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

Living with one's choices: Moral reasoning *in vitro* and *in vivo*.

Jeanette Kennett

Much of the recent research on moral judgment in the social and cognitive sciences focuses on subjects' responses to vignettes which impose a forced choice (approve/disapprove, appropriate/inappropriate) on them or which present them with highly unusual or disgusting scenarios – for example various versions of the trolley problem or sex with animals. No doubt many useful things can be learned from such studies- such as the ways in which moral opinions co-vary with socio-economic status – but it is at least not clear that they shed as much light as is claimed on the cognitive processes involved in moral reasoning - and so on implications for meta-ethical debates - in part because they do not and perhaps cannot take account of the cross temporal aspects of moral reasoning in everyday life. The moral verdicts subjects give in response to these vignettes are not decisions they must live with and for which they are accountable. Our moral choices have histories as well as consequences, and many of the most important moral decisions we make are not the work of a moment. I explore some representative cases of moral reasoning and revision across time and consider the implications.

Introduction: Moral judgment in vivo

During the 2008 election campaign Barack Obama and John McCain were asked what they regarded as their greatest moral failure. What was it that they most regretted in their lives? Obama named his teenage drug and alcohol use which he felt showed a disregard of others. I was more taken however with McCain's response. He named the failure of his first marriage.

Eleven years ago my then husband and I separated after a long marriage. It wasn't an easy thing to do – we had four children – but by the time we reached that point my defence for what I did indeed regard as a morally freighted course of action was one of necessity. I simply couldn't go on. The separation was a blessed relief and for a number of years I marvelled at the fact that I never for one moment missed the company of the man with whom I'd spent over 20 years of my life.

Readers will be relieved to find that I have no intention of regaling them with the detailed story of my marriage but it is true to say that when we split, despite publicly paying a bit of lip service to the mantra that there were faults on both sides, we were each much more

keenly aware of the other's deficiencies than we were of our own. I laid the blame for the breakdown of the marriage with him (if only he had….) while angrily rejecting his complaints about me. The failure of the marriage was a moral disaster certainly, but he was largely responsible for it. I was right, he was wrong. And it wasn't hard to find evidence to support my view. I replayed countless examples of his insensitivity, selfishness, unavailability etc. Meanwhile we each set about convincing ourselves that though the end of the marriage was not ideal for our children it was better than if we had remained unhappily married. In the course of time we each re-partnered with people much better suited to us in interests and dispositions and our children appear to have forgiven us. We now have a very cordial relationship – over the last few years we've even done Christmas Day together to save the kids (now adults) from having to divide their time between celebrations. So you might say there has been a happy ending and my original decision has been vindicated.

But that's not how it seems to me. Over the years I've revisited the scene of my marriage countless times and I now view many of the events I've reflected on quite differently. Now that the anger and resentment have dissipated I see the justice of many of my husband's criticisms of me and the reasonableness of his aspirations for our relationship (I'm not suggesting mind you that he's off the hook completely – just that I now appreciate what I then paid lip service to). By the time the marriage ended it was beyond saving but I can't comfort myself now, as I did then, that the split was always inevitable. It wasn't. If we'd addressed our problems – problems we were well aware of – earlier, and with greater commitment, goodwill, humility, and intelligence, we'd probably still be together. And I have no doubt that that would have been better for our children and would have avoided or ameliorated some of the difficulties they've faced in their lives. They bore the brunt of decisions they had no say in. I can't unequivocally say that I *wish* we'd worked harder at our marriage – that would mean wishing away my present relationship and the person I've become within it – but I can say that we *should* have. It's a part of what we owed to our children. Our failure to do so was a significant moral failure and is a source of guilt and regret to me at least.

I take it this story is not unusual though others may arrive at different moral conclusions about their particular situation. Whether to end or stay in a marriage, whether to give a child diagnosed with ADHD Ritalin, whether to put an elderly parent into a nursing home or care for them oneself, whether to accept a job offer in another city and move children away from their school and their friends, – these are decisions with significant moral dimensions and they are the kinds of decisions most of us will face. They are not usually snap decisions. We tend to spend a lot of time thinking about them and canvassing the options before deciding what to do and we often engage in a process of re-evaluation and revision after they are made. The past is a country we often return to and on each visit we find something new.

I tell this rather ordinary tale and suggest other equally ordinary tales as a backdrop to an examination of some well known recent empirical work on moral judgment carried out by Jonathan Haidt. I urge you to keep cases like this in mind as we proceed.

Haidt and Social Intutionism

Recent research into moral judgment in the cognitive and social sciences has focused largely upon subjects' responses to hypothetical moral dilemmas or morally charged situations and has sought by this means to elucidate the cognitive processes engaged in moral judgment. Examples of this research include Marc Hauser's Moral Sense Test (REF) and Jonathan Haidt and colleagues' work on disgust and on moral dumbfounding. Haidt has argued on the basis of a number of studies that "moral judgment is caused by quick moral intuitions and is followed (when needed) by slow, ex post facto moral reasoning' (Haidt 2001 p.817 see also Haidt and Bjorklund 2008a, p181). Haidt defines moral *judgments* as "evaluations (good vs bad) of the actions or character of a person that are made with respect to a set of virtues that are held to be obligatory by a culture or sub-culture" and moral *reasoning* as "conscious mental activity that consists of transforming given information about people in order to reach a moral judgment…the process is intentional effortful and controllable" (2001:817).

