Empathy for Prinz of the “Dark Side”

Abstract

Jesse Prinz has argued that empathy plays no important role in moral judgement, and further that it has a “dark side” which renders it by and large bad for morality. This paper challenges these conclusions and demonstrates that it is possible to meet Prinz’s objections by adopting a conceptualisation of empathy which combines elements of Martin Hoffman’s process-focussed definition of empathy with Michael Slote’s agent-centred approach to empathy’s functional role within morality. Beyond proving resilient in the face of Prinz’s attacks, such a conceptualisation of empathy also displays a degree of explanatory usefulness both within Prinz’s own brand of moral sentimentalism and the moral psychology literature more generally. Far from being bad for morality, empathy would appear to be a useful ally to a robust moral sentimentalism.

Table of Contents

Introduction

This paper is about empathy. More specifically, it is about empathy's role in our moral judgements. Although recent decades have produced a large body of research focused on the role of empathy and other 'fellow-feeling' constructs in prosocial behaviour, here I will be setting aside issues of moral motivation and restricting my attention to moral judgement only. My motivation for writing this is two papers by Jesse Prinz, '*Is empathy necessary for morality*' (2011b) and *'Against empathy*' (Prinz, 2011a) in which he challenges empathy's importance for morality, and even goes so far as to say that empathy is bad for morality. My hope is to exonerate empathy from Prinz's charges. Before beginning I should note that in general I am very sympathetic to Prinz's moral theory and share many of his sentimentalist views regarding morality (see Prinz, 2007). However in the aforementioned articles Prinz makes it quite clear that there is one area where our beliefs come apart: that is the issue of *how* our moral judgements come about. More precisely, I believe that empathy plausibly plays an important role in our moral judgements. Prinz, emphatically, does not.

The idea that empathy might play a role in our moral judgements is not a novel one. Simon Baron-Cohen, for one notable example, has argued that people's empathic capacity can be distributed along a bell-curve, and that those we would call 'evil' are those at one end of the distribution - at "zero degrees of empathy" (2011, p. 15). Nor is the link between empathy and morality a new one. Although the term 'empathy' didn't enter the English language until the early 19th century (as a translation of the German 'e*infülung*', literally translated as "feeling-into") something very much like it featured noticeably in the works of the British moralists, most notably David Hume (1740/1978) and Adam Smith (1759/1971), under the name 'sympathy'.

Today, definitions and theories of empathy abound. Some are quite narrow, stressing the importance of an affective match between the empathiser and the object of their empathy. Others are more relaxed and focus less on affective match, whilst some actually involve a two-step process where an affective state is followed by an other-directed response (e.g. Baron-Cohen, 2011). The result of this semantic multiplicity is that any hopes one might entertain of using the term 'empathy' in some generally accepted sense are most certainly in vain. For this reason one must aim for maximum clarity in stating what one means when one says ‘empathy’.

To this end, section 1 of what follows is an attempt to uncover exactly what Prinz is attacking; what is Prinz’s ‘empathy’? The answer to this turns out to be less clear than we might hope. Pushing on to section 2 however, I outline Prinz’s (many) objections to the importance of empathy (as he has defined it) for morality. In section 3, I introduce the theories of Martin Hoffman and Michael Slote, and sections 3, 4 and 5 are an attempt to construct and use a model of empathy in a way which can survive Prinz’s challenges. In sections 6 and 7 I consider some positive contributions which this conception of empathy might be able to make, both in explaining recent experimental results regarding intuitive moral judgement, and within Prinz’s own moral theory as well. I conclude that Prinz is wrong to deny the importance of empathy for morality. Far from being bad for morality, empathy can, should, and possibly already does play an important role in our moral judgements.

Before beginning however, I must clarify some terms. By ‘agent’ I mean simply someone who acts or performs an action. ‘Patient’ refers to any person other than the agent who is affected by the agent’s actions. This dichotomy applies in morally loaded situations, where certain peoples’ (agents) actions have consequences for others (patients). By ‘subject’ I mean a person who is empathising or attempting to empathise, and I use ‘object’ to refer to the person they are empathising with. Unlike ‘agent’ and ‘patient’ the terms ‘subject’ and ‘object’ can apply in situations which are not necessarily morally loaded. They refer to an empathic relationship, not a necessarily moral one. Preliminary orientation duties now discharged we can turn to Prinz’s attack on empathy, beginning with his characterisation of it. What exactly is Prinz’s ‘empathy’?

1. What is Prinz’s ‘Empathy’?

Alas, ascertaining this proves not to be the straightforward task we might hope. In ‘*Is empathy necessary for morality?’* Prinz rejects Stephen Darwall’s definition of ‘empathy’[1], stating that he wants to avoid Darwall’s appeal to the role of imagination in empathising, as this “seems overly intellectual” and “sounds like a kind of mental act that requires effort on the part of the imaginer” (Prinz, 2011b, p. 212). Darwall himself recognised, notes Prinz, that “empathy in its simplest form is just emotional contagion: catching the emotion that another person feels” (ibid., pp. 212-213). Similarly, Prinz rejects Darwall’s suggestion that empathy involves having a feeling that someone *should* feel. He provides the example of a cult member who is delighted by the cult leader’s nefarious plans. Prinz asserts that even though the cult member *should* feel fear, the fact that they

[1] I.e., “feeling what one imagines [another] feels, or perhaps should feel (fear, say), or in some imagined copy of these feelings, whether one comes thereby to be concerned for the [person] or not” (Darwall, 1998, p. 261)

do not means that "if I feel fear on their behalf, that is not putting myself in the cult member's shoes. As I will use the term, empathy requires a kind of emotional mimicry" (ibid., p. 213).

However he immediately goes on to write that:

> "I do not wish to imply that empathy is always an automatic process, in the way that emotional contagion is. Sometimes imagination is required, and sometimes we experience emotions that we think someone *would* be experiencing, even if we have not seen direct evidence that the emotion is, in fact, being experienced. For example, one might feel empathetic hope for a marathon runner who is a few steps behind the runner in first place, or anxiety for the first-place runner, as the second-place runner catches up. We can experience these feelings even if the runners' facial expressions reveal little more than muscular contortions associated with concentration and physical exertion. A situation can reveal a feeling" (ibid., p. 213, emphasis in original).

And finally concludes that:

> "Empathy is a kind of vicarious emotion: it's feeling what one takes another person to be feeling. And the 'taking' here can be a matter of automatic contagion or the result of a complicated exercise of the imagination" (*ibid*.).

I am curious as to how Prinz's allowance that empathy may involve a 'complicated exercise of the imagination' avoids his earlier charge of mental effort. Quite to the contrary, it seems to imply something *more* effortful and intellectual than anything suggested by Darwall – an idle daydreamer may engage in imaginative acts; 'complicated exercise of the imagination' sounds like something reserved for theoretical physicists.

Second, it is unclear how the example of the marathon runner differs in any important respect from the example of the cult member. The key difference between the two examples seems to be simply that in the case of the cult member the object's actual emotion is stipulated. Prinz draws the conclusion that the cult member *should* feel fear (i.e., fear is the appropriate emotional response given her situation), yet tells us that (for whatever reason) they don't; they feel delight. Thus if a subject feels fear on their behalf *we know* they have failed to mimic their emotion (and thus failed to empathise, according to Prinz), in spite of the fact that the cult member *should* feel fear. In the case of the marathon runner however, Prinz doesn't stipulate what she is feeling. Rather he allows that we can infer from her situation what she *would* be feeling. The question is: how can we 'mimic' the anxiety of a runner whose facial expression is not one of anxiety but rather of concentration and physical exertion? Prinz allows that "a situation may reveal a feeling". Given the runner's lack of affective facial expression (due to exertion and concentration), any empathic anxiety Prinz feels for the first-place runner must be entirely inferred from such situational cues. That is to say, Prinz infers that anxiety is the appropriate emotional response given the runner's situation – anxiety is what the first-place runner *should* feel, given the situation she's in. To substitute 'would' for 'should' as Prinz does only serves to go one step further and assert that, given one *should* feel anxious in situation *y*, any particular person in situation *y would* feel anxious.

To see this, simply imagine that the first place runner is entirely unaware of her approaching rival. Or perhaps she is aware of her rival, but isn't anxious because she knows she is a better runner, or simply thinks she is, or doesn't care if she loses, or wants the second place runner (her friend or sister perhaps) to win. In all these cases if Prinz feels anxiety for the runner he is not "putting himself in her shoes". Of course, perhaps Prinz simply forgot to add that the runner does, in fact, feel anxious. Yet without this stipulation it would perhaps seem more appropriate to say that Prinz is experiencing an emotion appropriate to the runner's situation. I will return to this idea in section 3 with Hoffman's definition of 'empathy'.

For now there is yet a third ambiguity to be found within Prinz's definition of 'empathy'. Prior to rejecting Darwall's definition of 'empathy', Prinz considers that offered by C. Daniel Batson[2]. He rejects Batson's definition also, however, arguing that due to its focus on another individual's welfare, it seems not to pick out empathy but rather the related construct concern. He notes that Batson uses the term 'empathic concern' throughout much of his work, but that this seems to combine two separable things: empathy and concern. Prinz observes that one can be concerned for the welfare of an insect, for example, or a beautiful building which has fallen into disrepair, however "empathy seems to connote a kind of feeling that has to be at least possible for the object of empathy" (Prinz, 2011b, p. 212). He concludes that 'empathic concern' actually conflates two distinct things – "a kind of feeling-for an object and a feeling-on-behalf-of an object" (*ibid.*) and suggests that much of the empirical literature, including Batson's, is confounded as it fails to make this distinction. Therefore, Prinz claims, subsequent conclusions drawn about empathy may actually be about concern instead. He also notes that we must distinguish empathy from sympathy for similar reasons, referring us to Darwall's observation that "sympathy is a third-person emotional response, whereas empathy involves putting oneself in another person's shoes"(Darwall, 1998, p. 261).

[2] I.e., as "other-oriented feelings congruent with the perceived welfare of another individual" (Batson et al., 1995, p. 621)

Prinz is correct to insist that we distinguish empathy from related constructs such as concern and sympathy, although I'm sceptical of his assertion that this distinction is rarely recognised. Such doubts aside however, if Batson's use of the term 'empathic concern' really conflates two distinct constructs - 'empathy' and 'concern' - what should we make of Prinz's use of the term 'empathetic hope' in the passage quoted above? Or, for that matter, of his use elsewhere of terms like 'empathetic reasoning' (2011b, p. 219), 'empathetic emotions' (2011a, p. 216), or 'empathetic distress', (2007, p. 37) to cite just a few? Is Prinz conflating all these constructs with empathy (and transitively, then, with each other as well)? Similarly, having set up the distinction between concern and empathy as "feeling-for" and "feeling-on-behalf-of" respectively, what ought we to make of Prinz's assertion that we may feel empathetic hope (or anxiety) *for* a marathon runner? Prinz might respond that 'empathic concern' is problematic because, as he said, one can feel concern for the welfare of a building which has fallen into disrepair yet buildings can't themselves feel concern, but 'empathetic hope' is problem free because one cannot feel hope for a building. However, this is obviously false. In fact I do not think it would be very hard show that it is a conceptual truth that if 'I am concerned for *x*', then 'I am hopeful for *x*' (or 'I entertain some hopes for *x*'). For it would seem very strange indeed were someone to assert that they felt concern for a building in disrepair, but did *not* feel any hope for its restoration (or at least hope that it not fall further into disrepair).

