Forthcoming in R. Langdon and C. Mackenzie, *Emotions, Imagination and Moral Reasoning*, New York: Psychology Press.

Emotions, Reflection and Moral Agency

Catriona Mackenzie

1. Introduction

The question of what role emotions play in moral deliberation, judgment and action has been a central concern of philosophical moral psychology since Plato, Aristotle and the Stoics. In Aristotle's view virtuous action is not just a matter of doing the right thing at the right time and in the right manner. It also involves correct perception: being able to discern the morally salient features of the eliciting situation and respond appropriately. This involves both rational and emotional capacities: knowing what principles are relevant to the situation and how they should be applied, but also having emotional responses that are appropriate to the situation – feeling anger or compassion or regret when the situation merits such responses. Hume also thought that virtue requires correct perception of the morally salient features of the eliciting situation. For Hume correct moral perception arises from the passions, or sentiments, but not from 'hot' passions; rather it requires adopting the point of view of what he refers to as the 'calm' passions or the 'corrected sentiments'. Hume compares the corrected, moral sentiments to corrective judgments in sense perception. Because we know that sense perception is not always veridical, in our judgments we learn to correct for common sensory illusions. Similarly the moral sentiments are passions or emotions that have been corrected by reflection. Thus, although Aristotle and Hume recognised that emotions play a key role in alerting us to relevant reason-giving considerations, they were well aware that emotions can also distort moral perception, deliberation and judgment. This is why, for both, moral agency requires that we reflect on and regulate our emotional responses.

Moral psychologist David Pizarro (Pizarro, 2000) points to three aspects of emotions that seem to pose a threat to good moral judgment. First, the partiality of emotions seems to conflict with the requirement that moral principles be impartial. Hume was aware of this problem. Although he anticipated the view now widely accepted in developmental and social psychology and in cognitive neuroscience that the capacity for affective empathy (or what Hume called sympathy) is a necessary precondition for moral motivation, he also recognised that the scope of our sympathies is limited. We are more likely to feel compassion for those we care about, or for people more like ourselves than for distant others. This is why empathy is not sufficient for morality and why the sentiments need to be corrected by general principles. Second, emotions can sometimes latch onto morally irrelevant features of a situation and influence our moral behaviour and judgments. Pizarro cites the literature in social psychology showing that something as trivial as finding a dime can have a positive influence on people's motivations to help others. Third, emotions are often conceptualised as mere feelings, passive affective states over which we exercise little voluntary control. This conception of the emotions is reflected in common linguistic usage, for example when we talk about being overcome by anger, and it is why emotions seem to conflict with the kind of reasoned, principled action that morality is thought to require of us.

It is these features of the emotions that seem to be behind Kant's view that emotions, even beneficent emotions such as sympathy, provide an unreliable foundation for morality. Contrary to a common misinterpretation, Kant does not deny that emotions such as compassion might play an important role in motivating actions that accord with moral requirements, but he thinks that what makes an action genuinely moral is that it is guided and motivated by a rational universal principle. Kant's influence on Kohlberg's account of moral development, and via Kohlberg, on contemporary moral psychology is well known. Kohlberg understands moral judgment as a process of reasoning from moral principles to a conclusion about how one ought to act in a specific situation. Moral development is marked by a progress from judgments guided by lower-level egocentric principles to judgments guided by principles based on social conformity to the most mature level of moral judgment guided by impartial, universal principles focused on issues of harm, justice and rights.

Recently this cognitivist tradition in moral psychology seems to have been challenged on a number of different fronts. Research in cognitive neuroscience on empathy deficits in psychopathy and autism, seems to show that empathy is a necessary condition for moral development; Damasio's work with patients with 'acquired sociopathy' arising from damage to the ventromedial pre-frontal cortex seems to show that while these patients' reasoning abilities are still intact, they suffer affective and emotional deficits resulting in highly impaired decision-making capacities and, in younger patients, moral incompetence; neuro-imaging studies conducted by Joshua Greene and colleagues, seem to show that regions of the brain associated with the emotions are highly active in processing moral judgments; Jonathon Haidt's social intuitionist model of moral judgment claims to show that moral judgments are not the result of effortful, conscious reasoning from principles but of moral intuitions, which Haidt characterises as automatic, affective responses or gut feelings, akin to perceptions.

I am very sympathetic to the view that emotions are crucial to moral agency and I think that philosophical moral psychology has much to learn from this empirical literature. I also think that moral psychology in the tradition of Kant and Kohlberg has over-emphasised the role of rational reflection and principle-based reasoning in moral cognition and has tended to construe emotions as mere feelings, lacking in cognitive content and as at least morally unreliable, if not in conflict with morality. And I agree with Haidt that this tradition has paid insufficient attention to the social dimensions of morality. However I am troubled by the work of Haidt and Greene for a number of reasons.