The words ‘intuition’ and ‘reasoning’ are intended to capture the contrast between two kinds of cognition – fast, automatic, affectively charged processes versus slow, effortful, controlled, conscious processing. Haidt’s work assumes this dual process model of cognition and is concerned to reject the ‘causal role of reflective conscious reasoning’ in moral judgment (2001:817) He argues that reasoning is rarely the cause of moral judgment, rather, “most of the action is in the intuitive process”. (2001:819). Haidt and Bjorklund (2008a) note a variety of evidence that suggests our everyday reasoning is a biased search “only for reasons that support one’s already favoured hypothesis”. Their argument is that we are misled by the facility with which people generate justifications for their moral judgments into thinking that the reasoning process is a cause of the judgment itself.

Haidt’s work has been of great interest to philosophers working in meta-ethics and moral psychology and has usually been taken to support non-cognitivism (emotivism or simple sentimentalism) accounts of moral judgment over their rationalist counterparts. As has been pointed out however (e.g., Jacobsen 2008, Kennett & Fine 2009) simply demonstrating that many or even most moral judgments are the product of automatic processing, or that reasoning processes themselves are subject to systematic biases, does not undermine rationalist claims about the role of reason in moral judgment, since these claims are conceptual and normative rather than descriptive. Nevertheless it would be a blow to rationalism, or indeed to any account of moral judgment which stresses a conceptual connection between moral judgment and *justification*, if reason *could not* play the role assigned to it in the theory – if our particular moral judgments and the moral intuitions upon which they often rely were cognitively impenetrable

Haidt does not go quite this far. Three of the links in his Social Intuitionist Model (SIM) involve conscious processing. Link 3, the Reasoned Persuasion Link, focuses on social striving to reach consensus on normative issues. In this link people offer each other reasons, but notably Haidt and Bjorklund claim that these “are best seen as attempts to trigger the right intuitions in others” In a piece of rhetoric against female circumcision

they cite as an example "each argument is really an attempt to frame the issue so as to push an emotional button...[r]hetoric is the art of pushing the ever-evaluating mind over to the side the speaker wants it to be on and affective flashes do most of the pushing" (2008a:192). The so-called Reasoned Persuasion Link is thus readily analysable in terms of the early emotivist position put by Ayer (1036) and Stevenson (1937) which held that the twin functions of moral judgment/discourse were to express one's feelings and influence the feelings of others. Links 5 and 6 of the model offer a little more to the empirically minded rationalist. Link 5, the Reasoned Judgment Link, acknowledges that "people may at times reason their way to a judgment by sheer force of logic, overriding their initial intuition' and acknowledges that '...in such cases reasoning truly is causal...[h]owever such reasoning is hypothesized to be rare, occurring primarily in cases in which the initial intuition is weak and processing capacity is high" and the intuitive judgment persists below the surface. (193-4). Link 6, the Private Reflection Link occurs when a person spontaneously generates a new intuition that conflicts with their initial intuitive judgment. The reflective process then, according to Haidt and Bjorklund, may involve weighing pros and cons or applying a rule or principle to the situation. However they insist that "all cases of moral reasoning probably involve a great deal of intuitive processing" (195) by which they appear to mean processing that is opaque to and largely impenetrable by, reason. In moral dilemmas the person ultimately decides based on "a feeling of rightness, rather than a deduction of some kind". Haidt suggests that only psychopaths and philosophers are honestly able to examine emotive issues dispassionately and this may not be cause for celebration since there is plenty of evidence (e.g., Damasio's work on 'acquired sociopathy') to suggest that "reasoning stripped of affective input becomes inept" (2008a:195)

The most general claim of the SIM then is that "the action in morality is in the intuitions, not in reasoning". (2008a:196) This is the claim that I will probe. In this paper I want to examine the strategy adopted by Haidt and colleagues to reveal, as they see it, the post hoc nature of (most) moral reasoning and focus upon their interpretations of their own data. I will suggest that their interpretations are not well supported by the data they present and that part of the problem is generated by the nature of the experimental

situation which does not model real life moral judgment and decision making. I will then provide an alternative account of the role which reasoning can play in governing our particular moral judgments, including many, and perhaps most, of our fast intuitive judgments. If I'm right then the meta-ethical conclusions which Haidt and others have drawn from his work will not stand even by their own lights.

Moral Dumbfounding and Moral Judgment

Haidt has argued that standard moral judgment interviews give a misleading impression of the role of reason in moral judgment because 'they create an unnaturally reasoned form of moral judgment' (2001:820) While according to the SIM model most of the action in moral judgment is to be found in links 1-4, (Intuitive Judgment, Post Hoc Reasoning, Reasoned Persuasion and Social Persuasion) "…if the person talking to you is a stranger (a research psychologist) who challenges your judgment at every turn…then you will be forced to engage in extensive, effortful, verbal central processing. ...leading to the erroneous conclusion that moral judgment is primarily a reasoning process" (2001:820) Haidt, Bjorklund, and Murphy (2000) found that while Kohlberg's Heinz dilemma (should Heinz steal the expensive drug to save his wife's life?) did indeed elicit moral reasoning and that people confronted with the dilemma were somewhat responsive to counter arguments, this was not the case when people were presented with "harmless taboo violations" designed to cause an immediate, strong, affective reaction, namely consensual adult sibling incest and harmless cannibalism of an unclaimed corpse. In the case of the taboo violations subjects' incapacity to justify their initial judgments did not cause them to change their minds even though they confessed they could not explain the reasons for those judgments – a phenomenon Haidt et al label 'moral dumbfounding'. Thus Haidt et al conclude that the dumbfounding evidence supports the SIM: reasoning is only effective when the initial intuition is weak.