Of course, I very much doubt it is the case that Prinz conflates all the above constructs with empathy. Prinz and Batson are not alone here; terms like 'empathic hope', 'empathic distress', 'empathic anger', and 'empathic concern' flow naturally from the fingertips when writing about empathy, and as a result thoroughly permeate the literature. What's more, it's rare that the appearance in a passage of terms such as these proves so surprising or confusing or ambiguous as to even interrupt the passage of one's reading - they float by quite well as we process the 'semantic flow' of the sentence. So although it is true that we can ask of Batson, "What exactly do you mean by 'empathic concern?'", it's also true that we can ask this of almost every author who has ever mused even briefly on the topic of empathy. Some will have an answer to this question (and in fact you will find mine later in this section), yet use of these terms is part of the vernacular of empathy-based literature; I think to ask this question of others demands that one be clear on the issue themselves. Few are, but that appears to be ok.

So having clearly established, via critique of Batson and Darwall, three features which he takes to be important to empathy – that it is distinct from related 'feeling-for' constructs, that it is non-imaginative, and non-normative (what they *should* feel) – Prinz ultimately comes to rest upon a definition of empathy which runs afoul of all three. What is Prinz's 'empathy'?

Turning to the related paper *'Against Empathy'* provides no clearer picture. In it Prinz explicitly restricts his attacks to what he calls a "broadly Humean" view of empathy, explaining that "following Hume, we can think of empathy as a kind of associative inference from observed or imagined expressions of emotion or external conditions that are known from experience to bring emotions about" (2011a, p. 215). Here again we find the reference to imagination and situational cues which proved so ambiguous previously. Furthermore, the definition of 'empathy' which emerges in the opening pages of this paper almost immediately, as in the previous case, runs into problems.

For example, Prinz begins the abstract by stating that "empathy can be characterized as a vicarious emotion that one person experiences when reflecting on the emotion of another" (*ibid*., p. 214) and, shortly thereafter, informs the reader that:

> "The core idea is that empathy is not the name of a specific emotion but refers, rather, to the experience of another person's emotional state, whatever that emotion might be. More precisely, I will say that empathy is a matter of feeling an emotion that we take another person to have." (*ibid*, p. 215)

Which then, one may well ask, is it? Is empathy a "vicarious emotion we experience when *reflecting on* the emotional state of another"? Or is it the "*experience of* [that] person's emotional state"? Or is it "feeling an emotion *we take* another person to have"? These are clearly three different things. The first would seem to imply something akin to concern or sympathy (recall Prinz rejected Batson's definition due, in part, to its focus on the *perceived* welfare of another); the second might be interpreted as dissolving the affective separateness of subject and object entirely (although I sincerely doubt this is what Prinz intends). Only the third is in keeping with Prinz's previous definition. However, it suffers the same ambiguities regarding imagination and feeling what another should feel. Finally, we have Prinz's earlier definition offered in his book '*The Emotional Construction of Morals'*: "in empathy, we feel the same emotion that someone else is feeling; we put ourselves in another person's shoes. If you are afraid, an empathetic person will experience fear too" (2007, p. 82). Here there is no reference to imagination or emotions we take others to have, the important thing appears to be an actual affective match between subject and object.

At this point, I don't really want to spend any more time on Prinz exegesis. I think the above has provided enough information for us to conclude that it is actually quite unclear exactly what Prinz takes 'empathy' to refer to. Still, we may hazard a guess. Given the above considerations,

I attribute to Prinz the following view: empathy is distinct from related 'feeling-for' constructs (i.e., concern and sympathy); doesn't refer to a specific (invariant) emotion; but to a 'matching' of emotion from object to subject (call this the ‘matching condition’); where the target emotion is that present in the object at the time when the subject’s empathising (attempt) takes place (call this the ‘mimicry condition’; it entails the ‘matching condition’); and may result from an automatic (contagion) or imaginative (situational inference) process, provided the mimicry condition is met. This definition in hand, we can now turn to Prinz’s case against the importance of empathy for morality.

2. Prinz Against Empathy

Although Prinz challenges the idea that empathy is a precondition for moral judgement, moral development and moral motivation (2011a, 2011b), my focus here will be primarily on the role of empathy in moral judgement (and moral development to the extent of its relation to moral judgement)[3]. The role of empathy in moral development and moral motivation already has its defenders (see e.g., Baron-Cohen, 2011; de Waal, 2010; Eisenberg-Berg & Neal, 1979; Hare, 1999; Hoffman, 2000; Preston & de Waal, 2002; Slote, 2010) Prinz identifies five ways in which empathy might prove to be a precondition for moral judgement. It may be a: constitutive, causal, developmental, epistemic or normative precondition. For our purposes here, we can consider some of these together.

2.1 Against the constitutive and the causal precondition theses

According to the constitution thesis, moral (dis)approbation is simply empathic (dis)pleasure redirected outwards: “when we think of the happiness in some group of people, we experience empathic pleasure. This pleasure then becomes a component of approbation, which is a pleasure we take in the character or action that has produced happiness in the group under consideration” (Prinz, 2011a, pp. 216 - 217); Prinz suggests that there is some hint of this in Hume. He gives us the causal precondition thesis stated counterfactually: “on any given occasion in which we experience a feeling of (dis)approbation, that feeling would not have arisen had there not been a prior empathetic response” (2011a, p. 219)

Prinz argues that we can easily see that disapprobation is not constituted or caused by empathy by considering victimless crimes, crimes against oneself, crimes where the victim dies instantly, crimes against groups, crimes against animals, and moral judgements at high levels of abstraction (eg, ‘murder is wrong’). He offers us an example of a person who uses their pet cat as an instrument of masturbation. Prinz writes that he may judge that act to be wrong, and that research suggests that

[3] For more on the role of empathy in moral development and moral motivation see Hoffman (2000), Preston & de Waal (2002) and Baron-Cohen (2011).

sexual transgressions elicit disgust (Rozin, Lowery, Imada, & Haidt, 1999). However it is absurd, says Prinz, to claim that his disgust comes from the vicarious disgust he experiences when contemplating the reaction of the cat. As he points out, "for all I know, little Tigger enjoys the experience" (2011a, p. 218). Similarly, Prinz offers the example of someone who paints graffiti in a public park. He notes that such crimes against the community elicit contempt (Rozin, et al., 1999), yet points out that his contempt for the vandal may not involve empathy for the victims because others in the community may think that graffiti is wonderful (how Prinz reconciles this with the immediately prior assertion that such crimes elicit contempt remains a mystery). Similarly, in cases where I myself am the victim, it makes as little sense to suppose that I might empathise with myself as it does to suppose that, in cases where the victim dies, I can empathise with a dead person. These examples all fail to meet the matching condition and Prinz claims they demonstrate that moral disapprobation can't be constituted or caused by empathy.

Typically, when we think of empathy in the context of morality, we think of empathy with patients. However Prinz also considers the empathy-based constitution thesis of Michael Slote (2007, 2010). Slote's position is interesting in that he argues that moral (dis)approbation is constituted by an empathic relation to agents as well as to patients. This is an appealing feature of Slote's account, and I shall have more to say on it a bit later. For the moment however, what does Prinz think of Slote's agent-centred version of the constitution thesis?

He notes that an initial benefit of Slote's view is that, due to its focus on agents, it avoids the problem of victimless crimes and those other similar cases above. For example, disapprobation for the cat-loving masturbator would involve a failure to empathise with his motives, Tigger's particular sexual proclivities notwithstanding. Prinz rejects Slote's view for two reasons however. First, he argues that there exist cases where moral disapprobation occurs in spite of empathy with the moral agent. For instance, a recovering paedophile may be able to empathise with another paedophile (or non-paedophiles may empathise with paedophiles who have been victims of prior abuse) yet still morally condemn their actions. Second, Prinz doubts whether Slote's theory can properly count as a constitution thesis at all. He writes that according to Slote "disapprobation involves *a lack of empathy* with a moral transgressor" (Prinz, 2011a, p. 219, emphasis in original), yet if one *fails* to empathise then disapprobation cannot be constituted by empathy. To posit some kind of 'dis-empathy' is implausible, says Prinz. He cites evidence which suggests that moral disapprobation involves emotions of blame such as anger, disgust and contempt (Prinz, 2007; Rozin, et al., 1999), and concludes that these emotions, not (dis)empathy, likely constitute moral judgements.

2.2 Against the developmental precondition thesis

Although denying any immediate causal role for empathy, Prinz considers that perhaps empathy might play a causal role earlier in life. For example, perhaps it is essential for developing a moral sense. Psychopathy, for which a lack of empathy is a diagnostic criterion, appears to provide strong support this view. Psychopaths are notoriously immoral and typically fail to make the moral/conventional distinction[4]. However Prinz argues that we cannot conclude from this that empathy is developmentally necessary as another plausible explanation exists, namely, his own theory based on (non-empathic) conditioning (see Prinz, 2007; or Prinz, 2008, for a very brief overview). He argues that "we might think of punishment as inculcating a sense of disapprobation directly without any essential empathetic involvement" (2011a, p. 222) and points out that flattened affect is another diagnostic criterion for psychopathy, but that affective arousal is an essential component of conditioning. He concludes it is probably these affective deficits (specifically, deficits in fear and sadness), not a deficit of empathy, which account for the psychopath's failure to obtain moral competence. Thus we cannot say that empathy is a developmental precondition for moral judgement.

2.3 *Against the epistemic and the normative precondition theses*

Prinz next considers whether empathy might be an epistemic or a normative precondition for moral judgement, that is, whether empathy might help us to see when (dis)approbation is appropriate by directing out attention towards the emotional wellbeing of those around us, or whether appeals to empathy might allow us to justify why an action is wrong. For example, without empathy we may still make moral judgements about things, but fail recognise when such judgements are warranted. Prinz doubts that this is the case. Although he allows it may sometimes be descriptively true that empathy plays this epistemic role, he argues it does so only contingently, and that it is ultimately unreliable as an epistemic guide. He provides the following counterexamples of alternative routes to moral judgement. First, cases where deontological considerations override utilitarian principles, e.g., one might judge that it is bad to harvest the organs of one healthy person in order to save five. It seems that we should feel cumulatively more empathy for five people in need than for one healthy person but the judgement that killing him is morally wrong does not track this empathic response. Second, cases involving moral judgements issued from behind a Rawlsian veil of ignorance. Such judgements are motivated by concern for the self, not empathy. Lastly, Prinz's own conditioning-

[4] This distinction will be explained later. For the moment suffice to say that psychopaths don't see a difference between conventional transgressions (e.g., cutting a queue, talking loudly in a library) and moral transgressions (e.g., assault, theft).

based theory, in which moral judgements are conditioned moral sentiments. Prinz argues that these alternatives all speak to the contingency of empathy's epistemic role in moral judgement.