First, although Haidt thinks his social intuitionist model of moral judgment poses a serious challenge to the Kantian/Kohlbergian tradition of moral psychology, the problem with the SIM is that it upholds rather than challenges, the tradition's impoverished conception of emotions as automatic affects over which we can exercise little reflective and self-regulatory control, just reversing the order of priority given within this tradition to affect and reason. Greene assumes a similarly impoverished conception of emotions. Second, Haidt and Greene also uphold the tradition's conception of moral reasoning as primarily principle-based reasoning and its rationalist conception of moral reflection. While reasoning does and should play a role in moral thinking, it is a mistake to conceptualise moral reflection so narrowly. The scope of moral reflection is much wider than Haidt and Greene allow and it involves the exercise of complex emotional, imaginative and agential capacities. It is also a social process. On this I am in agreement with Haidt, although I disagree with

Haidt's construal of the social dimensions of morality. Third, Haidt and Greene narrow the focus of moral psychology to moral judgments about others people's actions and characters and this is why their empirical research focuses on participants' one-off, snap judgments in response to abstract hypothetical moral dilemmas. But in my view it is not clear how much this research can tell us about moral reflection and decision-making in everyday contexts, which is extended over time and involves reflection on our responsibilities and commitments to others, our goals and values, our interpretations of and judgments about our own behaviour and emotional responses, and so on. To summarise these concerns, I think Haidt and Greene's approaches to moral psychology mischaracterize the moral emotions and moral reflection and present a skewed picture of moral agency.

In the following section of the paper I outline the central claims of Haidt and Greene and motivate these concerns. In the final section I outline an alternative picture of emotions, reflection and their role in moral agency that presents an intermediate position between moral intuitionism and rationalism. Just to clarify the parameters of my project here, Haidt and Green see their work as going beyond their descriptive claims about the role of intuitions and emotions in moral judgment and as having important implications for meta-ethics and normative ethics. I am not convinced that their empirical research provides support for a particular meta-ethical theory, such as sentimentalism (the view that certain evaluative concepts, including moral evaluations, are essentially constituted by specific emotional responses). Contrary to claims by Greene and Peter Singer, I'm even less convinced that it provides support for consequentialism as a normative theory. (Consequentalism holds that the rightness or wrongness of actions is determined solely by their overall beneficial or harmful consequences.) However, I won't be addressing these issues in this paper except in passing.

2. Haidt and Greene on moral judgment

Despite some important differences between the views of Haidt and Greene, which I'll point to, their views converge in many respects, particularly in their conceptions of emotions and moral reasoning. Haidt's social intuitionist model aims to describe the causal processes involved in moral reasoning. His thesis, in brief, is that moral intuition precedes and causes moral judgments and moral reasoning is a process of rationalisation, of searching for reasons to support and justify these judgments. This thesis not only reverses Kohlberg's account of the relationship between moral reasoning and moral judgment; the post hoc rationalisation claim debunks Kohlberg's conception of moral reasoning as the search for impartial and universal principles of justification of our moral judgments. Greene agrees with the claim that moral judgment is mostly driven by intuition. He also agrees that much moral reasoning, particularly deontological moral reasoning, which appeals to principles based on rights and duties, is post hoc rationalisation. In fact he suggests that 'what deontological moral philosophy really is, what it is *essentially*, is an attempt to produce rational justifications for emotionally driven moral judgments' (Greene, 2008: 39). However Greene argues that consequentialist moral reasoning employs affectively neutral cognitive processes, such as those involved in cost-benefit analysis. He regards it as genuine reasoning, rather than post hoc rationalisation, which can yield impartial, universal principles.

2.1. Moral Intuitions

Haidt claims that moral judgments of rightness or wrongness are automatic, affective responses or gut feelings, akin to perceptions. These automatic, intuitive responses to the actions or character of others are evaluative, but only thinly evaluative, involving judgments of good/bad, like/dislike, approach/avoid, and so on. Haidt compares intuitive moral judgments to aesthetic judgments. Just as we might respond automatically to a landscape, judging it as beautiful, so he claims we make automatic moral judgments about others' actions or characters. In neither case do we make the judgment on the basis of a process of conscious and deliberate reasoning – weighing evidence, using inferential or deductive reasoning to reach a conclusion. And, in both cases we are often not able to articulate the basis of our judgment – explaining why we find the landscape beautiful or why we regard an action as wrong. In support of this claim, Haidt cites evidence from social attitude and stereotyping studies, which suggest that many of our judgments about others arise from automatic, affective first impressions. I do not intend to assess whether this evidence does support the social intuition thesis, although Cordelia Fine has argued convincingly that Haidt is very selective in his use of social cognition studies and does not address the literature showing that people can exercise more control over these affective responses than Haidt claims (Fine, 2006).

Haidt and Greene both claim Hume as a philosophical predecessor for the moral intuitionist thesis. However, note that for Hume although the moral sentiments arise from natural sentiments, such as empathy or the love between parents and children, they are neither partial nor automatic affects; rather they are detached, impartial reflective emotional responses developed under the guidance of general principles. For Hume the education of the moral sentiments is thus a process whereby thinly evaluative, partial affective responses become more considered, more discriminating and more thickly evaluative as a result of social interaction and reflection upon our moral experience. Hume therefore has some basis for distinguishing correct moral perceptions and morally appropriate emotional responses from unreflective affective responses. A major problem for Haidt's social intuition model is that because he downplays the role of reflection in enabling us to regulate our emotional responses, as I'll explain later, he undercuts the ground for making these kinds of normative distinctions and for explaining the basis on which some moral intuitions are a reliable source of moral knowledge.