Haidt's criticisms if Kohlberg's methodology show that he is alive to the fact that work done in the laboratory might not accurately model what goes on in more naturalistic

settings and so might not provide a guide to people's everyday moral judgments or standard modes of reasoning. Curiously he does not apply the same level of critical scrutiny to his own experimental methods. I too am concerned that research in the laboratory on moral judgment gives a misleading picture of ordinary moral judgment and the role of reasoning. I also agree with Haidt that in some cases formal moral judgment interviews elicit more in the way of explicit reasoning than we would find from the same persons in vivo. But before airing some concerns about Haidt's own research methods, let's consider what there is to be said for the moral dumbfounding method used by Haidt which he claims vindicates the SIM and in turn supports a Humean sentimentalist picture of moral psychology and moral judgment over rationalist alternatives. To do so let us look more closely at one of the examples presented to subjects – that of consensual adult sibling incest.

> Julie and Mark are brother and sister. They are travelling together in France on a summer vacation from college. One night they are staying alone in a cabin near the beach. They decide it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills but Mark uses a condom just to be safe. They both enjoy making love but decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love? (Haidt 2001 p.814 from Haidt, Bjorklund and Murphy, 2000)

Subjects confronted with this story immediately said that it was not OK for Mark and Julie to make love. When asked to justify their responses they cited the possibility of deformed children or harm to Mark and Julie's relationship. When it was pointed out to them that these bad consequences were ruled out by the story they showed no disposition to withdraw their initial assessment even though they admitted they could not explain it.

This is the state Haidt dubs ‘moral dumbfounding’. Haidt thus concludes that moral reasoning is largely idle:

> Moral reasoning is usually an ex post facto process used to influence the intuitions (and hence judgement) of *other* people.[my emphasis] In the social intuitionist model one feels a quick flash of revulsion at the thought of incest and one knows intuitively that something is wrong. Then when faced with a social demand for a verbal justification, one becomes a lawyer trying to build a case rather than a judge searching for truth. One outs forward argument after argument never wavering in the conviction that Julie and Mark were wrong even after one’s last argument has been shot down (814)

Why use such confronting and unusual cases as incest, cannibalism, or masturbation with chicken carcasses to test the causal powers of moral reasoning? First the experimenter needs to elicit strong unopposed intuitions since these provide the clearest test case of the causal powers of explicit moral reasoning. Haidt allows that in rare cases where initial intuitions are weak reasoning may be ‘genuinely causal’. Moreover where there are conflicting intuitions (as may be the case in the Heinz dilemma) it will not be clear that reason alone is playing a role in overcoming the initial judgment. Second if we provide familiar or straightforward cases e.g., killing a relative for personal gain, sexual predation on young children, gambling away the rent, then it’s highly likely that the reasons subjects provide for their judgments would in fact support and justify those judgments, even if they have played no causal role.[i] We need extreme examples from which the standard justifications have been cleverly removed. If subjects were found to change their views and override their initial intuitions in cases of these ‘harmless taboo violations’, this would be evidence of the causal power of reasoning. That Haidt’s subjects did not change their views when their justifications were removed allegedly exposes the ex post facto nature of most moral reasoning.

But is Haidt's conclusion supported by the moral dumbfounding studies, or are there other, better, explanations of the subjects' responses?

Let's accept Haidt's claims that: a) much moral judgment is fast, automatic, intuitive and not directly preceded or caused by explicit reflection, and: b) that we often engage in ex post facto rationalisation of our judgments or actions. If all that is claimed by the SIM is that moral reasoning is most often not the *proximate* cause of individual moral judgments, then this is very likely true. However this does not provide a sufficient basis for the meta-ethical conclusions that Haidt and others wish to draw from his data. In particular it does not vindicate non-cognitivist accounts of moral judgment, such as simple sentimentalism or expressivism, over rationalism or neo sentimentalist accounts, which require reflective endorsement of the deliverances of affect and intuition.

My claim is that doesn't follow from the fact that the majority of our particular judgments are not preceded by a bout of explicit effortful reasoning that moral reasoning is mostly window dressing undertaken out of social motives. Much of what we see in the lab– both in the dumbfounding experiments and in work by Hauser and colleagues, and Greene and colleagues, testing responses to trolley problems and other moral dilemmas – doesn't give an accurate picture of the role of reason and reflection in everyday moral judgment and decision. I argue that these experiments aren't capable of measuring the role of reflection in the exercise of moral agency in judgment and decision, in significant part because they only seek and thus can only measure the proximate and not the distal causes of moral judgment. Before turning to this I highlight some other factors which should be taken into account in interpreting the data.

Moral judgment in vitro: reasons for caution

What experimental artefacts might affect the results of research into moral judgment and so the conclusions which are drawn from them? My focus here is on research which proceeds by presenting subjects with short vignettes and asks either for a moral verdict on the behaviour described or a judgment on the options presented in the vignette. As well

as the moral dumbfounding experiments described above, other researchers (Hauser et al, Greene et al) have presented subjects with a variety of dilemmatic situations in which they must decide whether it is right or permissible to sacrifice one life to save several others. The best known of these are the many variants on the trolley problem. In the original version a bystander may pull the lever to direct a runaway trolley which is hurtling towards five workers on a railway track down a disused side track instead where it will certainly kill one person who is walking along it. In another version the bystander may push a heavy man off a footbridge into the path of the trolley, the heavy man will die but the trolley will be stopped before it reaches the five workers. What should the bystander do? Trolley problems have been much discussed in the philosophical literature. I will highlight some shared features of the vignettes used in moral judgment studies which suggest that we should be cautious in drawing meta-ethical conclusions from them. I also draw attention to a particular feature of the moral dumbfounding studies which may also limit the usefulness of the results.