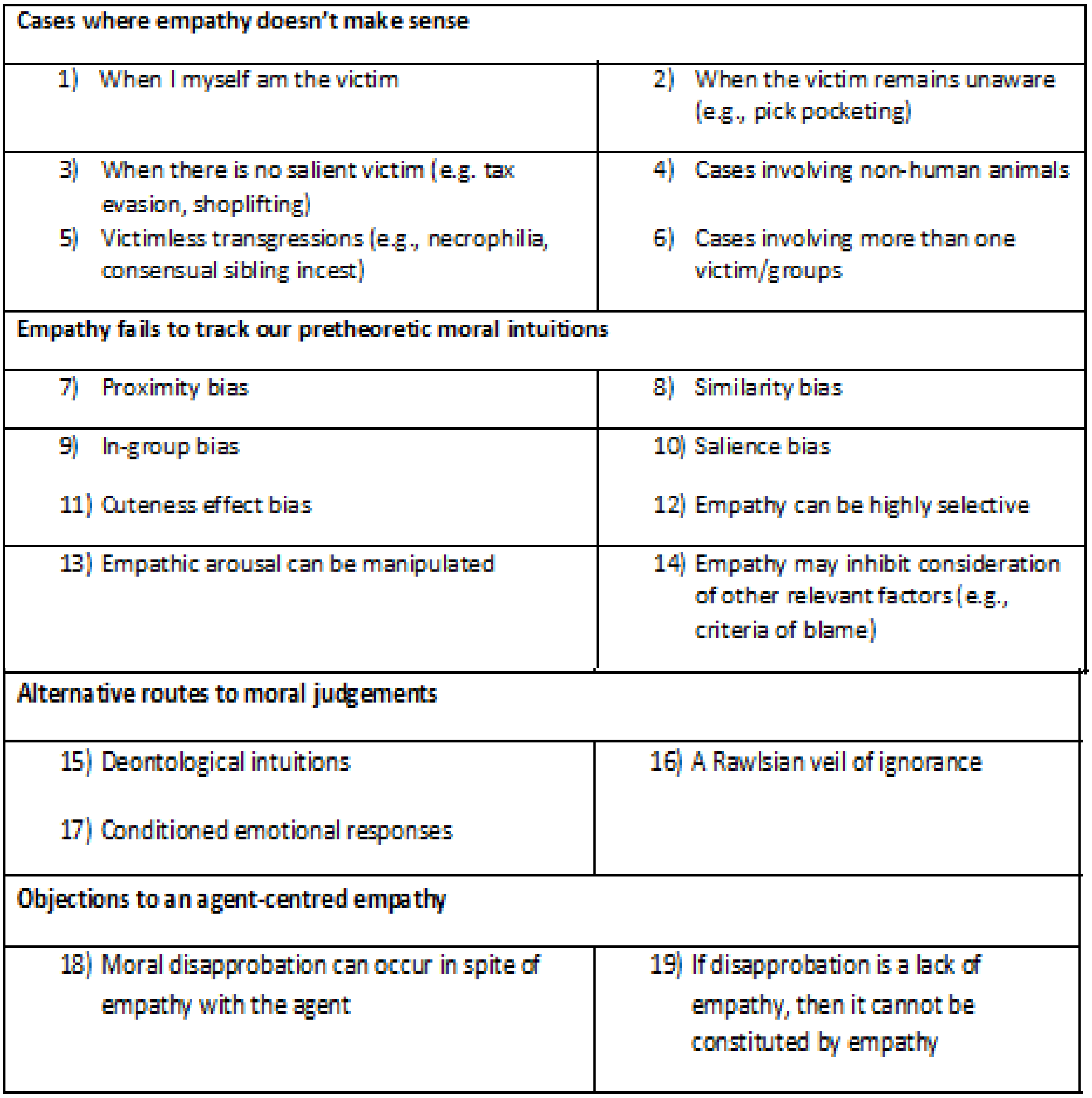
Prinz further observes that empathy is notoriously prone to bias: we empathise more with those who are physically near to us than with distant others, and more with those who are similar to us than those who are not (Hoffman, 2000); we empathise more with 'cute' things (Batson, Lishner, Cook, & Sawyer, 2005; Sherman & Haidt, 2011); empathy is easily manipulated and may lead to preferential treatment (Batson, Klein, Highberger, & Shaw, 1995); empathic arousal varies with the salience of the object's emotion (Tsoudis, 2002); and empathy is prone to 'in-group' biases (Brown, Bradley, & Lang, 2006). Although it may be descriptively true that we sometimes use empathy as an epistemic guide, Prinz argues that such biases prevent it from being a valid normative precondition for moral judgement. That is, empathy's many biased nature ruins its credibility to justify our moral judgements.

2.4 The "dark side" of empathy

Prinz allows that proponents of empathy might accept his arguments and agree that empathy is perhaps not *necessary* for morality, whilst still insisting that it is in general a pretty good thing. He argues that a morality based upon empathy would not be a good thing however, pointing to what he calls the 'dark side of empathy' (2011a, p. 227). Firstly, empathy's long list of biases cast doubt on whether empathically motivated judgements succeed in tracking basic pretheoretical moral intuitions: "maybe empathy is a bad thing. It does not track approbation, and if we use it in that capacity, we would make moral mistakes" (ibid., p. 228). Prinz notes that Hume recognised the potential dissociation between approbation and empathy arising from such biases, but is dissatisfied with Hume's reply: that we can align empathy with approbation by adopting "the general point of view" (Hume, 1740/1978, 3.3.1). Prinz asserts that the general point of view is something which we rarely adopt, and that empathy may in fact be its greatest impediment. Attempts to widen the scope of empathic focus often have some positive result but "too often land in the wrong place…with empathy, we ignore the forest fire, while watering a smouldering tree" (2011a, p. 228). He concludes that although empathy may be a good thing in the domain of friendship, where the norms are all about partiality, friendship is for this reason "not a paragon case of a moral relationship" (ibid., p. 229).

Finally, Prinz considers some real-life moral systems which arguably place some emphasis on empathy. Collectivist and liberal cultures are two examples. He argues that such empathy-based moral frameworks are compatible with (although don't necessarily entail) many moral ills. Whether such cultures really constitute empathy-based societies is a topic worth a paper of its own (at least!) and I won't pursue it here, except to say that Slote explicitly considers one area in which empathy and liberalism might come apart: issues surrounding freedom of speech. Slote argues that freedom of speech is something central to liberalism, but that his empathy-based ethics of care would sometimes morally permit certain actions which deny some people this right (Slote, 2007, pp. 67-87). And we might also question whether collectivist cultures are based upon other-focused processes like empathy, or something more self-focused like duty. Suffice to say that when Prinz concludes that such cultures arguably emphasise empathy, we should emphasise the word 'arguably'.

2.5 A lot to answer for: summing up



Cases where empathy doesn't make sense	
1) When I myself am the victim	2) When the victim remains unaware (e.g., pick pocketing)
3) When there is no salient victim (e.g. tax evasion, shoplifting) 5) Victimless transgressions (e.g., necrophilia, consensual sibling incest)	4) Cases involving non-human animals 6) Cases involving more than one victim/groups
Empathy fails to track our pretheoretic moral intuitions	
7) Proximity bias	8) Similarity bias
9) In-group bias 11) Cuteness effect bias	10) Salience bias 12) Empathy can be highly selective
13) Empathic arousal can be manipulated	14) Empathy may inhibit consideration of other relevant factors (e.g., criteria of blame)
Alternative routes to moral judgements	
15) Deontological intuitions 17) Conditioned emotional responses	16) A Rawlsian veil of ignorance
Objections to an agent-centred empathy	
18) Moral disapprobation can occur in spite of empathy with the agent	19) If disapprobation is a lack of empathy, then it cannot be constituted by empathy

Table 1: Prinz's case against empathy

3. Should we feel bad for proponents of empathy?

Prinz has certainly assembled an impressive battery of charges against the role of empathy in moral judgement, and it may seem as though proponents of empathy should be very worried at this point. However, I believe a robust conception of empathy and its functional role in moral judgement can be mustered to address these charges – we can feel hope for empathy yet.

Firstly, we might question whether Prinz's definition of 'empathy' is really up to scratch. He readily admits the possibility that proponents of empathy may disagree with his characterisation of it; that they may think he has "smuggled in some problematic property" (2011a, p. 216). In fact, this is entirely the wrong metaphor. The act of smuggling entails carrying something, in some sense, hidden or extra. Yet I believe Prinz's definition of 'empathy' is questionable precisely because it is too limited in the scope of behaviours which it designates. Far from smuggling anything in it seems a few bales shy of a straw-man. Indeed, there is some indication that even Prinz himself believes empathy to be something more than his characterisation of it. In '*Against Empathy'* he informs the reader that he is "restricting" his arguments to a broadly Humean view of empathy in order to focus the debate (p. 215). In '*Is Empathy Necessary for Morality?'* he notes that empathy "in its simplest form" is just emotional contagion (p. 212). And recall his reluctance to allow imagination a role in empathising.

Prinz denies that there is "anything anachronistic about [his] notion of empathy" (2011b, p. 213), citing David Hume's (1740) and Adam Smith's (1759) use of the term 'sympathy'. Yet an obvious question arises regarding the usefulness of 250 year-old definitions in deflecting the charge of anachronism. Prinz explicitly takes himself to be attacking the kind of empathy that he think these philosophers had in mind. Whilst it's certainly true that Hume's and Smith's use of 'sympathy' shares many features with conceptions of 'empathy' found today, since their time knowledge from areas of scientific enquiry such as neuroscience, cognitive science, evolutionary theory and developmental psychology, to name just a few, has contributed significantly to our understanding (or at least our conceptualisation) of the human mind and behaviour. Might not contemporary proponents of empathy be justified in wanting recourse to resources such as these, that weren't available to the British moralists, in constructing their definitions? This is not to say that Hume and Smith weren't insightful, or that they didn't have important things to say about empathy and its role in morality. Nor is it true that what Prinz defines *isn't* empathy. It's just that it's a very basic form of empathy, and while some contemporary authors may focus their attention on empathy at this level, others support the view that more sophisticated forms of (what they say is) empathising become available as an individual develops cognitively. Thus we might ask whether another, less simple and unrestricted characterisation of 'empathy' and its functional role in moral judgement might be better able to address the concerns and objections offered by Prinz. I believe it would, and that such a characterisation can be taken from the work of Martin Hoffman and Michael Slote.

3.1 Defining 'empathy': Prinz vs. Hoffman

Hoffman offers us an interesting take on empathy. He notes that in defining 'empathy', many authors focus purely on outcome (Prinz's reliance on the matching and mimicry conditions marks him as such a one). Yet Hoffman remarks that: "the more I study empathy, the more complex it becomes. Consequently, I have found it far more useful to define empathy not in terms of outcome (affect match) but in terms of the processes underlying the relationship between the observer's and the model's feeling. " (2000, p. 30). Thus Hoffman defines 'empathy' as, "an affective response more appropriate to another's situation than one's own" (ibid., p. 4) and for him an essential component of an empathic response is "the involvement of psychological processes that make a person have feelings that are more congruent with another's situation than with his own situation" (ibid. p.30).

The first interesting aspect of Hoffman's definition is obviously his focus on process rather than outcome. He maintains that although empathic processes may often result in an affective match, such a match need not necessarily occur for empathy to have taken place (no matching or mimicry condition). What is important is the engagement of particular psychological processes. Second, note Hoffman's reference to process*es*, plural. In fact he identifies five distinct modes of empathic arousal which increase in scope and decrease in automaticity. They are: motor mimicry and afferent feedback; classical conditioning; direct association; mediated association (verbal mediation); and role-taking.