This problem also infects Haidt's account of moral development. An obvious question about moral intuitions is whether they arise from prior moral socialisation or from prior moral reasoning that has now become automatic. Haidt and Bjorklund (2008) reject such explanations and propose that moral intuitions arise from innate, distinct moral modules that have become encoded in the brain through the evolutionary process. These modules, they argue, dispose us to be responsive to considerations related to five values: care and the avoidance of harm and suffering; reciprocity and fairness; hierarchy and authority; purity or sanctity; and loyalty to in-group members. These foundational values underpin all moral systems and conceptions of the virtues. But different cultures construct specific virtues, such as honesty or kindness, in different, although overlapping, ways. Moral development is a process of the endogenous unfolding, or externalisation of the moral modules at different developmental stages, assisted by socialisation processes and peer networks that enculturate individuals into culturally specific moral practices and understandings of the virtues. As Daniel Jacobson has argued, however, Haidt and Bjorklund's five foundational values are so abstract that they might figure in any moral view, no matter how morally heinous, so it is not clear how they are supposed to provide a basis for moral knowledge (Jacobson, 2008: 226). And, despite their denial that their position is a form of moral relativism, their account of moral development suggests that moral intuitions are biological response mechanisms that have been fine-tuned in different ways by cultural practices. To quote Jacobson, 'on this view, moral knowledge is simply the habituated ability to see things the way others see them in your parish: to have the same intuitions as others in your society' (Jacobson, 2008: 228). It is thus not clear how Haidt and Bjorklund can distinguish morality from social conformity.

Greene is not subject to this same objection because, for reasons that I'll explain shortly, he does not think that the kind of moral thought that arises from moral intuitions can provide the basis for genuine moral knowledge. But, like Haidt, he also conceptualises moral intuitions as automatic, emotional responses. And, like Haidt, he finds support for this conception of moral intuitions in dual process models of judgment and problem solving. According to dual process models, cognition involves two parallel processing systems: a default, hot, affective system that is both phylogenetically and ontogenetically primary and that operates largely automatically; and a cool, controlled cognitive system that is slow, conscious and involves verbal reasoning. Haidt and Greene extend dual process models to the domain of moral judgment: moral intuitions are part of the hot affective system while moral reasoning is a product of the cool cognitive system.

I am not in a position to assess dual process models of cognition more generally. However, I think that Haidt and Greene's extension of dual process models to the moral domain serves to perpetuate a false emotions/reason dualism in moral psychology. It also perpetuates an impoverished conception of emotions as mere feelings or affects. Greene, for example, characterises moral emotions as 'blunt biological instruments' (Greene, 2008: 71), and he describes moral intuitions as alarm-like emotional responses, which are rigid, inflexible and resistant to reason – the result of 'evolutionary adaptations that arose in response to the demands and opportunities created by social life' (Greene, 2008: 60). Because they are automatic, moral intuitions and emotions enable social creatures such as ourselves to respond quickly, efficiently and reliably to the needs of others. Yet Greene also claims that because they are the product of evolutionary forces, moral intuitions are suspect sources of moral knowledge that cannot ground adequate moral reasoning. In his view, moral intuitions are fundamentally unreliable because they involve emotions and because they 'appear to have been shaped by morally irrelevant factors having to do with the constraints and circumstances of our evolutionary history' (Greene, 2008: 75). As Neil Levy has argued, however, the fact that some of our basic moral intuitions might have evolved under non-moral selection pressures does not show, in and of itself, that these intuitions are suspect or that moral reasoning that is responsive to them ought to be distrusted (Levy, 2007: 300-306).

Greene cites neuro-imaging studies involving what he calls 'personal' and 'impersonal' dilemmas, as evidence both for his evolutionary claims about moral intuitions and for the dual process view of moral judgment. These dilemmas involve variants of Judith Jarvis Thomson's famous trolley and footbridge problems. Many philosophers would argue that the intention/foresight distinction seems to provide the most plausible principled basis for explaining the intuition that there is an important moral difference between these cases, namely that in the trolley case you foresee that by pulling the lever you will bring about the worker's death but you are not intentionally aiming to harm him, whereas in the footbridge case you are intentionally aiming to cause the death of the fat stranger. In experimental studies, participants also judged these cases as morally different without necessarily being able to articulate why.

Greene argues that there is no moral justification for drawing any distinction between the cases since both involve the death of one person to save five others and so can be justified on consequentialist grounds. His hypothesis is that people respond differently to these cases because the footbridge case involves a 'personal' moral violation, which involves directly bringing about bodily harm to another person. Such personal violations trigger negative emotional responses to the harm which conflict with the more sophisticated cognitive processing required to reason to a consequentialist judgment. On the other hand, the trolley case involves an 'impersonal' moral violation. In this case, because the harm is brought about by indirect means it does not trigger negative emotional responses and so people are able to reason logically to a consequentialist conclusion. Greene argues that our negative emotional responses to ‘personal’ moral violations are the result of innate but evolutionarily primitive responses to interpersonal violence that pre-date the development of our rational capacities. And he claims that neuro-imaging studies of participants’ brains as they responded to a range of personal and impersonal dilemmas provide support for this hypothesis, in two ways.

First, neuro-imaging showed increased neural activity in regions of the brain associated with emotional response and social cognition when participants were responding to cases involving personal harm, whereas when they were responding to cases involving impersonal harm there was greater activity in brain regions associated with higher cognitive functions. There has been considerable discussion in the cognitive neuroscience literature about whether neuro-imaging studies do support the dual process theory of moral judgment. For example, Jorge Moll and Ricardo de Oliveira-Souza (2007) dispute the dual process theory, and argue that neuro-imaging data shows that moral appraisals involve a complex interaction of cognitive and emotional mechanisms (2007, 2008). I am not competent to assess this debate.

