.* **17**(2), pp.117-40.
Haji, I. 2003. The emotional depravity of psychopaths and culpability. *Legal Theory.* **9**, pp.63-82.
Haji, I. 2005. Introduction: Semi-compatibilism, reasons-responsiveness, and ownership. *Philosophical Explorations.* **8**(2), pp.91-3.
Haksar, V. 1964. Aristotle and the punishment of psychopaths. *Philosophy.* **39**(150), pp.323-40.
Haksar, V. 1965. The responsibility of psychopaths. *The Philosophical Quarterly.* **15**(59), pp.135-45.

Hamilton, G. 2008. Mythos and mental illness: Psychopathy, fantasy, and contemporary moral life. *The Journal of Medical Humanities.* **29**(4), pp.231-42.
Hare, R.D. 1965. Psychopathy, fear arousal and anticipated pain. *Psychological Reports.* **16**(16), pp.499-502.
Hare, R.D. 1970. *Psychopathy: Theory and research.* New York: Wiley.
Hare, R.D. 1980. A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences.* **1**(2), pp.111-9.
Hare, R.D. 1982. Psychopathy and physiological-activity during anticipation of an aversive stimulus in a distraction paradigm. *Psychophysiology.* **19**(3), pp.266-71.
Hare, R.D. 1991. *The psychopathy checklist-revised.* Toronto: Multi-Health Systems.
Hare, R.D. 1995. *Without conscience: The disturbing world of the psychopaths among us.* New York: The Guilford Press.
Hare, R.D. 1998. The Hare PCL-R: Some issues concerning its use and misuse. *Legal and Criminological Psychology.* **3**, pp.99-119.
Hare, R.D., Frazelle, J. and Cox, D.N. 1978. Psychopathy and physiological responses to threat of an aversive stimulus. *Psychophysiology.* **15**(2), pp.165-72.
Hare, R.D. and Neumann, C.S. 2008. Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology.* **4**, pp.217-46.
Hare, R.D. and Neumann, C.S. 2010. Psychopathy: Assessment and forensic implications. In: Malatesti, L. and McMillan, J. eds. *Responsibility and psychopathy: Interfacing law, psychiatry and philosophy.* Oxford: Oxford University Press, pp.121-63.
Harenski, C.L. and Kiehl, K.A. 2011. Emotion and morality in psychopathy and paraphilias. *Emotion review.* **3**(3), pp.299-303.
Harlow, J.M. 1868. Recovery from the passage of an iron bar through the head. *Publications of the Massachusetts Medical Society.* **2**, pp.327-47.
Harold, J. and Elliott, C. 1999. Travelers, mercenaries and psychopaths. *Philosophy, Psychiatry and Psychology.* **6**(1), pp.45-8.
Harpur, T.J., Hakstian, A.R. and Hare, R.D. 1988. Factor structure of the psychopathy checklist. *Journal of consulting and Clinical Psychology.* **56**(5), pp.741-7.
Harris, G.T. and Rice, M.E. 2007. Treatment of psychopathy: A review of empirical findings. In: Patrick, C.J. ed. *Handbook of psychopathy.* New York: The Guilford Press.
Hart, C.H., DeWolfe, D.M., Wozniak, P. and Burts, D.C. 1992. Maternal and paternal disciplinary styles: Relations with preschoolers' playground behavioural orientations and peer status. *Child Development.* **63**, pp.879-92.
Hart, H.L.A. 1968. IX. Postscript: Responsibility and retribution. *Punishment and responsibility.* Oxford: Clarendon Press, pp.210–37.
Hatfield, E., Cacioppo, J.T. and Rapson, R.L. 1992. Primitive emotional contagion. In: Clarke, M. ed. *Review of personality and social psychology: Emotion and social behaviour.* Thousand Oaks, CA: Sage, pp.151-77.
Hatfield, E., Cacioppo, J.T. and Rapson, R.L. 1994. *Emotional contagion.* Cambridge: Cambridge University Press.
Haviland, J.M. and Lelwica, M. 1987. The induced affect response: 10-week-old infants' responses to three emotion expressions. *Developmental Psychology.* **23**, pp.97-104.