Hoffman's describes the first three modes as involving automatic, preverbal and essentially involuntary responses, which are available very early on in life. *Mimicry and feedback* involves two sequential steps: imitating the facial expression of the object, followed by afferent feedback which provides the relevant emotion. In *classical conditioning* distressed posturing by the mother (e.g. stiffening body) serves as an unconditioned stimulus for the infant's distress. The mother's facial and vocal distress may then become conditioned stimuli, which can be further generalised beyond the mother (e.g., facial and vocal displays of distress in people other than the infant's mother will serve as a conditioned stimulus and elicit the conditioned response of distress in the infant). *Direct association* involves cues in the victim's situation reminding the observer of similar events in their own past and cueing the associated emotions.

The last two modes involve higher degrees of cognitive sophistication, and allow the subject to empathise with an object beyond their immediate situation. *Mediated association* occurs when the victim's situation is related verbally. Semantic processing and decoding of the verbal message may lead to direct association or result in imagined sounds/images enabling mimicry. *Role-taking* involves putting oneself in another's place and imagining how they feel. This role-taking can be either 'self-focussed', i.e. imagining how I would feel in that situation or 'other-focussed', i.e. imagining how they (the other) would feel. Hoffman reports that of the two, self-focussed role-taking produces more intense empathic affect.

Note that not all five modes of empathic arousal appear to be congruent with Prinz's definition of 'empathy'. The key difference is Prinz's focus on outcome requiring the matching and mimicry conditions versus Hoffman's focus on process requiring just that "the observer attends to the victim and the feelings evoked in the observer fit the victim's situation rather than the observer's" (Hoffman, 2000, p. 40). The matching and mimicry conditions are entailed by Hoffman's first two modes of empathic arousal, and therefore compatible with Prinz's definition. However they are not entailed by the later three modes. Thus these features of Hoffman's account expand the scope of empathic possibility. And, as we shall see, it does so in a way which allows it to better deal with objections like Prinz's, as well as to fit with our intuitions, and lay and technical use of the term 'empathy'.

Some might ask what distinguishes concern from empathic concern, anger from empathic anger, hope from empathic hope, distress from empathic distress, and so on in the later modes, when the emotions of the subject don' match those of the object. What makes their anger empathic and not simply run-of-the-mill anger? It is a focus on outcome, however, that to require the identification of some qualitative difference between the object's affective state and the subject's 'empathic' affective state. A focus on underlying processes can immediately explain this distinction by ignoring issues of qualitative difference and explaining the distinction between affect and empathic-affect in terms of underlying process: empathic-anger is (perhaps) qualitatively just like regular anger, only has arisen via some specific (set of) psychological process(es). Of course it may still be that some qualitative difference can be identified, but a focus on process allows us to sensibly use terms such as 'empathic concern' without requiring this

A focus on process rather than outcome can also make sense of intuitive lay applications of the term 'empathy'. Empathy is not a moral domain-specific construct. We can empathise with a friend whose grandmother has recently died of old age, for example, without thereby construing the situation as a moral one. We also like to talk about empathising with fictional characters, such as those in novels. If a novel is well-written, when a cared about character finds themself in a dire or tragic situation, readers typically feel moved for them and this is often described in terms of empathy for the character. Yet it is quite unclear how a focus on outcome could allow empathy between a reader and a fictional character. With a focus on process we can also empathise beyond the immediate situation, as when we feel sad for a terminally ill person upon considering their life situation, in spite of the fact that they are currently engaged in some enjoyable activity.

Along similar lines Hoffman's definition of empathy can directly address some of Prinz's cases where empathy doesn't make sense. For example, in cases such as pick pocketing although the victim may be unaware of the theft, certainly we can say that anger is appropriate to her situation[5]. Thus regardless of whether she feels anger or not (though presumably she will eventually, once she becomes aware of the loss), if I feel anger, and if that anger has arisen via a specific (set of) psychological process(es), then I am empathising with her. The same goes for empathising with non-human animals and with groups. The case of empathy with groups is easy to see. What causes us to group them together in the first place is their shared situation, and we can certainly experience emotions appropriate to that situation regardless of what any particular group member might be feeling at the time. If those emotions come about in the right way then we have empathised. Hoffman's definition can also resolve the earlier confusion regarding the community of graffiti 'victims'. If Rozin et al are correct that crimes against the community elicit contempt then Prinz's contempt for the graffiti artist can certainly be called appropriate to the situation of a member of that community, regardless of any enjoyment which particular members of that specific community might find in the graffiti.

The case of non-human animals is perhaps harder to see as it may still seem strange to some people to suggest that one could empathise with a non-human animal. However I believe it is not implausible. Non-human animals can feel fear and pain, for example, and we can often recognise such feelings in their faces and behaviour. Further, given we can't communicate with them to any large degree, they often feel fear in situations which are perfectly safe simply because they don't understand what is happening[6]. We might suppose that being used as a masturbation aid would be

[5] That is, of being the victim of theft. Not of being the victim of theft yet being unaware that one is the victim of theft.

[6] Anyone who has ever tried to rescue an injured animal, or even bathe an unwilling cat or dog, can attest to this.

one such situation (it is possible the cat would feel pain as well, though we can't know as, thankfully, Prinz didn't go into specifics)[7].

Lastly, a focus on process can address Prinz's observation that psychopaths suffer general affective deficits which are more likely to contribute to their immoral behaviour than a lack of empathy. Affective arousal is equally important for empathic processes. To point to psychopaths' affective deficits is to point below the level of contention. Yes, psychopaths' have general affective deficits, and these most certainly have a negative effect their ability to be conditioned and to empathise. The question is, however, which of these two subsequent deficits (conditioning or empathy) is responsible for behaviour that we all seem to call 'immoral'?

Thus Hoffman's process-oriented and more complex definition of 'empathy' is easily able to handle an array of issues which prove problematic to outcome-oriented definitions such as that which Prinz attacks. However much of Hoffman's (excellent) work focuses on the positive influence of empathy on prosocial behaviour, yet my focus here is on moral judgement and I also want to find a conceptual basis for the role of empathy in such judgements. I believe the best route to this is the empathy-based moral theory of Michael Slote.

3.2 Using empathy: Understanding Slote

As mentioned earlier, Prinz does in fact consider Slote's agent-centred empathy. Although he ultimately rejects it, his rejection of it rejection is based upon a quite severe misunderstanding of Slote's view. Prinz attributes to Slote the following view: moral approbation is constituted by empathy with a moral agent. Moral disapprobation, therefore, is constituted by a failure to empathise (or a lack of empathy) with a moral agent.

This is not Slote's view.

In arguing against empathy, Prinz seems to assume that its proponents have two exclusive available alternatives: either they may say that moral judgement stems from empathy with patients, or that it stems from empathy with agents, but not both. Prinz places Slote in the latter camp. This exclusive dichotomy is unfair however, as it forces empathy into an unrealistic domain-specific role. The agent-patient dichotomy only applies in situations where one person acts and their actions affect at least one other person's welfare; for this reason they seems to designate specifically moral situations (more on this later). However, as already noted, empathy often occurs in non-moral situations. It makes no sense to talk of agents and patients in the case of the friend's deceased

[7] Prinz actually stated that I would feel disgust at the actions of the cat-masturbator. I do not mean to imply that a cat might feel disgust. I will say a bit more on empathy and disgust later.

grandmother, for example. To restrict empathy exclusively to either agent- or patient-centeredness is to miss the domain-general nature of empathy entirely - we empathise with *people*, not with agents or patients. Whilst it's true that in certain situations some of those people may be termed 'agents' and others 'patients', and that such considerations may well influence the arousal and outcome of empathic processes in those situations, it is not the case that they determine the empathic locus in the first place. An impressive feature of Slote's account is that it is not the case, as Prinz believes, that Slote has simply exchanged a patient-centred empathy for an agent-centred one, but that it has room for both. That being said, I will for ease of reference continue to use Prinz's term 'agent-centred' to refer to it.

According to Slote, then, moral approbation is *warranted* by an agent's actions displaying empathy for the patient. Moral approbation is *constituted*, on the other hand, by the judger's apprehension of that agent's empathy for the patient; that is, by the judger's empathy with the agent's empathy. Disapprobation, then, is not a case where the judger *fails to empathise* with the agent, or of some kind of dis-empathy. Rather it is when the judger does empathise with the agent, and this alerts them to (lets them feel) the agent's lack of empathy for the patient. As Slote writes:

> "We [judgers] sometimes see that someone else feels empathic concern for another and/or see that empathy reflected or expressed in their actions toward that other person, but our ability to see or notice such things may itself partly or wholly depend on our ability to empathize with such an empathic agential point of view, with the empathy of agents" (2010, p. 24)

And regarding disapprobation:

> "If a person's actions toward others exhibit a basic lack of empathy, then empathic people will tend to be chilled (or at least "left cold") by those actions, and I want to say that those (reflective) feelings toward the agent constitute moral disapproval. Thus empathy with an agent's lack of empathy or empathic concern for others, with an agent's cold indifference (or worse) toward others, yields a similar feeling in the person who *has* empathy, and that feeling, which I have just said amounts to a feeling of disapproval, is very different from the warmth or tenderness that is characteristically expressed in what an empathic person does as an agent. We are clearly, then, talking about two different points of view here: that of agents and that of someone who approves or disapproves of a given agent or agents" (ibid., p. 25).

This correction immediately negates Prinz's objections to agent-centred empathy; Slote's theory is not fundamentally flawed as a constitution thesis at all. Thus it remains a benefit of Slote's agent-centred account that it can, as Prinz noted, deal with objections 1) – 6) above. For example, when I myself am the victim my moral disapprobation results from my apprehension of my transgressor's lack of empathy for me, rather than empathy with myself. The others are handled in a similar fashion; it is our perception of an agent's empathic concern for the patient which is important in all these cases, regardless of whether the patient is aware, salient, a non-human animal or a group of individuals. Victimless crimes perhaps seems more problematic, given that if (dis)approbation is constituted by my (empathic) perception of an agent's (lack of) empathy for the patient, then Slote's theory still seems to require that there actually be a patient for the agent to (not) empathise with. So how can it account for victimless crimes?

Well, first we might note that Prinz plays rather free and loose with the distinction between moral transgressions and crimes. Necrophilia and sibling incest may be crimes (in today's Western society), but this doesn't necessarily make them moral transgressions. In fact, 'victimless' in and of itself seems to speak against the moral loading of a situation. Prinz would probably respond by pointing to the study undertaken by Murphy, Haidt and Björklund (2000; cited in Haidt, 2012; Nichols, 2004; Prinz, 2007). Murphy et al probed moral attitudes toward consensual sibling incest by asking American college students about a hypothetical case where a brother and sister have consensual sex. The scenario was carefully constructed so as to avoid all of the typical objections to incest (eg, they used contraception, only did it once, no one found out, and they felt closer as a result). 80 percent of subjects initially deemed the act to be morally wrong. However, when asked why it was wrong they were unable to justify their judgement. Each attempt at justification was shut down by the experimenter reminding them of the relevant aspect of the scenario (eg, if the subject pointed to deformed offspring as justification, the experimenter reminded them that the pair used contraception). In each case the subjects generally accepted the experimenter's counterarguments as valid, however only 17 percent were willing to change their initial judgement that the act was morally wrong. Most simply collapsed into assertions such as 'It's gross!' or 'Incest is nasty!'. These results seem to suggest that some victimless transgressions are nonetheless held to be moral transgressions.