Second, Greene claims that his data showed that the reaction time required to reach a consequentialist judgment in cases involving personal harm was greater than the reaction time required to reach a non-consequentialist judgment. There was no such difference in reaction time in cases involving impersonal harm. Greene argues that the differences in reaction time support the view that reasoning to a consequentialist conclusion in personal cases requires overriding strong negative emotional responses. However, Jonathon McGuire (2009 forthcoming) has recently re-analysed Greene’s data and conducted a detailed item analysis of participants’ reaction times to so-called personal and impersonal dilemmas. This re-analysis shows that Greene’s results showing differential reaction times are actually driven by a small sub-set of personal dilemmas, to which participants reacted very quickly, and which they almost universally judged to be impermissible (eg. not rescuing someone who has been involved in an accident because you don’t want to ruin the upholstery in your car).

Once these were excluded from the data set, there was no significant difference between reaction times to personal and impersonal dilemmas. Greene has recently conceded that the personal/impersonal distinction is problematic, but he thinks the imaging studies still provide support for a dual process model of moral judgment.

From the foregoing it should be clear that Haidt and Greene both conceptualise emotions as automatic affective processes and that Greene in particular thinks of them as primitive biological mechanisms that give rise to unreliable moral thinking. Just to clarify my argument at this point, I am not disputing that intuitions and emotions do play a role in moral cognition. What I want to dispute is the way Haidt and Greene conceptualise both the emotions and their role in moral thought.

2.2. Moral Reasoning

If moral judgments are just automatic affective responses, what is moral reasoning and what role does it play in moral cognition? Haidt distinguishes four different kinds of moral reasoning: post hoc rationalisation, reasoned social persuasion, reasoned judgment and private reflection. He argues that reasoned judgment and private reflection are rare species of moral reasoning; most moral reasoning either takes the form of *post hoc rationalisation* or *reasoned social persuasion*. In Haidt's view most moral reasoning is *post hoc rationalisation*. It is moral intuitions that do the real causal work in explaining people's moral judgments. It is only when pressed to provide reasons for our moral judgments that we appeal to moral principles. Haidt claims that the reason philosophers and psychologists like Kohlberg emphasise the role of principled reasoning in moral judgment is that they assume that reasoning aims to track the truth. However, in Haidt's view the 'goal' of reasoning – and here what he seems to mean by 'goal' is an evolutionary goal– is not to track the truth but to achieve social integration and harmony and to influence and persuade others. This conception of the goal of reasoning underpins Haidt's account of *reasoned persuasion*. Haidt thinks that we do sometimes engage in reasoned moral discussion and debate with other people, sometimes with the aim of reaching a community consensus. But he doesn't see this as a process of collective reasoning that is driven by concerns about moral justification or truth. It is rather a process of trying to persuade others to our partial and interested perspectives, a process of using rhetoric to trigger the desired affectively valenced intuitive responses in others. When it comes to moral judgments, he claims, our interest in reasons is not the disinterested interest of the scientist but the partial interest of the lawyer. In other words, most activities of reason-giving are nothing more than exercises in rhetorical persuasion.

Haidt does concede that sometimes people can engage in genuine moral reasoning, or *reasoned judgment*, which he characterises as systematic, step-by-step, conscious reasoning from first principles to reach consistent moral conclusions that may override our initial intuitions. Like Greene, he seems to think that consequentialist reasoning is the best exemplar of this kind of reasoning, citing Peter Singer's work as an example, and suggesting that most objections to Singer's conclusions arise from the recalcitrance of people's moral intuitions. Haidt also concedes that sometimes – for example when we have no clear intuitions, or when our intuitions conflict – we can engage in a process of *private reflection* or inner moral dialogue. He suggests that moral perspective taking – imaginatively putting oneself in another person's place – is one of the chief means of doing so. However he thinks such reflection plays very little role in moral cognition.

Haidt cites a range of different studies in social cognition as evidence for his claims about moral reasoning. These include studies which purport to show that when people are not aware of the cognitive processes causing their behaviour, for example because they have acted under post-hypnotic suggestion or subliminal priming, they search for plausible-sounding reasons to explain that behaviour; defensive motivation studies, which show that people adjust their beliefs and thinking to preserve coherence with self-definitional attitudes, such as their values and moral commitments; and biasing studies which show that people are not very good at understanding and assessing evidence or providing evidence for their views and that their assessment of evidence is biased, in other words they put greater weight on evidence that supports their beliefs while discounting other evidence that seems to question those beliefs. Haidt also cites evidence from his own moral judgment studies.

One study (Haidt, Koller and Dias, 1993), aimed to resolve a debate about whether the moral domain is universally limited to issues of harm, justice and rights or whether in some cultures it extends to issues more typically regarded as matters of social convention, such as practices concerning food, sex roles etc. Participants from different SES groups in Brazil and the US were presented with a series of vignettes involving what Haidt stipulates to be ‘harmless taboo violations’, such as a family who eats its pet dog after it has been killed by a car or a man who masturbates with a chicken carcass and then cooks and eats it. The researchers found that high and low SES groups in both Brazil and US, expressed disgust at these actions or at the least found them strange, but high SES groups did not regard them as morally wrong. On the other hand, low SES groups, while acknowledging that these actions did not cause harm to anyone, nevertheless regarded them as moral violations and as universally wrong. Some participants from both groups, while quick to make judgments, were often at a loss to explain why and became puzzled and confused when pressed for their reasons. Haidt refers to this inability to provide reasons for one’s moral judgment as ‘moral dumbfounding’. In a later study (Haidt, Bjorklund, & Murphy, 2000) participants seem to have been particularly dumbfounded when pressed to provide reasons for their judgments in cases involving other taboo violations which Haidt characterises as harmless – such as consensual adult incest, or cannibalism of an unclaimed corpse in a pathology lab – and in cases designed to elicit strong disgust reactions (taking a sip from a drink in which a dead, sterilized cockroach had just been dipped). Haidt also cites a study (Wheatley and Haidt 2005) in which participants were hypnotised to experience disgust in response to completely neutral words (eg. take or often) and then asked to read moral judgment stories. Those participants primed to experience disgust expressed more severe judgments.