Hemphill, J.F., Hare, R.D. and Wong, S. 1998. Psychopathy and recidivism a review. *Legal and Criminal Psychology.* **3**, pp.737-45.
Hobson, J. and Shines, J. 1998. Measurement of psychopathy in a UK prison population referred for long-term psychotherapy. *British Journal of Criminology.* **38**(3), pp.504-15.
Hoffman, M.L. 1960. Power assertion by the parent and its impact on the child. *Child Development.* **31**, pp.129-43.
Hoffman, M.L. 2000. *Empathy and moral development: Implications for caring and justice.* New York: Cambridge University Press.
Howe, D. 2013. *Empathy: What it is and why it matters.* Basingstoke: Palgrave Macmillan.
Howe, M.L. and Courage, M.L. 1997. The emergence and early development of autobiographical memory. *Psychological Review.* **104**(3), pp.499-523.
Ishikawa, S.S., Raine, A., Lencz, T., Bihrle, S. and Lacasse, L. 2001. Autonomic stress reactivity and executive functions in successful and unsuccessful criminal psychopaths from the community. *Journal of Abnormal Psychology.* **110**(3), pp.423-32.
Jaffee, S.R., Caspi, A., Moffitt, T.E. and Taylor, A. 2004. Physical maltreatment victim to antisocial child: Evidence of an environmentally mediated process. *Journal of Abnormal Psychology.* **113**(1), pp.44-55.
James, W. 1884. What is an emotion? *Mind.* **9**(34), pp.188-205.
Judisch, N. 2005. Responsibility, manipulation and ownership. *Philosophical Explorations.* **8**(2), pp.115-30.
Kane, R. 2002a. Free will: Reflections on Wallace's theory. *Philosophy and Phenomenological Research.* **64**(3), pp.693-8.
Kane, R. 2002b. Responsibility, reactive attitudes and free will: Reflections on Wallace's theory. *Philosophy and Phenomenological Research.* **64**(3), pp.693-8.
Kant, I. 2005. *Groundwork for the metaphysic of morals.* Published online at http://www.earlymoderntexts.com/pdf/kantgrou.pdf.
Kennett, J. 2006. Do psychopaths really threaten moral rationalism? *Philosophical Explorations.* **9**(1), pp.69-82.
Kiehl, K.A., Smith, A.M., Hare, R.D., Mendrek, A., Forster, B.B., Brink, J. and Liddle, P.F. 2001. Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biological Psychiatry.* **50**(9), pp.677-84.
Kirsch, L.G. and Becker, J.V. 2007. Emotional deficits in psychopathy and sexual sadism: Implications for violent and sadistic behavior. *Clinical Psychology Review.* **27**(8), pp.904-22.
Knausgaard, K.O. 2013. *My struggle book 2: A man in love.* London: Random House.
Kohlberg, L. 1981. *Essays on moral development, vol. I: The philosophy of moral development.* San Francisco, CA: Harper & Row.
Kolodny, N. 2005. Why be rational? *Mind.* **114**(455), pp.509-63.
Korsgaard, C.M. 1986. Skepticism about practical reason. *The Journal of Philosophy.* **83**(1), pp.5-25.
Kosson, D.S., Suchy, Y., Mayer, A.R. and Libby, J. 2002. Facial affect recognition in criminal psychopaths. *Emotion.* **2**(4), pp.398-411.

Krevans, J. and Gibbs, J.C. 1996. Parents' use of inductive discipline: Relations to children's empathy and prosocial behavior. *Child development.* **67**(6), pp.3263-77.

Kringelbach, M.L. and Rolls, E.T. 2004. The functional neuroanatomy of the human orbitofrontal cortex: Evidence from neuroimaging and neuropsychology. *Progress in Neurobiology.* **72**(5), pp.341-72.

Kroner, D.G., Forth, A.E. and Mills, J.F. 2005. Endorsement and processing of negative affect among violent psychopathic offenders. *Personality and Individual Differences.* **38**(2), pp.413-23.

Kuczynski, L. 1983. Reasoning, prohibitions and motivations for compliance. *Developmental Psychology.* **19**, pp.126-34.

Kutz, C. 2004. Chapter 14: Responsibility. In: Coleman, J. and Shapiro, S. eds. *Jurisprudence and philosophy of law.* Oxford: Oxford University Press, pp.548–87.