However, disapprobation in cases without patients could still in some sense be constituted by empathy with an agent in that we may still fail to resonate with the agent's motivations (motivating feelings). Slote holds that an agent's cold indifference yields a similar feeling in the empathic judger, and it is their rejection or dislike of this feeling that constitutes their disapprobation. Similarly then, in cases of sibling incest or necrophilia it seems reasonable to suggest that those who aren't inclined toward such activities would be disgusted upon considering having romantic or sexual feelings towards siblings or corpses, that is, upon considering the feelings of the agent, even though no patients exist. Thus even though empathy with the patient has dropped out of the picture, empathy with the agent might still function in basically the same way as it does in cases involving patients; in the latter case, the agent's cold-heartedness leaves us chilled, in the former their perverse sexual feelings leave us disgusted.

Within moral psychology there is a well-known and much discussed phenomenon known as the 'moral/conventional distinction'. From the age of about 3 years old, children begin to distinguish moral transgressions from conventional transgressions along four key dimensions. I will say more on this later, and should note that Prinz is sceptical of the importance of this distinction, but for the moment I simply want to point out that one of these key dimensions is that judgements regarding moral transgressions are typically justified by reference to the welfare of the victim (whereas conventional transgressions are not). This seems at direct odds with Murphy et al's finding that people typically refused to reverse their negative moral judgements regarding consensual sibling incest in spite of the absence of a victim.

However Haidt reports a follow-up study done by Joe Paxton and Josh Greene (Forthcoming) in which Harvard students were presented with the sibling incest scenario via computer[8]. In one experimental condition, the computer forced subjects to wait 2 minutes before responding with their judgement, and Haidt reports that in this case subjects became "substantially more tolerant toward Julie and Mark's (the brother and sister) decision to have sex" (Haidt, 2012, p. 69). Thus although an initial flash of disgust (which I suggested is can be explained in terms of agent-centred empathy anyway) may result in immediate moral condemnation in some cases of patientless transgression, when given time to reflect on the specific - patientless - nature of the scenario, subjects' moral condemnation is apparently substantially reduced. It may be that the idea of a victimless moral transgression doesn't make sense after all. Prinz asks us to consider 'victimless crimes' as evidence against the role of empathy in moral judgement, yet crimes aren't necessarily immoral, and perhaps we have reason to believe that there are no victimless moral transgressions.

3.3 Putting it together: …Sloffman?

In the previous section I attempted to show how Hoffman's process-focussed definition of empathy, and Slote's use of agent- (and patient- !) centred empathy, can address the cases in which Prinz

[8] The response patterns of the Harvard students conformed to those reported by Murphy et al (in the same condition).

objects that allowing empathy a role in moral judgement just doesn't make sense. It's rue however that if we view empathy as the engagement of particular psychological processes then we cannot, *pace* Slote, hold it to be a *constitutive* precondition of moral judgement. A judgement is clearly different from a process and would seem strange to say of a moral judgement (e.g. that *x*'s ϕ-ing is morally wrong) that it somehow *is* a process. It seems more correct to say that a judgement *results from* some process, and this is compatible with empathy having a causal role in moral judgements. Whilst I agree with Prinz that moral judgements are probably constituted by emotions such as anger, disgust and the like, I believe it is plausible that it is the prior occurrence of empathic (including agent-centred) processes which render these domain-general emotions *moral* emotions. We get angry with an agent *because* we have empathically felt their cold-heartedness, and, as Slote says, "that feeling (cold-hearted indifference), which I have just said amounts to (I say *causes*) a feeling of disapproval, is very different from the warmth or tenderness that is characteristically expressed in what an empathic person does as an agent" (Slote, 2010, p. 25, parentheses mine). I would argue that when our empathic processes alert us to an agent's disregard for the welfare of those around them (or disrespect for their rights, or perverse sexual feelings), then we feel anger (contempt, disgust) *because* it isn't reflective of *our own* empathic feelings toward the patient (or perhaps just our feelings in general, such is the case of necrophilia).

Thus the felt (empathically) cold-hearted lack of empathic concern an agent's actions display for a patient *causes* me to become angry (if I am an empathic person and thus myself empathically responsive to the patients). This assumes two things. First, that the welfare of the patient is somehow relevant, in that in cases involving patients moral judgements depend upon an agent's actions displaying empathic-concern for that patient (i.e., concern for their welfare). Support for this assumption will be offered below in section 5 on the moral/conventional distinction. The second assumption is that is that an agent's intentions are relevant in that moral judgements depend upon their intentions being in accord with empathic concern for the patient. The idea that an agent's intentions affect moral judgement will be discussed in section 6 on the moral faculty. Now I turn to Prinz's claims that alternative routes to moral judgement exist which don't involve empathy.

4. The contingency of empathy: non-empathic routes to moral judgement

Prinz argues that the existence of routes to moral judgement which don't involve empathy speak against the necessity of empathy's role therein. He offers us three examples of such alternative routes: cases where deontological intuitions override utilitarian considerations; when moral judgements are issued from behind a Rawlsian veil of ignorance; and moral judgements which are simply conditioned responses. I shall consider these in turn.

4.1 Deontological override

Firstly, consider cases where deontological intuitions override utilitarian considerations. Prinz offers up the classic case of the hypothetical surgeon who could kidnap and murder one healthy individual from the hospital waiting room and harvest his organs, thereby saving the lives of five dying others. Prinz claims that we intuitively want to say that murdering the healthy individual is morally wrong, yet a judgement based upon empathy should, he says, yield the contrary judgement, given that we feel cumulatively more empathy for five dying individuals than one healthy one.

Even if this is so, it is a strange assertion for Prinz to make given his prior assertion that we cannot empathise with groups. At the very least this notion of 'cumulative empathy' is in need of further explanation. What is more, the empathy Prinz has in mind here is obviously patient-centred empathy. But what might agent-centred empathy make of the case of the surgeon? In particular, how does it cope with cases of multiple (potential) victims? The answer is: quite easily. Recall that on an agent-centred model my moral disapprobation is constituted by my apprehension of an agent's lack of empathy for the victim. Thus in the case of the surgeon, although it may well be that his actions (laudably) display empathy for the five dying people, the clear lack of empathy displayed for the healthy individual "leaves me chilled" and this feeling constitutes my disapprobation and I judge the murder of the healthy individual to be morally wrong.

In fact Slote himself argues at some length that empathy correlates highly with commonsense deontological intuitions. For example he argues that the distinction between doing vs. allowing harm is something central to deontology, and that we emotionally "flinch" from causing harm more than from allowing harm (2007, p. 44) such that harm we cause has a greater causal immediacy for us than harm we simply allow. Killing, for example, puts us closer to another's harm than simply letting die, and is thus prone to stronger empathic arousal. Slote argues that his empathic model is perfectly compatible with cases of deontological override.

4.2 Rawlsian veil of ignorance

John Rawls famously proposed that principles of justice for the ordering of the basic structure of society should be selected from what he termed the original position (Rawls, 1971). When in the original position one selects principles of justice from behind a veil of ignorance, i.e. in (feigned) ignorance of one's *actual* social position, *actual* abilities, *actual* wealth, and so on. Rawls' idea was that if I am blind to my actual circumstances when I select principles of justice I will be motivated by purely rational self-interest to order the basic structure as fairly as possible. Prinz argues that moral judgements issued from behind such a veil of ignorance are based upon rational self-interest and therefore don't involve empathy. As such, they speak to the contingency of empathy's role in moral judgement.

Hoffman explicitly considers his model of empathy in relation to Rawls' veil of ignorance and principles of justice (Hoffman, 2000, pp. 231 - 238). Although he agrees with the fairness of Rawls' principles, Hoffman questions the assumption that people will be willing to abide by those principles once the veil has been lifted and they are made aware of their actual circumstances. In Rawls' case, the assumption is that people are both rational and reasonable, where 'reasonable' refers to a willingness to subordinate one's own interests to those of others in certain circumstances (e.g., when doing so conforms to principles of justice one recognises as objectively fair). Hoffman doubts (rightly, in my opinion) that consideration of abstract principles will be sufficient motivation for high producers to conform to Rawls' difference principle (for example). However he argues that due to empathy's functioning as a prosocial motive (see Hoffman, 2000, chapter 2) it stands as a plausible vehicle for the post-veil stability of the basic structure[9]. Thus Hoffman suggests that empathy is important in motivating behaviour which conforms to abstract principles (of justice, for example). I said my focus here was on moral judgements, but still Hoffman's arguments go some way to suggesting that empathy may not in fact be bad for morality.

4.3 Conditioned moral sentiments

This alternative route to moral judgement is Prinz's own theory: that moral (dis)approbation is constituted by moral emotions, and moral judgements are constituted by moral sentiments (dispositions to experience particular moral emotions in response to particular actions). Such sentiments are the result of a process of conditioning, whereby parental discipline (punishment, withdrawal of affection) conditions us to have negative emotional responses (fear, sadness) to particular action-types. Prinz holds that the plausibility (he would no doubt opt for a stronger term) of this model speaks against the necessity of empathy for making moral judgements.

I will say a bit more about the positive role empathy might be able to play in Prinz's theory in section 7. First however, recall that Hoffman's third mode of empathic arousal is 'classical conditioning'. Now this is different to what Prinz has in mind but it does point to the beginnings of an explanation of conditioned moral judgement in terms of empathy. Also, Hoffman devotes a lot of attention to outlining the ways in which "empathic induction" as a disciplinary technique increases children's helping and prosocial behaviour. Empathic induction is when "parents highlight the other's perspective, point up the other's distress, and make it clear that the child's action caused it"

[9] For Rawls stability is an essential criterion for principles of justice to be valid.

(Hoffman, 2000, p. 143). The resuling empathic distress could then play the role Prinz reserves for fear and sadness. Further, having the conditioned negative emotional response caused by the victim rather than by authority figures might help avoid issues of state-dependant learning. For example, it is possible on Prinz's model that a child will learn that stealing is wrong only when authority figures might find out and punish them. Perhaps we could explain the difference between career criminals and non-criminals by parental discipline techniques focussed on either punishment or empathy-induction. This is mere speculation and drifting toward the realm of moral behaviour so I won't pursue it here, except to say that although it's possible that Prinz is right and some moral judgements are conditioned responses still there may be reasons to prefer empathy's playing a role in that conditioning process.

5. The importance of patient welfare: the moral/conventional distinction

So far I have attempted show that Hoffman's characterisation of empathy and a proper understanding of Slote's view of empathy's functional role in moral judgement can address Prinz's counterexamples of cases where empathy doesn't make sense, and of alternative routes to moral judgement. This still leaves unaccounted for the substantially longer list of charges relating to empathy's biased arousal and functioning, which Prinz argues void any epistemic or normative value it may have, and make it "bad for morality". Before arguing that there is reason to hold that empathy does play an important epistemic role in moral judgement, I should note that Slote has already done this, and one might follow him in simply biting the bullet and simply embrace empathy's biased nature. Slote holds empathy's inherent biases actually correspond to commonsense, intuitive, normative moral distinctions we typically want to make, and for this reason he prefers the term 'partial' to 'biased'.