Haidt thinks moral dumbfounding provides evidence that because we are not aware of the cognitive processes that give rise to our moral judgments, when pressed to justify them we engage in a post hoc search for reasons to bolster these judgments. Dumbfounding occurs when we are unable to find such reasons. In other cases we appeal to prior moral theories to justify our judgments. Haidt defines moral theories as ‘a pool of culturally supplied norms for evaluating and criticizing the behavior of others’ (Haidt, 2001:16). As a descriptive claim, Haidt might be correct that people may not be very good at providing good reasons for their judgments. He is undoubtedly also correct that moral reasoning can be biased, defensively motivated, and sometimes a matter of post hoc rationalisation of prejudice. However, just because people can be bad reasoners, or dumbfounded when it comes to explaining the reasons for their judgments, this is not sufficient to show that moral reasoning is a matter of post hoc rationalisation.

There are several problems with Haidt's claims about moral reasoning. First, as Daniel Jacobson has pointed out, Haidt conflates a causal claim about the origin of evaluative judgments with a justificatory claim about what justifies those judgments. Haidt thinks that so-called rationalists are making the causal claim that our moral judgments arise from reasoning, to which his response is that they arise from moral intuitions and that reasoning is just post hoc rationalisation. However, rationalism is a justificatory claim to the effect that moral judgments can only be justified by rational principles. So one can hold this view about justification while also acknowledging that people can be bad reasoners or not able to articulate the principles on which their judgments are implicitly based. As I have already indicated, Greene's view is not subject to this objection because he does think that moral judgments must be rationally justified and he thinks that consequentialist reasons can provide such justification. I don't think the evidence he presents supports this claim, but this is a separate issue. Second, although Haidt does concede that sometimes a process of giving and exchanging reasons with others will yield better understanding and judgment, it is not clear on what grounds he can in fact distinguish good or justified moral reasons from mere rationalisation or rhetorical persuasion since Haidt does not explain what constitutes moral knowledge, as distinct from socially agreed norms, as I explained earlier. Third, while I agree with Haidt that moral reflection and reasoning is a social process, I fundamentally disagree with his debunking characterisation of this process as primarily rhetorical persuasion aimed at influencing other people to one's point of view. In my view Levy provides a much better characterisation of the sense in which moral reasoning is social – namely that moral knowledge, like all knowledge, is an ongoing, distributed, community-wide enterprise in which, through moral debate and under the pressure of objection and argument, our judgments are tested and revised (Levy, 2007: 308-316).

I now want to press a more general objection to Haidt and Greene's claims about moral judgment and moral reasoning. The exclusive focus of their work in moral psychology is judgments about the rightness or wrongness of particular actions in hypothetical moral dilemma situations and moral reasoning is restricted to reasoning about such judgments. However, it is far from evident that participants' one-off responses to such hypothetical dilemmas can tell us much about ordinary moral reasoning and reflection, for a number of reasons. First, many of the scenarios describe situations that people are likely to regard as unrealistic, abstract and under-specified, because they are far removed from every day moral concerns and they lack the kind of contextual information that usually feeds into our decision-making and judgments. This includes information about the surrounding circumstances of an action, such as the events leading up to it, about people's characters and patterns of behaviour over time, and about the possible effects of an action on people's relationships and self-concepts. In the absence of such information, it is not particularly surprising that people might be dumbfounded when asked whether it is morally right for adult siblings to have allegedly harmless consensual sex, or to cannibalise an unclaimed corpse in a mortuary. It is noteworthy that participants in Haidt's (2000) study were not similarly dumbfounded when presented with Kohlberg's Heinz dilemma, which presents a scenario that most people can relate fairly easily to their moral experience.

Second, the scope of our everyday moral reasoning and reflection is much broader than these experimental situations allow. Because moral decision-making and action usually has significant repercussions for ourselves and others, deliberation needs to take account of multiple factors as relevant to the specific situation in question: moral intuitions and principles certainly, but also responsibilities and commitments to others, our own and others' needs, our short and long-term goals, our values, the ideals we want to live up to, our understanding of our own characters – our virtues and limitations – and so on. Take a familiar example. Let's say I have a pressing deadline to meet and I work out a timetable for meeting it consistent with meeting my obligations to my students and my family. But then a colleague becomes ill and some arrangement needs to be made for his lectures next week. I feel an obligation to help out and I could teach the material, but I'm feeling under pressure already. One course of action that is open to me is to say to the head of department that I can't take on the extra load and that the lectures should be cancelled or that he needs to find someone else to take them. But I find that decision hard to square with my self-conception as a reliable and responsible colleague. On the other hand, I know that agreeing to do this is going to require me to work extra on the weekend and that I may not be able to go to my son's cricket match as I had promised. I'm worried that either way I'm going to let somebody down. The kind of moral reflection required to respond to this rather mundane moral dilemma involves emotional, imaginative, agential and reasoning skills that are more complex than either Greene or Haidt allow. I'll talk more about these skills a bit later. Here I want to point out that in response to a similar objection raised by Darcia Narvaez (Narvaez, 2008), Haidt has recently conceded that moral decision-making and deliberation does require this kind of broad-ranging moral reflection (Haidt & Bjorklund, 2008a). However he claims that the social intuitionist model was originally designed only to describe the causal processes involved in moral judgment, not moral decision-making and that moral judgment and moral decision-making are quite different cognitive processes that are not closely related functionally. I can see no good reason for accepting this claim. I would suggest that moral judgment and moral deliberation recruit the same emotional, imaginative, reasoning and reflective capacities, so in my view the objection still stands.