Third personal, hypothetical judgments

First the vignettes elicit only a subset of moral judgments – those that require verdicts on the actions of (unknown) third parties in hypothetical situations. They do not probe first personal moral judgments or verdicts about one's own moral choices. While we do regularly make judgments about other parties it is not altogether clear that we can generalise from evidence about the cognitive underpinnings of these judgments – especially those made in response to highly artificial or unlikely scenarios (to be addressed next) – to first person online moral judgments and decisions that will inevitably be subject to a much more extensive and subtle range of informational inputs, as well as a range of practical constraints[ii]. What we say in the lab in response to a vignette might not match what we would judge or do if placed in the situation described. Haidt and Bjorklund acknowledge this limitation in their response to commentators (2008b). There they distinguish between moral *judgments* which are "about whether *another person* did something right versus wrong" (my emphasis p.243) and moral *decision making*, and they stipulate that the SIM describes the phenomenology and causal processes of (third personal) moral judgment only. They argue that moral judgment is

functionally distinct from moral decision-making and adopt the suggestion that it would have been shaped by different selection pressures. They difference as they see it is that in moral judgment, as opposed to moral decision, there is very little at stake for the self. "We can make a hundred judgments each day and experience them as little more than a few words of praise or blame, linked to flashes of feeling, that dart through consciousness. But moral...decisions are different: they have real consequences for the self and others" (243). They concede that moral decisions must take into account many factors besides the initial intuitive response.

Haidt and Bjorklund may be correct in making a functional distinction between third personal judgments and practical decisions in this way but they seem not to notice that such a distinction, and the consequent restriction of the scope of the SIM to third personal moral judgment, limits both the interest and explanatory power of their model and seriously undercuts the Humean meta-ethical conclusions they and others wish to draw from their work. First and second personal moral judgments are not counted though they would ordinarily be considered to be primary instances of moral judgment; moreover two important conceptual features of moral judgments, their universality and their authority, which are used in both the philosophical and social psychology to distinguish them from non-moral judgments, are ignored. Yet these features help to explain the link between judgment and decision-making. The universality of moral judgment means that it applies to me too on pain of hypocrisy. Arguably, if I think that my judgment that a certain kind of action is impermissible doesn't apply to me in relevant circumstances, then I am not making a moral judgment at all.

Our moral judgments are supposed to underwrite our moral *decisions* If we find, at the moment of decision, that we cannot choose in accordance with our prior judgment, we are either straight-forwardly weak-willed, or we will have reason to revisit the judgment itself in the light of the additional inputs (and gut responses) experienced in the context of decision. So if Huck Finn finds that he cannot bring himself to turn Jim in to the authorities he might berate himself as weak or he might be prompted by his sympathies to revise his views on the morality of slavery.[iii] On rationalist accounts of the process he

could treat his reluctance as a kind of data that alerts him to the possibility of error in his original judgment. There is thus an internal process by which principles, intuition, judgment, and choice are brought into reflective equilibrium. On sentimentalist accounts too the functional dissociation between moral judgment and moral decision making now posited by Haidt and Bjorklund is significant, for it means that their evidence about the nature of third personal moral judgments, even if unchallenged, can provide little support for the Humean position they have claimed to vindicate. These "flashes of feeling" cannot explain what Prinz calls "the rapid move from thinking an action is wrong to thinking I ought to prevent or avoid that action." Moral judgments are apparently now not posited by Haidt to be the kind of thing that "vie for control of the will" such that " [w]hen they occur we are thereby motivated to act"(Prinz 2006:36) If anything Haidt and Bjorklund now assume a version of externalism that claims a merely contingent connection between moral judgment, decision, and action, mediated perhaps by social pressures.

Extreme or fantastic examples

The second limitation of the studies is their use of highly unusual, decontextualised, or artificial examples. Subjects must consider situations that they are unlikely ever to encounter in real life and choose among options that would probably never occur to them if they did. Take the footbridge version of the trolley problem. How likely is it, as you become aware of the impending tragedy below you, that the option of using the stranger next to you to stop the train would occur to you? How likely is it, at the moment of decision that you would know all the consequences of your action? The experimental scenarios stipulate what in real life may be very unclear and then impose a forced choice on the subject, and subjects may resist either the scenario or the choices offered.

I suspect that what I will call *cognitive resistance* plays a role in responses to many of the highly artificial, stylised scenarios and dilemmas beloved by both philosophers and moral judgment researchers. Consider the Mark and Julie case. A quick internet search indicates that even sibling incest overwhelmingly occurs in circumstances of significant family dysfunction and abuse and involves power imbalances and long term damage to

relationships and individuals. (I exclude here exploratory play between young children). This fits with what I take to be the lay understanding of incest. Yet the story presented to the subjects implicitly invites them to think of Mark and Julie as normal, functioning, happy siblings from a well off family (they are college students holidaying in France). But how then did the subject of having sex with each other even arise for them, let alone seem like a feasible and inviting option for passing the evening together. What is the context that could possibly explain such a conversation? Why would they engage in behaviour that anyone could reasonably suspect would be damaging to that relationship and possibly harmful in other ways as well? At the very least having sex with one's sibling, given the taboo on incest, seems like a profoundly morally risky thing to do and their decision to do so was not justified by any other considerations (boredom perhaps) that were available to them. On these grounds alone they would be well advised not to proceed. How could Mark and Julie be so sure they would not feel shame and disgust when reflecting on what they did? What kind of people would casually contemplate a one night stand with their sibling as providing 'at the very least a new experience', and remain unaffected by engaging in it? The story of Mark and Julie just doesn't ring true.