Leichsenring, F., Leibing, E., Kruse, J., New, A.S. and Leweke, F. 2011. Borderline personality disorder. *The Lancet.* **377**(9759), pp.74-84.

Levenson, R.W., Ekman, P. and Friesen, W.V. 1990. Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology.*

Levenston, G.K. and Patrick, C.J. 2000. The psychopath as observer: Emotion and attention in picture processing. *Journal of abnormal psychology.* **109**(3), pp.373-85.

Levy, N. 2008. The responsibility of the psychopath revisited. *Philosophy, Psychology and Psychiatry.* **14**(2), pp.129-38.

Levy, N. 2011. Expressing who we are: Moral responsibility and awareness of our reasons for action. *Analytic Philosophy.* **52**(4), pp.243-61.

Levy, N. 2014. Psychopaths and blame: The argument from content. *Philosophical Psychology.* **27**(3), pp.351-68.

Lieb, K., Zanarini, M.C., Schmahl, C., Linehan, M.M. and Bohus, M. 2004. Borderline personality disorder. *The Lancet.* **364**(9432), pp.453-61.

Lykken, D.T. 1957. A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology.* **55**(1), pp.6-10.

Maibom, H.L. 2005. Moral unreason: The case of psychopathy. *Mind and Language.* **20**(2), pp.237-57.

Maibom, H.L. 2008. The mad, the bad, and the psychopath. *Neuroethics.* **1**(3), pp.167-84.

Malatesti, L. and McMillan, J. eds. 2010. *Responsibility and psychopathy: Interfacing law, psychiatry and philosophy.* Oxford: Oxford University Press.

Mason, M. 2011. Blame: Taking it seriously. *Philosophy and Phenomenological Research.* **LXXXIII**(2), pp.473-81.

Maudsley, H. 1873. *Body and mind.* London: Macmillan and Co.

Maudsley, H. 1874. *Responsibility in mental disease.* London: H.S. King.

Maxwell, B. and Sage, L.L. 2009. Are psychopaths morally sensitive? *Journal of Moral Education.*

McKenna, M.S. 1998. The limits of evil and the role of moral address: A defense of Strawsonian compatibilism. *The Journal of Ethics.* **2**(2), pp.123-42.

McMillan, J.R. 2003. Dangerousness, mental disorder, and responsibility. *Journal of Medical Ethics.* **29**(4), pp.232-5.

Meffert, H., Gazzola, V., den Boer, J.A., Bartels, A.A. and Keysers, C. 2013. Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. *Brain.* **136**(8), pp.2550-62.

Mikhail, J. 2011. Emotion, neuroscience, and law: A comment on Darwin and Greene. *Emotion Review.* **3**(3), pp.293-5.

Millgram, E. 1996. Williams' argument against external reasons. *Nous.* **30**(2), pp.197-220.

Millon, T., Simonsen, E., Birket-Smith, M. and Davis, R.D. eds. 1998. *Psychopathy: Antisocial, criminal and violent behaviour.* New York: The Guilford Press.

Minzenberg, M.J. and Siever, L.J. 2006. Neurochemistry and pharmacology of psychopathy and related disorders. In: Patrick, C.J. ed. *Handbook of psychopathy.* New York: Guilford, pp.251-77.

Mitchell, D.G.V., Colledge, E., Leonard, A. and Blair, R.J.R. 2002. Risky decisions and response reversal: Is there evidence of orbitofrontal cortex dysfunction in psychopathic individuals? *Neuropsychologia.* **40**, pp.2013-22.

Montmarquet, J.A. 2002. Wallace's 'Kantian' Strawsonianism. *Philosophy and Phenomenological Research.* **64**(3), pp.687-92.

Morse, S.J. 2008. Psychopathy and criminal responsibility. *Neuroethics.* **1**(3), pp.205-12.

Motzkin, J.C., Newman, J.P., Kiehl, K.A. and Koenigs, M. 2011. Reduced prefrontal connectivity in psychopathy. *Journal of Neuroscience.* **31**(48), pp.17348-57.

Muller, J.L., Sommer, M., Dohnel, K., Weber, T., Schmidt-Wilcke, T. and Hajak, G. 2008. Disturbed prefrontal and temporal brain function during emotion and cognition interaction in criminal psychopathy. *Behavioral Sciences and the Law.* **26**(1), pp.131-50.