Third, in real world moral contexts, as this example shows, moral deliberation, reflection, decision-making and judgment are usually not one-off, single decisions but processes that are extended over time and are shaped by prior histories. My decision to take the lectures or not is likely to be shaped to some extent by my relationship with my colleague and the patterns of our prior interactions. If I suspect the genuineness of his illness and see it as yet another instance of a common pattern of reneging on his obligations and expecting other people to pick up the slack then my response to the head of department's request is likely to be quite different than if I know him to be a highly responsible person who would only cancel his classes if he is really ill. If I feel resentful towards my colleagues because I think they think of me as the departmental dogsbody, then again my thinking about the reasonableness of the request will be quite different than if I think that everyone in the department shares the load. Further, any decision I make will be nested in a complex set of interconnected decisions and judgments, both my own and that of other people, so that the process of reflection, deliberation and decision-making will be an iterative one, involving ongoing moral interaction and negotiation with others. Now, its just not clear to me how much neuro-imaging or moral judgment studies of one-off responses to abstract hypothetical dilemmas can illuminate the cognitive processes involved in such temporally extended, iterative, real world moral decision-making situations.

To recap, I have argued that Haidt and Greene operate with impoverished conceptions of moral intuitions and emotions as automatic affective processes, and an overly narrow conception of moral reasoning that doesn't do justice to the complexity of moral reflection. I have also objected to the post hoc rationalisation claims and Haidt's construal of the social dimensions of moral cognition. However, I haven't outlined an alternative picture of the role of emotions and reflection in moral agency. This is what I'm aiming to do in the final section of the paper. What I hope to do is to find a way of acknowledging Haidt's insights that intuitions, emotions, affects and social interaction play a far more central role in moral cognition than rationalist models of reasoning and reflection recognise, without simply reversing the rationalist view, as I think he does, or embracing his debunking conclusions about morality.

3. Emotions, Reflection and Moral Agency

Much contemporary philosophical emotions theory rejects the view that emotions are 'blunt biological instruments', as Greene describes them, or non-rational affective responses, mere feelings. This is not to deny that emotions involve characteristic affective and physiological components and give rise to characteristic action tendencies, some of which involve biological response mechanisms, for example, the fight or flight responses characteristic of fear. But the view is that the emotions also have cognitive content. There is a debate in the literature about how the cognitive dimensions of emotions should be understood. Judgmentalist theories construe the cognitions involved in emotions as beliefs or judgments. Evaluative appraisal theories argue that the cognitive component of emotions should be thought of as evaluative attitudes or appraisals that frame the way we perceive and interpret the eliciting situation. I am sympathetic to evaluative appraisal theories. On this view, jealousy for example, is an affect that might involve certain bodily and psychic feelings, but it also involves an evaluation of the eliciting situation that picks out certain features of the situation as salient, others as less salient. So a person who is feeling jealous is likely to notice and attach significance to aspects of his partner's behaviour that he might otherwise not pay attention to at all – such as her making a phone call or wearing a new piece of clothing. Or, in the example I gave earlier, construing my colleague's illness as part of a pattern of reneging on his commitments is likely to be an evaluative appraisal bound up with the emotion of resentment.

Evaluative appraisal theories can accommodate many of the insights that seem to underlie Haidt's moral intuitionism. Appraisal theories can accommodate the insight that emotions are quasi-perceptual in the sense that they are ways of seeing or construing a situation from a particular point of view or perspective. They can also accept that these appraisals are often quick responses to the eliciting situation that are not always amenable to deliberative control (Greenspan, 2003:117). And they can allow for affective dissonance – namely, that our emotional responses do not always cohere with our all things considered judgments. Sometimes this may be because this dissonance is an expression of deep-seated implicit evaluative attitudes that we do not endorse and that conflict with our more considered, conscious judgments (eg. racial stereotyping). These are the kinds of cases that drive Haidt's approach to moral intuitions. However, as Karen Jones has argued, sometimes affective dissonance – recalcitrant or 'outlaw' emotions – can be important for practical rationality, alerting us to reasons, including moral reasons, that we may not be consciously aware of or able to articulate (Jones, 2003). Compassion, for example, or friendship can help unseat racist convictions that agents might consciously endorse, as the example of Huck Finn attests. Because recalcitrant emotions, which may be no more than gut feelings, can attune us in this way to reasons to which we ought to attend, it need not be irrational to act on the basis of these emotions rather than on the basis of our all things considered judgments.

Evaluative appraisal theories thus hold that emotions can be rational, not just in the strategic sense that they can aid decision-making but in the normative sense that they can attune us to relevant reason-giving considerations. To say that emotions are rational in this sense is to say that they are responsive to features of the situation to which we ought to be responsive. It is also to say that the evaluation expressed in the emotion is correct and therefore that the emotion is an appropriate, or fitting, response to the situation. Grief is an appropriate or fitting response to the death of a loved one, compassion is an appropriate response to another's suffering, remorse is an appropriate response to wrongdoing. By the same token, emotions can be assessed as unreasonable or disproportionate, which is to say that the evaluations expressed in the emotion are not warranted by the situation. Road rage is a good example of an irrational emotional response, a response that is not fitting or warranted and that blinds us to the relevant reason-giving considerations.