For example recall Prinz's objection that relying on empathy as an epistemic guide may cause us to help one, salient, proximate victim while ignoring the cause of the problem: "…the focus on affected individuals distracts us from systemic problems that can only be addressed by interventions at an entirely different scale…with empathy, we ignore the forest fire, while watering a smouldering tree" (Prinz, 2011a, p. 228). Slote's response is a double thumbs-up. He asserts that such partiality tracks our moral intuitions very well (recall the case of deontological override above). Take the 'here-and-now' bias for example. We experience greater empathic arousal for victims who are proximate and contemporaneous to us. Slote offers the example of miners who are trapped underground and observes that "we have a reaction to their plight that impels us to help them and affects us more than the consideration that we could spend the same money we use to rescue them to instead install safety devices in the mine that would save more lives in the long run. The temporal immediacy of the need, the clear and present danger, evokes a stronger empathic reaction than we would have in regard to dangers and lives to be lost at some time in the future" (Slote, 2010, p. 14). Slote maintains that the suggestion that we should ignore the trapped miners in favour of installing safety devices which would ultimately save more lives in the long run would actually horrify most people. I believe Slote has done an excellent job of illustrating how empathy can track our intuitive moral judgements and distinctions (for more see Slote, 2007, 2010). However, I want to approach empathy's epistemic and normative moral value from another direction – that of the moral/conventional distinction. I briefly mentioned this distinction earlier and will now say some more about it.

For some time, researchers have been impressed with the fact that from about three years of age, children begin to distinguish moral and conventional transgressions along several key dimensions (e.g. see Smetana & Braeges, 1990; Turiel, 1983, 2008). When presented with paradigm cases of moral transgressions (e.g. hitting or hair pulling) and conventional transgressions (e.g. chewing gum in class, talking without raising your hand) children judge moral transgressions to be more serious, less dependent on authority, generalisable across contexts, and they typically refer to the victim's welfare in justifying their judgement. For example, when asked about a case of hitting, children tend to say the act is very wrong, would be wrong even if the school rules allowed hitting, would be wrong in other places (e.g. in Japan), and justify its wrongness with reference to the victim's welfare. Chewing gum in class, on the other hand, is rated at less serious, would be ok if the school rules permitted it, may be permissible in other places (e.g. Japan), and its wrongness is justified by reference to rules or convention.

Given this, a *prima facie* obvious argument for empathy's epistemic value immediately emerges – moral transgressions are denounced for the negative impact on the victim's welfare, and empathic processes provide relatively direct and reliable access to knowledge of a victim's welfare. Further, one might even attempt to argue directly from the 'welfare' dimension of the moral/conventional distinction to empathy's normative moral value. It appears to be an empirical fact that moral judgements are justified with reference to welfare. If I have used empathy as an epistemic guide (to the patient's welfare), then I can point to my empathising in justifying my actions (if because I empathised, my actions were such that they showed consideration or concern for the patient's welfare).

5.1 Can we rely on the moral/conventional distinction?

Prinz would not accept this. He challenges the importance of the moral/conventional distinction and suggests that its dimensions may in fact be learned (conditioned) in much the same way as he thinks our moral sentiments are. Although holding the distinction to be a real one, Prinz believes that 'moral' and 'conventional' may in fact be orthogonal dimensions. He argues that although there may appear to be paradigm cases of *abstractly stated* moral rules, once we begin to make the rule specific enough to apply in practice conventional aspects begin to appear (Prinz, 2008, p. 385). Consider a moral rule such as 'do not harm a member of your in group'. Formulated thus, it does not have any apparent conventional component; however Prinz argues that all cultures exhibit exceptions to such rules. For example harming members of one's in-group is permissible in initiation rites, or in sporting events. Prinz argues that it is ultimately cultural convention which determines the scope of harm prohibitions, and harm norms cannot be fully specified without reference to contingent features of culture.

Thus according to Prinz harm norms (for example) are authority contingent. In this culture it is morally wrong to scar a teenager's face with a stone tool but in another culture where scarification is embraced it will be morally permissible. Likewise, says Prinz, rules which are patently conventional have moral dimensions. For example it is conventionally wrong to wear shoes inside in Japan. Yet a failure to comply with this conventional rule is a form of disrespect, and respecting others is a moral precept. So Prinz doubts that the four dimensions of the moral/conventional distinction actually carve out a domain which deserves to called 'morality'. Rather, he concludes that the moral and conventional domains apply at different levels of abstraction (Prinz, 2008, p. 386) and thus would deny that we can infer empathy's epistemic and normative value directly from the moral/conventional distinction. For Prinz, moral rules are simply those rules toward which we have conditioned moral sentiments.

I remain unconvinced by Prinz's arguments. Whilst admittedly it seems obviously true that an abstract rule such as 'do not harm a member of your in-group' will have conventional exceptions, rules at this level of abstraction do not reflect the items typically presented in empirical assessments of the moral/conventional distinction. Most of the interest in the moral/conventional distinction stems from the fact that children seem to develop this capacity at a young age. Consequently much of the research samples young children, and assessment items frequently involve transgressions which take place in the classroom or schoolyard. It is important to remember that researchers such as Turiel are not interested in determining whether children can *label* transgressions as moral or conventional; rather, the assessment items are carefully selected *as* paradigm cases of moral or conventional transgressions – they come pre-labelled – and what is of interest is whether children's judgements regarding the seriousness, authority independence, etc of the transgressions reliably tracks this preconceived division. Moral transgressions are not presented in the abstract form 'Billy harmed a member of his in-group, is that wrong?', nor do they involve elaborate social mores or institutions, but rather are simple scenarios involving concrete transgressions such as schoolyard hitting or hair-pulling. Some studies have taken place in schools and used observed events as assessment items. Such cases are certainly not abstract. And they are obviously specific enough to apply in practice as, well, they literally have. Turiel reports that of 33 recorded events classified as moral[10], "48% pertained to issues of fairness and rights (e.g., taking another's property, only sharing with some, revealing a secret about another); 40% to physical or psychological harm (e.g., hitting, name-calling); and 12% to actions taken to prevent unfairness" (2008, p. 140). No mention of face-scarring or cage-fighting, however.

Indeed there seems to be a common theme running through Prinz's examples of conventional exceptions to moral rules; that of consent. In sporting events and (arguably) in initiation rites the participants know that they will likely be harmed and yet consent to participate anyway. Perhaps consent has some bearing on issues of welfare insofar as the psychological and emotional state of someone being harmed would differ greatly depending on whether or not they consented to that harm. Consider the psychological state of a boxer compared to that of someone being beaten on the street (or a child being hit in the playground). It seems that issues of consent may be relevant to our judgements in such cases. Pursuing this thoroughly would take us too far afield, but I will point out that Slote's theory could accommodate such a view. In cases of consensual harm, an agent's causing harm is consistent with their having empathised with the patient.

So I think that given negative judgements of moral transgressions are typically justified with reference to the victim's welfare, the moral/conventional distinction provides good reason to attribute some epistemic and normative moral value to empathy, and thus that it provides support for the first assumption (the importance of the patient's welfare) of an agent-centred empathic moral theory. Accepting that Prinz may remain unconvinced however, I now want to move on to consider the second assumption of an agent-centred empathic moral theory, that an agent's intentions are somehow relevant to our moral judgements.

[10] Events were coded by four trained independent observers.

6. The importance of agential intent: empathy the moral faculty

Some moral psychologists pursue what is known as the 'linguistic analogy' and take the moral/conventional distinction, paired with a 'poverty of the stimulus' argument, to support the notion that we possess an innate moral faculty. The idea of 'innateness' certainly has its detractors (see e.g. Griffiths (2002) and Griffiths & Machery (2008)). Prinz is one for obvious reasons. However that debate rages elsewhere, and I'd like to talk past these people for a moment and speak instead to those of a moral nativist persuasion.

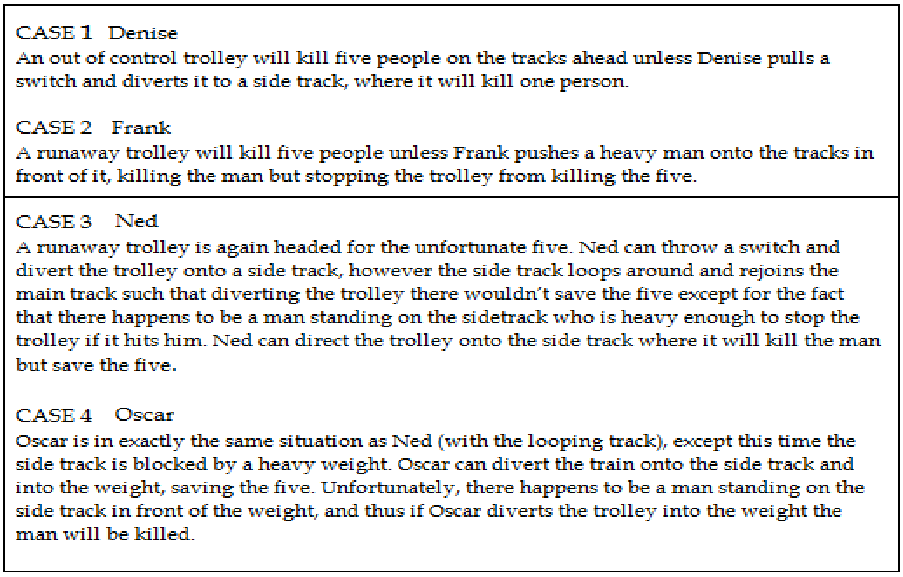
6.1 The linguistic analogy

First considered by Rawls (1971), the linguistic analogy is based on work done by Noam Chomsky and other linguists who posit an innate 'language acquisition device' (Chomsky, 1965). Put very simply, the idea is that there is an insufficient amount of positive and negative evidence (a poverty of stimulus) available to children to explain their linguistic competence in purely behaviourist terms (cf. Skinner, 1957), and thus they must be born with some innate brain mechanism which allows them to acquire language quickly and accurately in spite of this 'poverty of stimuli'. Further, according to Chomsky et al a set of abstract, unconscious rules or 'universal grammar' underlie all human languages. All human beings are supposed to have the same universal grammar, and although variation occurs within the universal grammar (i.e., different languages) it also limits what form a human language may take.

Some authors, e.g. Dwyer (1999, 2006, 2008) Hauser et al (2008) and Mikhail (2000), propose that we should understand children's emerging moral competence in similar terms. They argue that there is a poverty of the *moral* stimulus, and that we can best understand moral competence as underpinned by a moral faculty, analogous to the language acquisition device and subject to principles analogous to a universal grammar: a universal moral grammar; a set of unconscious operative principles which permit moral variation but also constrain what form human moralities can possibly take (and thus what judgements they can possibly issue). Dwyer asks us to consider the moral judgement that 'it is good to torture small babies for fun' and suggests that such a judgement "has the feel of something no "normal" moral creature could generate" (Dwyer, 2008, p. 414). She holds that the reason for this is that certain 'cognitive movements' are simply not allowed by the underlying principles of the universal moral grammar.