Muller, J.L., Sommer, M., Wagner, V., Lange, K., Taschler, H., Roder, C.H., Schuierer, G., Klein, H.E. and Hajak, G. 2003. Abnormalities in emotion processing within cortical and subcortical regions in criminal psychopaths: Evidence from a functional magnetic resonance imaging study using pictures with emotional content. *Biological Psychiatry.* **54**(2), pp.152-62.

Murphy, J.G. 1972. Moral death: A Kantian essay on psychopathy. *Ethics.* **82**(4), pp.284-98.

Nagel, T. 1970. *The possibility of altruism.* Princeton, NJ: Princeton University Press.

Nagel, T. 1986. *The view from nowhere.* Oxford: Oxford University Press.

Newman, J.P., Patterson, C.M. and Kosson, D.S. 1987. Response perseveration in psychopaths. *Journal of Abnormal Psychology.* **96**(2), pp.145-8.

Nichols, S. 2002. How psychopaths threaten moral rationalism: Is it irrational to be amoral? *Monist.* **85**(2), pp.285-303.

Nussbaum, M. 2001. *Upheavals of thought.* Cambridge: Cambridge University Press.

Nussbaum, M. 2004. Emotions as judgments of value and importance. *Thinking about feeling: Contemporary philosophers on emotions.* Oxford: Oxford University Press, pp.183-99.

Ogloff, J.R.P. and Wong, S. 1990. Electrodermal and cardiovascular evidence of a coping response in psychopaths. *Criminal Justice and Behavior.* **17**(2), pp.231-45.

Oksenberg Rorty, A. 2004. Enough already with 'theories of the emotions'. In: Solomon, R.C. ed. *Thinking about feeling.* Oxford: Oxford University Press, pp.269-78.

Oshana, M.A.L. 1997. Ascriptions of responsibility. *American Philosophical Quarterly.* **34**(1), pp.71-83.
Oshana, M.A.L. 2004. Moral accountability. *Philosophical Topics.* **32**(1), pp.255-74.
Owens, D. 2008. Rationalism about obligations. *European Journal of Philosophy.* **16**(3), pp.403-31.
Parfit, D. 1984. *Reasons and persons.* Oxford: Clarendon Press.
Parfit, D. and Broome, J. 1997. Reasons and motivation. *Proceedings of the Aristotelian Society.* **Supp. Vol. 71** (May), pp.99-146.
Patrick, C.J. ed. 2006. *Handbook of psychopathy.* New York: The Guilford Press.
Patrick, C.J., Bradley, M.M. and Lang, P.J. 1993. Emotion in the criminal psychopath: Startle reflex modulation. *Journal of Abnormal Psychology.* **102**(1), pp.82-92.
Patrick, C.J., Cuthbert, B.N. and Lang, P.J. 1994. Emotion in the criminal psychopath: Fear image processing. *Journal of Abnormal Psychology.* **103**(3), pp.523-34.
Pickard, H. 2012. *Philosophy Bites: Hannah Pickard on responsibility and personality disorder*. Warburton, N. July 7. Available from: http://philosophybites.com/2012/07/hanna-pickard-on-responsibility-and-personality-disorder-originally-on-bioethics-bites.html
Preston, S.D. and de Waal, F.B.M. 2002. Empathy: Its ultimate and proximate bases. *Behavioural and Brain Sciences.* **25**, pp.1-72.
Prinz, J.J. 2004a. Embodied emotions. In: Solomon, R.C. ed. *Thinking about feeling: Contemporary philosophers on emotions.* Oxford: Oxford University Press, pp.44-60.
Prinz, J.J. 2004b. *Gut reactions: A perceptual theory of emotion.* Oxford: Oxford University Press.
Prinz, J.J. 2011. Is empathy necessary for morality? In: Coplan, A. and Goldie, P. eds. *Empathy: Philosophical and psychological perspectives.* Oxford: Oxford University Press, pp.211-29.
Pritchard, J.C. 1835. *A treatise on insanity.* London: Sherwood, Gilbert and Piper.
Pritchard, M.S. 1974. Responsibility, understanding, and psychopathology. *The Monist.*
Pujol, J., Batalla, I., Contreras-Rodriguez, O., Harrison, B.J., Pera, V., Hernandez-Ribas, R., Real, E., Bosa, L., Soriano-Mas, C., Deus, J., Lopez-Sola, M., Pifarre, J., Menchon, J.M. and Cardoner, N. 2012. Breakdown in the brain network subserving moral judgment in criminal psychopathy. *Social Cognitive and Affective Neuroscience.* **7**(8), pp.917-23.
Robinson, J. 2004. Emotion: Biological fact or social construction? In: Solomon, R. ed. *Thinking about feeling.* Oxford: Oxford University Press, pp.28-43.
Robinson, P. 1996. Competing theories of justification: Deeds v. reasons. In: Simester, A.P. and Smith, A.T.H. eds. *Harm and culpability.* Oxford: Clarendon Press, pp.45-70.
Rollins, B.C. and Thomas, D.L. 1979. Parental support, power and control techniques in the socialization of children. In: Burr, W.R., et al. eds. *Contemporary theories about the family: Vol. 1, research-based theories.* New York: Free Press, pp.317-64.
Rolls, E.T. 2004. The functions of the orbitofrontal cortex. *Brain Cognition.* **55**(1), pp.11-29.
Rosen, G. 2014. Culpability and duress: A case study. *Proceedings of the Aristotelian Society.* **Supp. Vol. 88**(1), pp.69-90.