A central focus of the moral socialisation of children is to train their capacities for recognising the specific emotional responses that are appropriate or fitting to the eliciting situation. Ronald de Sousa has developed an influential analysis of how this kind of socialisation works (de Sousa, 1987). De Sousa argues that we learn our emotional repertoires – our emotional vocabulary and our emotional responses – through what he calls paradigm scenarios. Paradigm scenarios involve two aspects: 'first, a situation type providing the characteristic objects of the specific emotion-type...and second, a set of characteristic or "normal" responses to the situation' (de Sousa, 1987:182). In de Sousa's view, we learn which properties are relevantly motivating for particular emotions and which situation types should elicit a particular emotional response partly as a function of in-built biological mechanisms, but largely through socio-cultural norms. Emotional education involves teaching children to identify and name particular emotional responses, teaching them which objects characteristically warrant those responses in which typical situations, and also teaching them how to respond, that is, which action tendencies are appropriate.

So the basis for our emotional understanding is developed in childhood, through the paradigm scenarios that taught us the meaning of particular emotions. However this understanding is layered by our subsequent encounters with a range of emotional paradigms. As a result, our emotional responses incorporate sometimes extended and complex scenarios, which may not be accessible to conscious introspection. The complexity explains both the degree to which emotions are intelligible and communicable to ourselves and others, and the degree to which they are opaque. Their intelligibility and communicability derives from the biological basis of many emotions, but also from the socio-cultural norms they embody. Their opacity derives from the fact these norms are also interpreted and played out in a variety of subtly or significantly different scenarios, reflecting the complex individual and familial biological and psychic histories that generated them.

Evaluative appraisal theories are thus quite consistent with a naturalistic outlook, with recognition of the social dimensions of the emotions and with acknowledging that we may be unaware of the cognitive processes that give rise to our emotions. But they are also responsive to the normative dimensions of emotional evaluations, to the fact that we don't just regard emotions as biological responses or outbursts of feeling. Our moral practices and norms, and our expectations of our own and other people's behaviour assume a normative conception of the emotions; that some emotional responses are appropriate or fitting and provide accurate evaluations of the eliciting situation and others do not. And our conception of moral agency assumes that people are responsible for reflecting upon and regulating their emotional responses. Appraisal theories thus provide an intermediate position between the kind of rationalism that discounts the capacity of emotions to attune us to relevant moral reasons (and I think Greene is actually a rationalist in this sense) and views like Haidt's which reduce emotions to gut responses and discount our capacities to reflect upon and regulate our emotional responses.

I want to conclude by saying something about the conception of moral agency and reflection that I think is assumed, and reasonably so, by our moral practices. My claim is that our moral practices operate under the assumption that as moral agents we regard ourselves and other moral agents as what Karen Jones calls 'reason-responders' (Jones, 2003). To be a reason-responder is to be the kind of creature who is capable of rationally guiding her actions in light of her best reasons, reasons that she regards as providing a *justification* for those actions. This involves a capacity to conceive of reasons as reasons, as distinct say from mere desires, automatic emotional responses or rationalizations. And this in turn requires complex second-order capacities to reflect on our desires, automatic emotional responses, and reasoning processes, and to consider whether they do provide us with genuine reasons for action, reasons that we should regard as authoritative in determining what we do. Say I find myself feeling enraged by the poor driving skills of another motorist and experience a very strong desire to shout at him abusively and make rude gestures. To think of myself as a reason-responder is to think of myself as an agent who is capable of reflecting on this desire, considering whether I have good reasons to act on it and regulating my actions accordingly. It is also to regard myself as committed to what Jones describes as 'the on-going cultivation and exercise of habits of reflective self-monitoring' (Jones, 2003:194).

Now this second-order capacity for reflective self-monitoring is sometimes understood in highly intellectualist terms, as requiring some kind of conscious, reflective endorsement of our desires and emotional responses. Sometimes reflective self-monitoring does take this form, however there are a range of other ways in which we engage in this kind of self-monitoring, deploying emotional and imaginative as well as rational skills. For example, what I call 'reactive emotions' (psychologists refer to these as meta-emotions) are second-order emotional evaluations of our first-order emotional responses. Let's imagine a young mother of a toddler who is being particularly trying as the mother is doing the grocery shopping in a supermarket. She finds herself growing increasingly irritable and impatient and more and more inclined to lose her temper and shout. Regulating this first-order emotional response need not be a matter of consciously reflecting on whether she endorses this response; it might take the form of experiencing reactive emotions of distress or shame about this response, prompting her to re-frame her perspective on her child's behaviour. Or she might take up her child's perspective, perhaps self-consciously, but perhaps not, coming to see the shopping trip from the child's perspective as boring and interminably long when all the child wants to do is to go to the playground. This re-framing of the situation might then prompt her to be more patient. Let's also imagine that the mother begins to notice a common pattern in her emotional responses to her child on shopping trips. She might then deliberately take steps to try to ensure that she is not required to take her child with her when she does the grocery shopping. This might be because she knows that she cannot rely on herself to control her emotional responses once the child starts nagging about lollies. In the same way an alcoholic might try to ensure that she avoids social situations in which alcohol is going to be consumed. These are all very ordinary ways in which we regulate our emotional responses, either by re-appraising the situation emotionally or imaginatively, or by trying to exercise control over the circumstances we put ourselves in. There is a significant empirical literature demonstrating that people do employ such strategies to exercise second-order control over their emotional reactions (see Pizarro and Bloom, 2003, Fine, 2006), so what I'm saying here is not new. However, my point is also that the conception of moral agency embodied in our norms and practices not only assumes that we regard ourselves and others as reason-responders capable of this kind of reflective self-monitoring, but also rightly holds us responsible for cultivating habits and skills of reflective self-monitoring and for exercising these skills. This does not mean that we always succeed – quite evidently we don't always do so.