Is there any evidence that these scenarios encounter cognitive resistance? Anecdotally, there is, at least in the philosophy classroom. Those of us who teach philosophy find plenty of resistance to the use of fanciful or impossible examples; it can take years to get students to see their methodological usefulness – a usefulness that is increasingly questioned in the field of moral philosophy. Students frequently challenge the supposed omniscience of the protagonist in the moral examples and seem to think that this matters for moral decision making. They say things like: 'but you couldn't know that'. They try to fill out thin stories, reject bits they find unbelievable, ask for more detail, and look for third options if they don't like those on offer. But does this carry over to the lab? Interestingly Haidt et al tell us that his subjects continued to argue "that Julie and Mark will be hurt, even though the story makes it clear that no harm befell them".(614) This is taken by them as further evidence for the post hoc reasoning thesis. I think it is, rather, evidence that the participants don't buy the story (or don't buy the very limited account of harm assumed by the experimenters, or don't buy the assumption implicit in the study

design that utility is all that matters morally). Their clear resistance to the claim of harmlessness suggests that their responses and justifications might be tracking their estimation of the probable impact of incest on siblings in the actual world rather than in the scenario as given. For these nearby cases their responses may be justified. When their responses are disallowed by the experimenter, rather than explored, they are, unsurprisingly, left with nothing to say. Haidt et al may have succeeded at inducing a state of dumbfounding in their subjects but given that it is likely engendered at least in part by disbelief it does not unambiguously support the post hoc reasoning hypothesis.

Experimenter effects

Third, it is also likely that Haidt's subjects in particular are disempowered by the experimental situation. In other words, they lack: a) the skill to play the game and the ability to see the point of it – a skill we do indeed often find in philosophers where we play the game for certain theoretical purposes e.g., to reveal the structure of normative theories, and b) the confidence or motivation to outright reject the scenario and say to the experimenter – 'But it wouldn't be like that. How could their relationship not be affected? What about when they meet and fall in love with other people? The secret could become a terrible burden'. Moreover they almost certainly lack the skill to identify and challenge the utilitarian assumptions in the scenario with which they may inchoately be disagreeing. It is surprising that Haidt and Bjorklund who cite Milgram's studies as an example of the power of the situation to induce 'obedience without persuasion' (2008a:192) do not acknowledge the possible influence of the experimental situation – the authority of the experimenter and the desire of the subject to cooperate with the experimenter, not to appear rude, or even to get course credit for participation – on their results.[iv]

Bias towards the proximate cause of judgment

Fourth, the speed and automaticity of moral responses to many (but not all) of the vignettes given in the laboratory and also to many moral situations encountered in daily life can only show that reflective processes were not engaged at that time. It does not and cannot show that moral reflection is idle since the research does not distinguish between

proximate and more distant causes of moral judgment. Work in the laboratory will tend to pick up on the proximate cause so the results will be biased against both rationalist accounts of morality and neo-sentimentalist accounts which also emphasise the process of reasoned reflection and justification in refining and expanding our moral sensibilities and schooling our intuitions.

Recovering the role of reflection in a manner consistent with the data requires

- an understanding of the processes by which reasoned judgments become automatized over time
- attention to the subject matter of moral thought
- attention to the history of moral judgments and their cross temporal aspect
- acknowledgment of the full range of inputs to reflection

It is to these issues that I now turn.

Moral reasoning and cross temporal moral guidance

My claim is that it doesn't follow from the fact that the majority of our particular judgments are not immediately preceded by a bout of explicit effortful reasoning that moral reasoning is mostly window dressing. In order to see this we need to distinguish between clear cut, core moral judgments and the ways in which they guide us, and more difficult, nuanced and complex judgments (such as those we began with).

Reasoning and core moral judgments

Core moral judgments include those based on simple straight forward and uncontentious rules about physical harm, cheating, and fairness, which we learn as children and in general do come to reflectively endorse. Moral education involves both empathy induction (Hoffmann) and reason giving. The child must come to see the point of core moral rules and indeed the reason giving practices of children indicate that they are active participants in this process providing increasingly differentiated evaluations of various violations. (Pool et al 1983) However, as others have pointed out, while the acquisition

of any skill (eg., driving) may initially require a lot of cognitive effort, once a skill is mastered it becomes largely automatised. (Salzberg & Kasachkoff 2004 and implicitly acknowledged in Haidt & Bjorklund 2008a). Many of our moral judgments could be like this. In support of this idea Fine (2006) reviews a variety of studies which indicate that "at least some automatic processes reflect the action of prior controlled processes". She argues that the repeated explicit selection of a goal leads to its being triggered automatically by eliciting situations. For example Fishbach et al (2003) found that for successful dieters, temptation stimuli led to automatic activation of their goal of staying slim, and Moskowitz et al (1999) found that subjects with egalitarian goals exerted pre-conscious control over stereotype activation. Fine concludes that "the SIM overlooks some of the important subtlety of how ...some automatic processes arise." (93)

In the realm of moral judgment it is surely true that we don't waste time wondering whether and how core moral rules and principles apply in straightforward cases. Experienced moral agents don't need to expend conscious effort to judge that it's impermissible to hit someone over the head and steal their wallet, just as experienced drivers don't need to expend conscious effort on turning the steering wheel in the direction they want to go. Ingrained moral rules or principles can give rise to fast automatic responses in situations where they apply. These responses can present themselves as strong intuitions which may be resistant to challenge. Does this tell us anything deep about moral reasoning? In particular does it tell us, as Haidt and colleagues argue, that reasoning is almost always post hoc and plays no role in moral judgment itself? Surely not. At best it indicates that in easy cases or where we are required to make a quick decision, our decision will be governed by automated responses – hardly a surprise.