Roskies, A.L. 2003. Are ethical judgments intrinsically motivational? Lessons from "acquired sociopathy" [1]. *Philosophical Psychology.* **16**(1), pp.51-66.
Roskies, A.L. 2011. A puzzle about empathy. *Emotion Review.* **3**(3), pp.278-80.
Salmela, M. 2011. Can emotion be modelled on perception? *Dialectica.* **65**(1), pp.1-28.
Sawin, D.B. and Parke, R.D. 1980. Empathy and fear as mediators of resistance-to-deviation in children. *Merrill-Palmer Quarterly of Behaviour and Development.* **26**, pp.123-34.
Scanlon, T.M. 1998. *What we owe to each other.* Cambridge, MA: Harvard University Press.
Scanlon, T.M. 2008. *Moral dimensions: Permissibility, meaning, blame.* Cambridge, Mass.: Harvard University Press.
Schachter, S. and Singer, J.E. 1962. Cognitive, social, and physiological determinants of emotional state. *Psychological Review.* **69**(5), pp.379-99.
Schroeder, M. 2007. Reasons and agent-neutrality. *Philosophical Studies.* **135**(2), pp.279-306.
Seara-Cardoso, A. and Viding, E. 2014. Functional neuroscience of psychopathic personality in adults. *Journal of Personality.* **83**(6), pp.723-37.
Shand, A.F. 1918. Emotion and value. *Proceedings of the Aristotelian Society.* **19**, pp.208-35.
Shoemaker, D. 2007. Moral address, moral responsibility, and the boundaries of the moral community. *Ethics.* **118** (October), pp.70-108.
Shoemaker, D. 2009. Responsibility and disability. *Metaphilosophy.* **40**(3-4), pp.438-61.
Shoemaker, D. 2011. Attributability, answerability, and accountability: Toward a wider theory of moral responsibility. *Ethics.* **121**(3), pp.602-32.
Sifferd, K. and Hirstein, B. 2013. On the criminal culpability of successful and unsuccessful psychopaths. *Neuroethics.* **6**(1), pp.129-40.
Skodol, A.E., Siever, L.J., Livesley, W.J., Gunderson, J.G., Pfohl, B. and Widiger, T.A. 2002. The borderline diagnosis II: Biology, genetics, and clinical course. *Biological Psychiatry.* **51**, pp.951-3.
Slobogin, C. 2000. An end to insanity: Recasting the role of mental disability in criminal cases. *Virginia Law Review.* **86**(6), pp.1199-247.
Smart, J.J.C. 1969. Free-will, praise and blame. *Mind.* **78**(3), pp.337-61.
Smith, G.T. and Oltmanns, T.F. 2009. Scientific advances in the diagnosis of psychopathology: Introduction to the special section. *Psychological Assessment.* **21**(3), pp.241-2.
Smith, M. 1994. *The moral problem.* Oxford: Blackwell.
Smith, M. 1995. Internal reasons. *Philosophy and Phenomenological Research.* **55**(1), pp.109-31.
Smith, R.J. 1984. The psychopath as moral agent. *Philosophy and Phenomenological Research.* **45**(2), pp.177-93.
Sneddon, A. 2005. Moral responsibility: The difference of Strawson, and the difference it should make. *Ethical theory and moral practice.* **8**(3), pp.239-64.
Sobel, D. 2001. Subjective accounts of reasons for action. *Ethics.* **111**(3), pp.461-92.
Solomon, R. 1976. *The passions: Emotions and the meaning of life.* New York: Doubleday.