6.2 Universal moral grammar and the doctrine of double effect

The discovery of these principles and constraints is taken to be an empirical matter, and to this end Hauser et al (2007) probed people's judgements of moral permissibility in paired hypothetical moral dilemmas. They hypothesised that moral judgements are mediated by an unconscious appraisal system which considers causal and intentional properties of human actions. They were interested in contrasting these unconsciously operative principles with subjects' expressed principles (their justifications of their judgements). Subjects from a wide demographic and international range (N≈ 5,000) were presented with nineteen 'trolley dilemmas'. These included two pairs in which the cases differ in only one key aspect: whether harm caused is intended (i.e., a necessary means to the agent's end) or merely foreseen (see fig. 2). Hauser et al wanted to see if subjects' moral judgements were sensitive to the doctrine of 'double effect'.



CASE 1 Denise
An out of control trolley will kill five people on the tracks ahead unless Denise pulls a switch and diverts it to a side track, where it will kill one person.

CASE 2 Frank
A runaway trolley will kill five people unless Frank pushes a heavy man onto the tracks in front of it, killing the man but stopping the trolley from killing the five.

CASE 3 Ned
A runaway trolley is again headed for the unfortunate five. Ned can throw a switch and divert the trolley onto a side track, however the side track loops around and rejoins the main track such that diverting the trolley there wouldn't save the five except for the fact that there happens to be a man standing on the sidetrack who is heavy enough to stop the trolley if it hits him. Ned can direct the trolley onto the side track where it will kill the man but save the five.

CASE 4 Oscar
Oscar is in exactly the same situation as Ned (with the looping track), except this time the side track is blocked by a heavy weight. Oscar can divert the train onto the side track and into the weight, saving the five. Unfortunately, there happens to be a man standing on the side track in front of the weight, and thus if Oscar diverts the trolley into the weight the man will be killed.

Figure 2: Brief description of trolley dilemmas presented in Hauser et al (2007).

In the pair involving Denise and Frank, Hauser et al found that 85% of subjects judged that it would be permissible for Denise to throw the switch and kill the one, but only 12% judged it permissible for Frank to push the man onto the tracks. In the second pair involving Ned and Oscar, Hauser et al report that 56% of subjects judged that it was morally permissible for Ned to divert the trolley, whereas 72% judged it permissible for Oscar to do so (a smaller but statistically significant difference; $p < 0.001$). Hauser et al conclude that the pattern of subject's moral judgements is sensitive to to the doctrine of double effect[11], and that there is surprisingly little difference in the employment of this principle across possible sources of variation (different e.g., different religions, demographics (Hauser, et al., 2007, pp. 15-16)).

6.3 Dissociation between judgement and justification

Just as interesting is Hauser et al's report that when many of the subjects whose judgement differed within a scenario pair were subsequently unable to provide adequate justification for this difference

[11] Although they consider that the pattern of responses in the first pair could also be due to redirection versus introduction of threat, or personal versus impersonal harm.

in their own judgements (let alone cite the doctrine of double effect). For example, participants who judged one scenario of a pair to be permissible yet its partner to be impermissible were asked to justify this difference. Their justifications were coded into one of three categories: ‘sufficient’, ‘insufficient’, and ‘failed’. ‘Sufficient’ justifications were those that correctly identified the (or ‘a’, see footnote 11) target principle and stated it as the reason for the difference. ‘Insufficient’ justifications were those which identified some morally arbitrary difference, such as the agent’s gender, as the reason. ‘Failed’ responses were those which made any assumptions beyond what was stipulated in the scenario, those which were blank, or those which admitted to being unable to explain the difference.

Of those subjects whose judgements in the first pair of cases differed, almost half (267) provided ‘failed’ justifications for this difference. Of the remainder, only 30% were able to provide ‘sufficient’ justifications. Of those whose judgements differed between cases in the second pair, 45 out of 68 justifications were coded as ‘failed’ and only 3 subjects provided ‘sufficient’ justification for this difference. Hauser et al concluded that these results support their hypothesis that our moral judgements operate over a causal-intentional cognitive structure, which is sensitive to what we call the ‘doctrine of double effect’, and that such principles, although operative, are not expressed (not available to conscious awareness).

These are certainly very interesting and surprising conclusions (especially if you are an empirical moral rationalist) yet they are compatible with an ethics based upon agent-centred empathy. In all these cases it is Subjects’ (in this case judgers) agent-centred empathy (with their cold indifference, their intention to kill) which causes their negative judgement. And given a process-focussed definition of empathy, this would be true regardless of what the agent might actually feel. Thus an agent-centred empathic morality appears to be compatible with Hauser et al’s hypothesis that moral judgements are sensitive to agents’ intentions. What is more, I believe that moral nativists should be attracted to the idea of viewing our capacity for empathy as a kind of ‘moral faculty’, as empathy checks many boxes which fit well with such an idea.

6.4 The empathy faculty

First, empathic behaviour seemingly adheres to fixed stages of development. For example, Hoffman relates his five modes of empathic arousal to five stages of empathic development in children (I won’t go into them here, but see Hoffman (2000, pp. 6-7, 63-93)) and these stages begin from birth (with ‘reactive newborn cry’; arguably a building block of empathic capacities).

Second, early stages of empathic arousal rely on facial and vocal cues from others. Human infants seem naturally disposed to attend to such things. For example babies look longer at collections of shapes arranged as faces than collection of shapes arranged randomly (Maurer & Barrera, 1981).

Third, empathy appears to have evolutionary roots. Early forms of empathy are (arguably) observed in non-human animals such as rats, monkeys, and others (for examples see Bekoff & Pierce (2009, pp. 85-109) and Preston & de Waal (2002)).

Lastly, viewing empathy as a process, and with Hoffman's theory of 'inductions' in mind (p. 21 above), we have a solution to poverty of the moral stimulus problems regarding (at least one dimension of) the moral/conventional distinction. If empathy is the engagement of particular psychological processes, then we might regard it as something like a skill. If so, then it is something which can be (and is) developed (through practice when we're older, and inductions when we're younger). So we can answer the question of how people (including children) uniformly justify negative judgements of moral transgressions, of varying and novel content, with reference to patient welfare. Rather than having to somehow learn the specific content of various moral transgressions, children develop a skill which they can apply in novel contexts.

Furthermore, practice increases the automaticity of cognitive and motor skills (think of driving a car or adding up change). Empathic processes might develop a degree of automaticity also. If I am right and empathic processes cause moral emotions, then this could explain Hauser et al's results that many people are unable to justify their judgements, as well as results like Murphy et al's regarding sibling incest and disgust. In some cases the empathic processing of agents and patients may occur so quickly that it escapes conscious awareness altogether such that all that subjects are left with is the experience of the subsequent moral emotion (anger, disgust, contempt).

Of course this is mere speculation. But I do think that the above considerations should give those pursuing the linguistic analogy and seeking a 'moral faculty' reason to at least consider empathy as a candidate, and this would involve experiments like Hauser's and Murphy's being constructed with empathy in mind.

7. Empathy and moral progress

Before concluding, there is one further area in which I believe empathy may have a positive role: that of moral progress. In defending his particular brand of subjective moral sentimentalism, Prinz acknowledges what is probably the most strongly-held objection to moral relativism in general: morality's being relative to some intuitively morally arbitrary factor such as culture or response-dependency prohibits us from making cross-cultural moral judgements, or judgements regarding past moralities within our own culture. If morality is relative, then different moralities cannot be better or worse; they are simply lateral repositionings in moral space. But surely we can truthfully assert that the abolishment of slavery in the United States (for example) represents an instance of moral *progress*? In comparing two possible worlds which differ only in that one permits slavery whilst the other does not, we surely want to say that the one which does not is the morally superior of the two. Prinz certainly wants to be able to say this, however moral relativism seems to preclude such ordinal evaluations and this often leads to an outright rejection of relativism in all forms, including Prinz's subjectivism.

In response to this concern, Prinz argues that although different moralities are simply lateral repositionings in moral space and there is no transcendental stance outside our own moral perspective from which to compare moral norms, we can in fact hope to secure moral progress by way of extra-moral criteria: criteria of evaluation which we (empirically) seem to value in non-moral domains. He lists these as consistency, basis in accurate factual information, rules which constitute minimal imposition, facilitate social stability, welfare and wellbeing, exhibit a larger degree of generality and universality, are resistant to genealogical critique, and consistent with our premoral biological norms and states that "each of these points can be treated as a standard of assessment. Each provides a sense in which one rule can be better than another. In that respect, they provide us with tools for measuring moral progress. Of two competing moral rules, the one that does better by these standards will be judged the better rule. This is an empirical claim. The list is subject to empirical alteration" (2007, p. 292).

He illustrates how this may work by taking the example of Smith, a hypothetical, well-to-do, mid-nineteenth century, white American who believes that owning slaves is permissible. In evaluating this moral norm, Smith may come to realise that it is based upon inaccurate factual information (e.g. that black people are biologically inferior), or fails a test of universality. He may realise that slavery leads to social instability. And the practice of slavery obviously impacts negatively on the welfare of the slaves, which conflicts with universality and may reduce Smith's own welfare insofar as he is "naturally sympathetic to the suffering and moved by the biological predisposition to help those in need" (Ibid., p. 294). Thus even though Smith "probably wouldn't become an abolitionist overnight…[he] would probably start to suspect that slavery is bad in an extramoral sense. This change in view would be a first step toward a new set of moral values about slavery — values that Smith would recognize as better. Once he started to form negative emotional attitudes toward slavery on extramoral grounds, he would be primed for moral reconditioning" (ibid.).

Prinz himself realises that evaluating moral progress by way of extra-moral criteria raises a serious *prima facie* concern: it seems to permit what he calls "coherent evil", i.e. a system of values which is "utterly abhorrent" yet does better by these extra-moral values than one which is humane (2007, p. 297). He notes that "those who pursue genocide often construct elaborate moral systems that are more consistent than the liberal moralities they replace" (ibid., p. 299). In comparing two competing moral norms, nothing seems to prohibit an 'evil' morality scoring higher on measures of consistency, universality, etc. In addressing this concern, Prinz reminds us that coherence isn't everything: "we assess moral progress by appeal to welfare, well-being, universality, and social stability" (ibid.). However, it seems to me that it is welfare/wellbeing[12] which ultimately does all the work in cases of coherently non-evil moral progress – it is the only 'non-moral' criterion which stands in direct opposition to what we might pretheoretically consider to be a 'bad morality'.

Social stability, for example, plausibly depends at least partly upon issues of welfare. Prinz explicitly considers the question of whether someone who is not suffering can identify moral dysfunction and push for reform (he concludes that they can; Smith being one such individual. Ibid., p. 293). Yet the reverse of this, that those who *are* suffering will identify moral dysfunction and (ultimately) push for reform, seems quite likely. And the fact that those who might be suffering typically can't make use of orthodox routes to such reform (consider slaves, for example) suggests that their suffering may likely lead to social instability; if not outright revolt, at least a social tension which must be accommodated. Considerations of welfare thus bear heavily upon the issue of social stability. Likewise, Prinz states that it is the negative impact on the slaves' welfare which conflicts with universality, not slavery *simpliciter*. Smith believes that slavery is entirely fair - it's a lottery and those who are unlucky enough to be slaves simply have to put up with it. Nothing here conflicts with universality, as Smith *could* be a slave, he's just lucky he's not. This suggests that a form of slavery which doesn't impact negatively on the welfare of the slaves (if such a thing is conceivable) is universalisable. Again here, welfare plays the key role.