It is not difficult for those who endorse rationalist views of moral judgment and agency to offer an alternative explanation of such ingrained moral responses that is consistent with an account in which reason plays a significant causal role. To do so one needs to acknowledge the cross-temporal aspects of reasoning that are substantially ignored by moral judgment research . Consider this summary of Michael Bratman's account of the

relation between planning and decision: “A rational intentional action ...is one which is part of a plan ... that is rational for an agent to adopt and not irrational for her to fail to reconsider. In that way, first-order desires that are not reflectively considered at the time of action are nevertheless rational ...ongoing or recurring reflection is not a plausible requirement for rational action. Most action involves habit and automatic response that not only fails to involve reflection, it sometimes precludes it.” (Christman, 2008:153). On this kind of account reason may guide action *at a distance* via the reflectively endorsed establishment of habit and automatic response in accordance with our principles, plans, and goals. This philosophical view of agency is in line with the evidence from social psychology referred to above.

Moral revision and unusual or difficult cases

Haidt and colleagues might protest at this point that the moral dumbfounding cases they present challenge the subject to reconsider their automatic moral responses but it is rare for subjects to do so even when the justifications they rely on have been shown not to apply to the case at hand. These are cases where it *is* irrational, by their own lights, for subjects to fail to reconsider and they demonstrate that reason has precious little to do with driving moral judgment.

There are two responses available to the defender of the role of reason in moral judgment here. The first is similar to that made by the indirect utilitarian to charges that her utilitarianism is compromised by her endorsement of the following of rules or policies or the inculcation of dispositions, even in situations where following those rules, policies or expressing those dispositions may not bring about the best consequences. It accepts that there may be extreme and unusual circumstances, such as those presented in experimental vignettes, where actions which are normally very wrong might be justified. But it seems like a good thing that we are intuitively averse to such actions even in exceptional circumstance. It might not be irrational for someone to resist revising their core moral judgments *considered as policy* with respect to, for example, the killing of innocent strangers, incest, or torture, in the face of an exceptional or fantastic scenario. Overall we

will do better by adopting policies by which we fashion ourselves into the kinds of people for whom committing incest, pushing people off bridges, or cutting up healthy people for spare parts, are not even options.

The second response connects with the previous discussion of the Mark and Julie scenario, which identified reasons for concern over Mark and Julie's conduct that were not neutralised within the story. Karen Jones (2006) argues that ordinary subjects tacitly take their intuitive moral responses to be *tracking* reasons which once articulated would justify their moral judgments. That they cannot presently 'put their finger' on the reason may not rationally compel them to the view that no such reason exists, for as Jones points out, experience teaches us that sometimes "emotions can key us to the presence of real and important reason-giving considerations" even though it is only later that we can reflectively access and articulate those reasons.

Of course we are sometimes mistaken in our tacit assumption that our intuitive responses are reason tracking; there is certainly evidence that a person's incidental mood or emotional state can 'contaminate' her moral judgments (e.g., Forgas & Moylan, 1987; Wheatley & Haidt, 2005). However there is also evidence that this bias can be corrected more or less accurately (see Wilson & Brekke, 1994), when the individual's attention is drawn to their mood as a possible source of bias (e.g., Schwartz & Clore, 1983)[v]. Data such as this suggests that not only do we tacitly *assume* that our moral judgments are responsive to reasons, (something Haidt et al could concede) they are, often enough, so responsive. We will revise them if and when we become convinced that our gut reactions are, in a particular case, irrelevant to the issue at hand or, though relevant not decisive.

According to the SIM, revision of a moral judgment will usually involve developing a competing intuition as a result of social pressure (and this is certainly one path to changing one's mind). Haidt thinks private reasoning only rarely plays a role in this process. However, Fine points out that the SIM's prediction that private moral reasoning is rare and that in the absence of social pressure people simply engage in post hoc justification of automatic moral attitudes, does not sit well with the finding that "the vast

majority of individuals (over 90%) report discrepancies between their privately experienced 'should' and 'would' responses to stereotyped groups" (Fine 2006:94) If moral intuitions lead directly to moral judgment or constitute the moral judgment we would not expect such discrepancy between gut reactions to e.g., homosexual practices or racial minorities and one's judgments about how one should respond. Evidence cited by Kennett & Fine (2009) of individuals explicitly discounting their intuitive responses to homosexuality, further suggests that reasoning, both privately and with others, may often enough override intuition. [vi]

But what do we mean by 'reasoning' here? The dual processing model of cognition characterizes reasoning in terms of "abstract thinking and high level cognitive control" (Greene, 2007 p398) As we've seen this is the model presupposed by the SIM. As Haidt has put it controlled processing is…"a tool used by the mind to obtain and process information about events in the world or relations among objects". This view of reason is clear in Greene's interpretation of subject's responses to 'The Crying Baby Dilemma'. (Greene et al 2004)

> It's war time, and you are hiding in a basement with several other people. The enemy soldiers are outside. Your baby starts to cry loudly, and if nothing is done the soldiers will find you and kill you, your baby, and everyone else in the basement. The only way to prevent this from happening is to cover your baby's mouth, but if you do this the baby will smother to death. Is it morally permissible to do this?

In this dilemma participants 'answer slowly and exhibit no consensus' indicating according to Greene that negative social-emotional responses compete with a strong "cognitive" case. The cognitive case involves explicit reasoning about the consequences: it involves the 'processing of information about the world' to reach the utilitarian conclusion, whereas the non-utilitarian conclusion is on this account driven by affect.