I want to make two final points. First, cultivating and exercising the skills of reflective self-monitoring requires extensive social scaffolding not just when we are learning these skills but in order to maintain them. To give just a few examples, we often need the assistance and encouragement of others to devise and maintain strategies for exercising regulative self-control; we rely on others' responses, and on broader social norms, values and institutions, to gauge whether our emotional evaluations are reasonable and appropriate; we engage in social debate and discussion concerning the adequacy of norms of appropriateness for particular emotions; and through art, literature, film, or interaction with people from other social, ethnic or racial groups, we can expand our emotional repertoires and imaginative horizons, and challenge automatic or habitual patterns of emotional response, such as racist attitudes. Second, reflective self-monitoring as I have characterised it, is quite different from the kind of conscious reasoning from universal principles that Haidt and Greene think of as paradigmatic moral reasoning. I agree with them that this kind of conscious reasoning plays a less central role in moral agency than some rationalist models suggest. However, reflective self-monitoring and moral deliberation more generally may implicitly be guided by principles, and is implicitly or explicitly guided by our values, ideals and self-concepts, as I hope to have made clear through my examples. Haidt seems to think that trying to bring our emotional responses and behaviour into line with our self-concepts, values and moral commitments is a species of rationalisation. However, I would argue that this quest for normative coherence is evidence of the reflective capacities that are necessary for moral agency.

REFERENCES

Ronald de Sousa (1987), *The Rationality of Emotion*, Cambridge, Mass.: MIT Press

Cordelia Fine (2006), ‘Is the Emotional Dog Wagging Its Rational Tail, Or Chasing It?’, *Philosophical Explorations*, 9(1): 81-98

J. Greene, R.B. Somerville, L.D. Nystrom, J.M. Darley & J.D. Cohen (2001), ‘An fMRI investigation of emotional engagement in moral judgment’, *Science*, 293: 2105-8

Joshua Greene & Jonathon Haidt (2002), ‘How (and where) does moral judgment work?’, *Trends in Cognitive Sciences*, 6(12): 517-523

Joshua Greene (2008), ‘The Secret Joke of Kant’s Soul’, in Walter Sinnott-Armstrong (ed), *Moral Psychology*, vol. 3, Cambridge, Mass.: MIT Press: 35-79

Patricia Greenspan (2003), ‘Emotions, Rationality and Mind/Body’, in Anthony Hatzimoysis (ed), *Philosophy and the Emotions*, Cambridge: Cambridge University Press: 113-125

Jonathon Haidt (2001),‘The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment’, *Psychological Review*, 108 (4): 814-834

Jonathon Haidt & Fredrik Bjorklund (2008), ‘Social Intuitionists Answer Six Questions about Moral Psychology’, in Walter Sinnott-Armstrong (ed), *Moral Psychology*, vol. 2, 2008, MIT Press: 181-217

Jonathon Haidt & Fredrik Bjorklund (2008a), ‘Social Intuitionists Reason, In Conversation: Reply to Jacobson and Narvaez’, in Walter Sinnott-Armstrong (ed), *Moral Psychology*, vol. 2, 2008, MIT Press: 241-254

Daniel Jacobson (2008), ‘Does Social Intuitionism Flatter Morality or Challenge It?, in Walter Sinnott-Armstrong (ed), *Moral Psychology*, vol. 2, MIT Press: 219-232

Daniel Jacobson & Justin D’Arms (2003), ‘The significance of recalcitrant emotion (or anti-quasijudgmentalism)’ in Anthony Hatzimoysis (ed), *Philosophy and the Emotions*, Cambridge: Cambridge University Press: 127-145

Karen Jones (2003), ‘Emotion, Weakness of Will, and the Normative Conception of Agency’, in Anthony Hatzimoysis (ed), *Philosophy and the Emotions*, Cambridge: Cambridge University Press: 181-200

Neil Levy (2007), *Neuroethics*, Cambridge: Cambridge University Press

Jorge Moll & Ricardo de Oliviera-Souza (2007), ‘Moral judgments, emotions and the utilitarian brain’ *Trends in Cognitive Sciences* 11(8): 319-321.

Jorge Moll, Ricardo de Oliviera-Souza, Ronald Zahn, and Jordan Grafman (2008), ‘The Cognitive Neuroscience of Moral Emotions’, in Walter Sinnott-Armstrong (ed), *Moral Psychology*, vol. 3, 2008, MIT Press: 1-17

Darcia Narvaez (2008), ‘The Social Intuitionist Model: Some Counter-Intuition’, in Walter Sinnott-Armstrong (ed), *Moral Psychology*, vol. 2, 2008, MIT Press: 233-240

David A. Pizarro (2000), ‘Nothing More than Feelings? The Role of Emotions in Moral Judgment’, *Journal for the Theory of Social Behavior*, 30:4: 355-375

David A. Pizarro & Paul Bloom (2003), ‘The Intelligence of the Moral Intuitions: Comment on Haidt (2001), *Psychological Review*, 110 (1): 193-196

Peter Singer (2005), ‘Ethics and Intuitions’, *Journal of Ethics* 9: 331-352