Solomon, R. 2004. Emotions, thoughts and feelings: Emotions as engagements with the world. In: Solomon, R. ed. *Thinking about feeling.* Oxford: Oxford University Press, pp.76-88.

Sommer, M., Sodian, B., Dohnel, K., Schwerdtner, J., Meinhardt, J. and Hajak, G. 2010. In psychopathic patients emotion attribution modulates activity in outcome-related brain areas. *Psychiatry Research.* **182**(2), pp.88-95.

Southwood, N. 2011. The moral/conventional distinction. *Mind.* **120**(479), pp.761-802.

Speak, D. 2005. Semi-compatibilism and stalemate. *Philosophical Explorations.* **8**(2), pp.95-102.

Stiles, J. and Jernigan, T.L. 2010. The basics of brain development. *Neuropsychology Review.* **20**(4), pp.327-48.

Stocker, M. 1987. Emotional thoughts. *American Philosophical Quarterly.* **24**(1), pp.59-69.

Stocker, M. 1994. Emotions and ethical knowledge: Some naturalistic connections. *Midwest Studies in Philosophy.* **19**(1), pp.143-58.

Stout, M. 2005. *The sociopath next door.* New York: Random House.

Strawson, P.F. 1980. P. F. Strawson replies. In: Straaten, Z.V. ed. *Philosophical subjects: Essays presented to P. F. Strawson.* Oxford: Clarendon Press, pp.260-7.

Strawson, P.F. 2008. Freedom and resentment. *Freedom and resentment and other essays.* Abingdon, Oxon: Routledge, pp.1-28.

Strayer, J. 1989. What children know and feel in reponse to witnessing affective events. In: Saarni, C. and Harris, P. eds. *Children's understanding of emotions.* New York: Cambridge University Press, pp.259-89.

Talbert, M. 2008. Blame and responsiveness to moral reasons: Are psychopaths blameworthy? *Pacific Philosophical Quarterly.* **89**, pp.516-35.

Talbert, M. 2012. Accountability, aliens, and psychopaths: A reply to Shoemaker. *Ethics.* **122**(3), pp.562-74.

Todd, C. 2014. Emotion and value. *Philosophy Compass.* **9**(10), pp.702-12.

Torgerson, S., Czajkowski, N., Jacobson, K., Rechborn-Kjennerud, T., Roysamb, E.Neale, M.C. and Kendler, K.S. 2008. Dimensional representations of DSM-IV Cluster B personality disorders in a population-based sample of Norwegian twins: A multivariate study. *Psychological Medicine.* **38**, pp.1617-25.

Vargas, M. and Nichols, S. 2007. Psychopaths and moral knowledge. *Philosophy, Psychiatry and Psychology.* **14**(2), pp.157-62.

Viding, E. and Larsson, H. 2010. Genetics of child and adolescent psychopathy. In: Salekin, R.T. and Lynam, D.R. eds. *Handbook of child and adolescent psychopathy.* New York: The Guilford Press, pp.113-34.

Vincent, N.A. 2011. A structured taxonomy of moral responsibility concepts. In: Vincent, N.A., et al. eds. *Moral responsibility.* Dordrecht: Springer Netherlands.

Vitacco, M.J. 2007. Psychopathy. *The British Journal of Psychiatry.* **191**, pp.357-.

Waldman, I.D.S.H.R. 2007. Genetic and environmental influences on psychopathy and antisocial behavior. *Handbook of psychopathy.* pp.205-28.

Wallace, R.J. 1994. *Responsibility and the moral sentiments.* Cambridge, Mass.: Harvard University Press.