I rather doubt that this interpretation does justice to the case. Greene's couching of the dilemma as a simple competition between the two processes with one eventually proving

dominant, sells the reflective process short. What's notable about the Crying Baby case and what distinguishes it from many other dilemmas used in moral judgment research is:

1) It is first personal. The subjects are not being asked to make a judgment on the actions of an unknown third party
2) It is a morally difficult case. It is no small thing to contemplate killing your innocent and defenceless infant even if the child will almost certainly die anyway. Consequentialist considerations are weighty, but they are not the only moral considerations at stake. Williams (1973) canvasses some of these in his discussion of negative responsibility and integrity. [vii] And while Greene has suggested that the 'up close and personal' factor pushes the social-emotional response in many moral dilemmas I doubt that this plays a big role here. Killing one's child remotely by flicking a switch is not obviously less morally problematic than killing her by smothering her.
3) The scenario, while not one that most people are likely to personally encounter, is realistic. Participants are unlikely to experience cognitive resistance to the vignette. Rather they are able to *imaginatively occupy the scenario* and use this to inform their responses.

Mental time travel, imagination, and moral reasoning

A consideration of the Crying Baby case points to a lacuna in the accounts of moral judgment made available by dual processing models and drawn upon in interpreting moral cognition research. The role of the moral agent, the one who makes the judgment, is left out of the picture. This decoupling of moral judgment from agency is, I suggest, a mistake. Moral judgments must be made by moral agents. We will miss much of what is most interesting about moral judgment if we ignore its agential aspect. The reflective self-awareness that makes us agents capable of moral judgment, and of the regulation of our moral responses, requires the exercise of additional capacities not explicitly encompassed by dual processing theory or recognised by the SIM.

Real life moral choices, often, as Haidt now acknowledges, engage our sense of self. A crucial aspect of this sense of self is given by what is known as autonoetic awareness. This is "awareness of oneself as a continuous entity across time... (Levine et al 1998). Such temporally extended self-awareness seems to be a necessary condition of the kind of reflection – on the worth of possible goals, activities, and on the type of person we want to be – that provides us with *normative* reasons, including moral reasons, and so establishes us as agents capable of moral judgment. More generally *any* kind of agential planning requires the capacity to imaginatively project ourselves forwards and backwards in personal time, a capacity which has been dubbed mental time travel.

Mental time travel is a controlled activity which we undertake for the purpose of evaluating the past, choosing for the present, or planning for the future. In mental time travel the agent recalls and re-experiences episodes involving her past self, or imagines herself as taking part in some future episode. Mental time travel, then, includes what are sometimes called episodic, or personal memories, in the backward looking cases, and what is sometimes called prospection, in the forward looking cases. Episodic remembering is the familiar category of memory in which a person replays a past experience in which she was personally involved (Tulving 1972/1983). Prospection involves the simulation of future events in which we mentally rehearse a situation. Both memory and prospection are essential to agency and are intimately connected with planning and reflection and so, I claim, with the capacity for genuine moral judgment.

For example, in planning for this year's family Christmas dinner I might recall last years' disaster when the turkey took too long to cook with the result that the children got overtired and irritable and Uncle Ray got drunk and argued with Grandad. On the basis of my trip to the past I judge that things will go better this year if we eat earlier and limit alcohol and so I plan to get the turkey in the oven by 8.00am and to serve no alcohol until everyone is seated for dinner. Or, I might upon reflecting how my life is going, decide that living up to my principles requires that I do more than I have been doing to help the needy. As a result I might commit my future self in various ways: arranging for automatic

donations to charity from my pay, or volunteering my time at a charity for the homeless and making the required forward looking revisions to my busy schedule.

Because I see myself as a diachronic agent and my choices as interconnected, events that occur now can also prompt reflection on past choices and behaviour leading to reinterpretation of the past and revision of relevant judgments and principles. The examples given in this paper suggest that this kind of private reflection is not limited to an exotic minority of professional philosophers and the like or provoked only by unusual circumstances. It is the constant companion of many, if not most, of us as we move through our lives. Such reflection may be prompted directly by some consequence of past decisions but also by other trigger events such as illness, the birth of a child, or the ageing and death of a parent. Who has not revised their view of their parents and of their own behaviour as, say, a teenager, in the light of their own parenting experiences? To be sure such reflection might not often meet ideal standards of deliberation. Imagination may fall short, we may not adequately discount for cognitive biases, and the quality of our moral judgments may reflect these shortcomings in reasoning. But if we significantly lack these capacities for reflection our very status as moral agents is called into question.

The ongoing activities of planning, monitoring, judging good or bad, moral reflection and moral revision all require memory, imagination and prospection. They all require a diachronic conception of self and others. Without such cognitive resources an individual's verbal judgments of right and wrong would be so impoverished and unsupported as to seriously undermine any claim to be even minimally competent moral judges and interlocutors. Severe amnesics who lack autonoetic awareness and the capacity for both forward and backward mental time travel may retain some capacity for synchronic moral judgment since provided that their semantic memory is normal they will be able to apply a learned rule to a situation and may have normal affective responses to present pleasant or unpleasant stimuli. Clearly however they cannot count as full moral agents and do not meet the conditions for responsibility. They cannot reflect upon their behavior or revise their moral judgments and principles. Their concept of a reason and their capacity to both track and respond to reasons is minimal. To the extent

that they could count as unimpaired moral judges on dual processing accounts of the cognitive foundations of moral judgment this will surely be a problem for such accounts and for those meta-ethical positions which claim support from this picture of moral cognition. (Kennett & Matthews 2009, Gerrans and Kennett forthcoming)

Conclusion

Christine Korsgaard claims that it is "from the standpoint of practical reason that moral thought and moral concepts... are generated" (Korsgaard 1992:132). That is the conclusion to which this discussion has led. The processes of memory, imagination, projection and rehearsal described here can take as their objects our gut reactions, our more abstract moral principles, conflicts between them and much more besides. These processes can deliver justifications which are neither mere rationalizations of gut intuitions or effortful rule application. They enable us to respond to our reasons *as* reasons and so vindicate a rational reflective conception of moral agency. While we don't need to invoke these processes in many easy cases in which we are required to render a moral verdict we commonly do so in the more complex and ambiguous situations with which we began. Moral judgments can be made prospectively, synchronically, or retrospectively and plausibly they require an agent who can see things diachronically. Perhaps the SIM could be even further revised to incorporate a role for mental time travel and the reflection made possible by it, but I suggest that the distinctiveness of the SIM will then be lost, since it will be an impossible task to show that a mature agent's moral judgments, many of which will have been revisited and modified over a prolonged period, are the product solely or primarily of intuitive processes as Haidt wishes to understand them. This is why the thin, de-contextualised, encapsulated scenarios and the restricted, time limited, choices presented to subjects in moral judgment research may not tell us very much about the processes underlying our more interesting real life, interconnected, moral judgments and choices – choices we must live with.