So it seems that welfare is ultimately the only extra-moral criterion incompatible with a coherently evil morality; and we can easily see how the moral/conventional distinction may be relevant here. In the previous section I argued that the welfare dimension of the moral/conventional distinction supports the notion that considerations of welfare are important to us in making moral judgements. This strongly suggests that welfare is not, contra Prinz, an 'extra-moral' criterion. I also argued that empathy plausibly gleans some epistemic and normative value from the this dimension and I would say now that this hints at the value which empathy may have in securing Prinz's moral progress as, ultimately, it is only considerations of welfare which can preclude coherently evil moral 'progress'. Thus although Prinz is strongly opposed to allowing empathy a role in morality I believe that

[12] Prinz seems to use these terms interchangeably.

empathy actually stands as a likely and promising avenue to securing the (non coherently evil) moral progress which Prinz so desires. Empathy may in fact have a positive role to play in Prinz's own theory.

This point is highlighted in a similar example offered by Jonathan Bennett in his famous paper *The Conscience of Huckleberry Finn* (Bennett, 1994). Bennett recounts the moral dilemma which Huck faces whilst helping his slave friend Jim escape to freedom. Huck accepts the culturally predominant morality of his contemporaries (including the permissibility of slavery) as valid, yet his actions (helping a slave to escape), which are motivated by sympathy[13] for Jim, stand in direct opposition to that morality. In the story, Huck feels that he has morally failed and views his actions as a total abandonment of morality and its principles. He doesn't come to question the validity of the moral principles regarding slavery, but rather views his sympathetically motivated aid to Jim as a failure of his own moral character – for Huck, the decision to abandon moral principles amounted to an abandonment of morality in general. However Bennett argues that what Huck fails to realise is that one can live by moral principles yet still have some control over their content; and "one way such control can be exercised is by checking of one's principles in the light of one's sympathies. This is sometimes a pretty straightforward matter. It can happen that a certain moral principle becomes untenable—meaning literally that one cannot hold it any longer—because it conflicts intolerably with the pity or revulsion or whatever that one feels when one sees what the principle leads to" (ibid., p. 303). Thus whilst it is conceivable that a principle endorsing the moral permissibility of slavery could be reconciled with issues of consistency, universality, etc, it is very unlikely that such a principle would survive empathic scrutiny. Bennett's example of Huck reinforces the epistemic moral value of empathy and the way in which it might help secure Prinz's goal of moral progress: empathy might serve as a moral anchor. Although Huck didn't come to reverse his moral judgement regarding slavery, with enough empathic exposure to the negative impact it has on the slave's welfare perhaps he might. At the very least empathy might prevent agents from acting in accordance with consistent "abhorrent moralities" which the agent nonetheless holds to be right. Even Prinz writes that slavery might reduce the quality of Smith's life, "insofar as he is naturally sympathetic to the suffering and moved by the biological predisposition to help people in need" (2007, p. 294). To me, this sounds suspiciously akin to something like empathy.

Conclusion

This paper has been an attempt to meet Jesse Prinz's challenge that empathy is not important for, and in fact is bad for, morality. I admit that I have not succeeded in showing that empathy is

[13] Bennett uses 'sympathy' to "cover every sort of fellow-feeling" (Bennett, 1994, p. 295).

necessary for morality, but that is fine; that is not what I set out to do. What I hope I have done is to demonstrate that allowing empathy a role in moral judgement (in both a descriptive and normative sense) may, in spite of Prinz's objections, hold some promise still. Not only can a process-focussed, agent-centred empathic ethics overcome most of Prinz's objections, it may also in fact have a positive explanatory role to play in the recent philosophical and experimental literature, including Prinz's own sentimentalism.

Lastly, I want to briefly revisit Prinz's worry that, due empathy's inherent partiality, a society based upon an empathic morality would be a bad thing. Setting aside Slote's plausible conclusion that such partialities actually track a lot of our typical moral intuitions, I would stress the point that for a society to embrace an empathic morality is not simply for them posit the construct in a moral theory and then continue along with things, business as usual. Rather, embracing an empathic morality involves facing those partialities head-on, and doing our best to overcome them. Prinz sees the fact that empathy has so many observed biases only as a reason for concern, but the flip side of the coin is that the fact that we know so much about empathy's biased nature means we know a significant amount about what we can do to improve its social application.

To this extent moral education in a society which embraced empathy would involve inductions like those described by Hoffman, but would also require institutions which could facilitate such inductions in contexts beyond empathy's usual partial sphere. The internet, social media, as well as television, film, music and the like are to some extent already working toward this end. Charities display heart-wrenching pictures on TV of people who are starving, or who are afflicted by natural disaster or war. Internet sites like YouTube allow people from all over the world to share their experiences in a real and vivid way. All these mediums can provide material for inductions and work towards increasing the "temporal immediacy of the need, the clear and present danger" in a way that can work toward overcoming the here-and-now bias (for example). I don't want to go so far as to say that we might eliminate empathy's partiality altogether – in fact I think that Slote is correct in asserting that empathy's partiality tracks many of our pretheoretic moral intuitions. Yet I do believe that by embracing an empathic morality in the ways just described we might at least raise the baseline of empathic arousal, thereby widening the net of our empathic concern for the welfare of others. In this way perhaps a morality based in empathy may not be so bad after all.

References

Baron-Cohen, S. (2011). *The science of evil: on empathy and the origins of cruelty*. New York: Basic Books.

Batson, C. D., Batson, J. G., Todd, R. M., Brummett, B. H., Shaw, L. L., & Aldeguer, C. M. R. (1995). Empathy and the Collective Good: Caring for One of the Others in a Social Dilemma. *Journal of Personality and Social Psychology, 68*(4), 619-631.

Batson, C. D., Klein, T. R., Highberger, L., & Shaw, L. L. (1995). Immorality From Empathy-Induced Altruism: When Compassion and Justice Conflict. *Journal of Personality and Social Psychology, 68*(6), 1042-1054.

Batson, C. D., Lishner, D. A., Cook, J., & Sawyer, S. (2005). Similarity and Nurturance: Two Possible Sources of Empathy for Strangers. *Basic and Applied Social Psychology, 27*(1), 15-25.

Bekoff, M., & Pierce, J. (2009). *Wild justice: the moral lives of animals*. Chicago: University of Chicago Press.

Bennett, J. (1994). The conscience of Huckleberry Finn. In P. Singer (Ed.), *Ethics* (pp. 294 - 306). Oxford: Oxford University Press.

Brown, L. M., Bradley, M. M., & Lang, P. J. (2006). Affective reactions to pictures of ingroup and outgroup members. *Biological Psychology, 71*(3), 303-311.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

Darwall, S. (1998). Empathy, sympathy, care. *Philosophical Studies, 89*, 261 - 282.

de Waal, F. B. M. (2010). *The age of empathy: nature's lessons for a kinder society*. London: Souvenir.

Dwyer, S. (1999). Moral competence. In K. Murasugi & R. Stainton (Eds.), *Philosophy and Linguistics* (pp. 169-190). Boulder: Westview Press.

Dwyer, S. (2006). How Good Is the Linguistic Analogy? In P. Carruthers, S. Laurence & S. Stich (Eds.), *The Innate Mind* (Vol. 2, pp. 237-256): Oxford University Press.

Dwyer, S. (2008). How not to ague that morality isn't innate: comments on Prinz. In W. Sinnott-Armstrong (Ed.), *Moral Psychology* (Vol. 1, pp. 407-418). Cambridge: MIT Press.

Eisenberg-Berg, N., & Neal, C. (1979). Children's moral reasoning about their own spontaneous prosocial behavior. *Developmental Psychology, 15*(2), 228-229.

Griffiths, P. E. (2002). What is innateness? *Monist, 85*(1), 70.

Griffiths, P. E., & Machery, E. (2008). Innateness, Canalization, and 'Biologicizing the Mind'. *Philosophical Psychology, 21*(3), 397-414.

Haidt, J. (2012). *The righteous mind: why good people are divided by politics and religion*. Camberwell: Penguin Group.

Hare, R. D. (1999). *Without conscience: the disturbing world of the psychopaths among us*. New York: Guilford Press.

Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language, 22*(1), 1-21.

Hauser, M., Young, L., & Cushman, F. (2008). Reviving Rawls's Linguistic Analogy: operative principles and the causal structure of moral actions. In W. Sinnott-Armstrong (Ed.), *Moral Psychology* (Vol. 2, pp. 107-144). Cambridge: MIT Press.

Hoffman, M. L. (2000). *Empathy and moral development: implications for caring and justice*. New York: Cambridge University Press.

Hume, D. (1740/1978). *A treatise of human nature*. Oxford: Oxford University Press.

Maurer, D., & Barrera, M. (1981). Infants' Perception of Natural and Distorted Arrangements of a Schematic Face. *Child Development, 52*(1), 196-202.

Mikhail, J. M. (2000). *Rawls' linguistic analogy: A study of the "generative grammar" model of moral theory described by John Rawls in "A Theory of Justice".* Unpublished 9967447, Cornell University, United States -- New York.

Murphy, S., Haidt, J., & Björklund, F. (2000). Moral Dumbfounding: When intuition finds no reason. In Preparation, Department of Philosophy, University of Virginia.

Paxton, J., & Greene, J. (Forthcoming). Cited in Haidt, J. *The Righteous Mind: why good people are divided by politics and religion*, p. 69.

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*(1), 1-20.

Prinz, J. (2007). *The emotional construction of morals*. New York: Oxford University Press.

Prinz, J. (2008). Is morality innate? In W. Sinnott-Armstrong (Ed.), *Moral Psychology* (Vol. 1, pp. 367 - 406). Cambridge: MIT Press.

Prinz, J. (2011a). Against empathy. *The Southern Journal of Philosophy, 49*(Spindel Suppliment), 214-233.

Prinz, J. (2011b). Is empathy necessary for morality? In P. Goldie & A. Coplan (Eds.), *Empathy: Philosophical and Psychological Perspectives* (pp. 211 - 229). New York: Oxford University Press.

Rawls, J. (1971). *A theory of justice*. Cambridge: Belknap Press of Harvard University Press.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD Triad Hypothesis: A Mapping Between Three Moral Emotions (Contempt, Anger, Disgust) and Three Moral Codes (Community, Autonomy, Divinity). *Journal of Personality and Social Psychology, 76*(4), 574-586.

Sherman, G. D., & Haidt, J. (2011). Cuteness and Disgust: The Humanizing and Dehumanizing Effects of Emotion. *Emotion Review, 3*(3), 245-251.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.

Slote, M. (2007). *The ethics of care and empathy*. New York: Routledge.

Slote, M. (2010). *Moral sentimentalism*. New York: Oxford University Press.

Smetana, J. G., & Braeges, J. L. (1990). The Development of Toddlers' Moral and Conventional Judgments. *Merrill-Palmer Quarterly, 36*(3), 329-346.

Smith, A. (1759/1971). *The theory of moral sentiments*. New York: Garland Pub.

Tsoudis, O. (2002). The Influence of Empathy in Mock Jury Criminal Cases: adding to the affect control model. *Western Criminology Review, 4*(1), 55-67.

Turiel, E. (1983). *The development of social knowledge: morality and convention*. Cambridge [Cambridgeshire]: Cambridge University Press.

Turiel, E. (2008). Thought about actions in social domains: Morality, social conventions, and social interactions. *Cognitive Development, 23*(1), 136-154.