Wallace, R.J. 2002. Replies to reviews of 'responsibility and moral sentiments'. *Philosophy and Phenomenological Research.* **64**(3), pp.707-27.

Wallace, R.J. 2007. The argument from resentment. *Proceedings of the Aristotelian Society.* **CVII**(3), pp.295-318.
Ward, T. 2010. Psychopathy and criminal responsibility in historical perspective. In: Malatesti, L. and McMillan, J. eds. *Responsibility and psychopathy: Interfacing law, psychiatry and philosophy.* Oxford: Oxford University Press, pp.8-30.
Waters, S.F., West, T.V. and Mendes, W.B. 2014. Stress contagion: Physiological covariation between mothers and infants. *Psychological Science.* **25**(4), pp.934-42.
Watson, G. 1987. Responsibility and the limits of evil: Variations on a Strawsonian theme. In: Schoeman, F. ed. *Responsibility, character and the emotions: New essays in moral psychology.* Cambridge: Cambridge University Press, pp.256-86.
Watson, G. 1996. Two faces of responsibility. *Philosophical Topics.* **24**, pp.227-48.
Watson, G. 2013. Psychopathic agency and prudential deficits. *Proceedings of the Aristotelian Society.* **113**(3pt3), pp.269-92.
Whiting, D. 2012. Are emotions perceptual experiences of value? *Ratio.* **25**(1), pp.93-107.
Williams, B. 1981. Internal and external reasons. *Moral luck: Philosophical papers 1973-1980.* pp.101-13.
Wispé, L. 1987. History of the concept of empathy. In: Eisenberg, N. and Strayer, J. eds. *Empathy and its development.* Cambridge: Cambridge University Press, pp.17-37.
Woodworth, M. and Porter, S. 2002. In cold blood: Characteristics of criminal homicides as a function of psychopathy. *Journal of Abnormal Psychology.* **111**(3), pp.436-45.
Wootton, B. 1959. *Social science and social pathology.* London: G. Allen and Unwin.
Yang, Y., Raine, A., Colletti, P., Toga, A.W. and Narr, K.L. 2010. Morphological alterations in the prefrontal cortex and the amygdala in unsuccessful psychopaths. *Journal of Abnormal Psychology.* **119**(3), pp.546-54.
Yang, Y., Raine, A., Colletti, P., Toga, A.W. and Narr, K.L. 2011. Abnormal structural correlates of response perseveration in individuals with psychopathy. *The Journal of Neuropsychiatry and Clinical Neurosciences.* **23**(1), pp.107-10.
Yang, Y., Raine, A., Lencz, T., Bihrle, S., LaCasse, L. and Colletti, P. 2005. Volume reduction in prefrontal gray matter in unsuccessful criminal psychopaths. *Biological Psychiatry.* **57**(10), pp.1103-8.
Yang, Y., Raine, A., Narr, K.L., Colletti, P. and Toga, A.W. 2009. Localization of deformations within the amygdala in individuals with psychopathy. *Archives of General Psychiatry.* **66**(9), pp.986-94.
Yen, S., Shea, M.T., Battle, C.L., Johnson, D.M., Zlotnick, C., Dolan-Sewell, R., Skodol, A.E., Grilo, C.M., Gunderson, J.G., Sanislow, C.A., Zanarini, M.C., Bender, D.S., Rettew, J.B. and McGlashan, T.H. 2002. Traumatic exposure and posttraumatic stress disorder in borderline, schizotypal, avoidant, and obsessive-compulsive personality disorders: Findings from the collaborative longitudinal personality disorders study. *The Journal of Nervous and Mental Disease.* **190**(8), pp.510-8.
Zanarini, M.C., Yong, L., Frankenburg, F.R., Hennen, J., Reich, D.B., Marino, M.F. and Vujanovic, A.A. 2002. Severity of reported childhood sexual abuse and its relationship to severity of borderline psychopathology and psychosocial

impairment among borderline inpatients. *Journal of Nervous and Mental Disease.* **190**(6), pp.381-7.

Zarpentine, C. 2007. Michael Smith, rationalism, and the moral psychology of psychopathy. *Florida Philosophical Review.*

Zavaliy, A.G. 2008. Absent, full and partial responsibility of the psychopaths. *Journal for the Theory of Social Behaviour.* **38**(1), pp.87-103.