References

Ayer A J (1936) *Language, Truth, and Logic*, London: Gollancz. (2nd. Edition, 1946.)

Bennett, J. (1974). The Conscience of Huckleberry Finn *Philosophy*, *49*, 123-34.

Bratman, M. (2000). 'Reflection, Planning, and Temporally Extended Agency'. Philosophical Review, 109:35-61.

Christman, (2008) in Kim Atkins and Catriona MacKenzie (eds) *Practical Identity and Narrative Agency* Routledge, New York.

Fine, C. (2006). Is the emotional dog wagging its rational tail, or chasing it? *Philosophical Explorations, 9,* 83-98.

Fishbach, A., R.S. Friedman, and A.W. Kruglanski. 2003. Leading us not into temptation: momentary allurements elicit overriding goal activation. *Journal of Personality and Social Psychology* 84: 296-309.

Forgas, J. P., & Moylan, S. J. (1987). After the movies: the effects of transient mood states on social judgments. *Personality and Social Psychology Bulletin, 13*, 478-489.

Gerrans P & Kennett J. Neurosentimentalism and Moral Agency (forthcoming in *Mind)*

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., Cohen, J.D. (2004) The neural bases of cognitive conflict and control in moral judgment. *Neuron*, Vol. 44, 389-400.

Greene, J.D. (2007) Why are VMPFC patients more utilitarian?: A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*. Vol 11, No. 8, 322-323.

Haidt, Bjorklund, and Murphy (2000)

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814-834.

Haidt, J., & Bjorkland, F. (2008a). Social intuitionists answer six questions about moral psychology. (In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and diversity* (pp. 181-218) Boston: MIT Press.)

Haidt, J., & Bjorkland, F. (2008b).
(In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and diversity* (pp. 181-218) Boston: MIT Press.)

Haidt, J., & Hersh, M.A. (2001). Sexual morality: the cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology, 31*, 191-221.

Hauser

Jacobsen D (2008) (In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and diversity* (pp. 181-218) Boston: MIT Press.)

Jones K (2006) Metaethics and emotions research: A response to Prinz *Philosophical Explorations* 9 45-54

Kennett J and Fine C (2009) ‘‘Would the real moral judgment please stand up? The implications of social intuitionist models of cognition for meta-ethics and moral psychology’, *Ethical Theory and Moral Practice* 12: 77-96

Kennett J & Matthews S (2009) Mental Time Travel, Agency and Responsibility’ (2009) In Matthew Broome and Lisa Bortolotti (eds) *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives* Oxford University Press

Korsgard C (2002) Internalism and the Sources of Normativity. In Herlinde Pauer-Studer (ed) *Constructions of Practical Reason: Interviews on Moral and Politicasl Philosophy*, Stanford: Stanford University Press

Levine, B., Black, S.E., Cabeza, R., Sinden, M., Mcintosh, A.R., Toth, J. P., Tulving, E., Stuss, D. T. (1998), “Episodic Memory and the Self in a Case of Isolated Retrograde Amnesia,” *Brain*, 121, 1951-1973.

Moskowitz, G. B., P.M. Gollwitzer, W. Wasel, and B. Schaal, B. 1999. Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology* 77: 167-84.

Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations, 9,* 29-43.

Salzburg and Kasachkoff (2004)

Stevenson C L. (1937) The Emotive Meaning of Ethical Terms, *Mind*, 14-31

Tulving, Endel (1972), “Episodic and Semantic Memory,” In *Organization of Memory*, ed. E. Tulving, W. Donaldson, New York: Academic, 381-403.

Tulving, Endel (1983), *Elements of Episodic Memory*, Oxford: Clarendon.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science, 16,* 780-784.

Williams B (1973) A critique of utilitarianism in J J C Smart and Bernard Williams *Utitlitarianism: For and Against* Cambridge University Press

Wilson, T.D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117-142.

[i] Subjects would have access to these reasons by social processes of distributed reasoning.
[ii] For a discussion see Kennett& Fine (2008b)
[iii] As Bennett (1974) has persuasively argued, Huck is described as weak-willed in giving into his sympathies. This is an interesting case for Haidt's account in a variety of ways, including his account of virtue.
[iv] Jonathan McGuire (personal communication) reports that subjects he has tested on the Greene & Haidt style dilemmas often later make remarks such as 'that wouldn't happen''. If subjects don't find the scenarios credible their responses may not give us much information on how they make moral judgments in vivo.
[v] All cited in Kennett& Fine 2009
[vi] Note that Haidt's own studies produce evidence in line with this. They found that the capacity to discount or override ones gut reactions in reaching a judgment on so-called harmless disgust scenarios is correlated with higher socio-economic status and lack of religious belief.
[vii] Greene is a consequentialist and is dismissive of deontological considerations as anti-rational. Most of his dilemmas are set up so as to pit consequentialist considerations against deontological ones. But it seems possible to construct dilemmas with competing deontological demands or demands of virtue, which would also elicit slower and more conflicted responses.