Empathy, Emotion Regulation, and Moral Judgment

Antti Kauppinen

Trinity College Dublin

Final Draft, May 24, 2013

For Heidi Maibom (ed.) *Empathy and Morality*, Oxford University Press.

Abstract

Empathy's role in moral judgment has not received as much attention as its role in moral motivation. Yet given that emotions have at least a causal influence on moral belief, it is plausible that empathy makes an important difference. However, critics like Jesse Prinz point to empathy's inherent partiality and limitations as reasons to think that it has only a limited explanatory role and that it is normatively problematic. Drawing on the classical sentimentalist tradition in philosophy and recent psychological literature, I defend the view that when empathic reactions are subjected to emotion regulation by reference to an ideal perspective, they are after all fit to play a fundamental role in explaining why we make non-self-interested, interpersonally acceptable moral judgments. Getting along with others puts pressure on us to down-regulate empathy with the near and dear and up-regulate empathy with the distant and different. When we successfully do so, there's good reason to think that the resulting judgments are vindicated rather than undermined.

In recent years, some striking claims have been made about the importance of empathy – roughly, the capacity to share the feelings of others – to morality and prosocial action. Perhaps most notably, Michael Slote (2007, 2010) maintains that empathy is the "cement of the moral universe" that "arguably constitutes the basis of both metaethics and normative ethics" (2010, 4). As inevitably happens with philosophical enthusiasms, there has also been a backlash, even among those who believe emotions are central to moral thought. Within the sentimentalist camp, Jesse Prinz (2011a, b) makes a thorough case against empathy, arguing it's neither constitutively, causally, developmentally, epistemically, nor motivationally necessary for moralizing. Indeed, Prinz argues that empathy is likely to lead us astray in moral thought, however important it is for personal relationships. Shaun Nichols (2004) and

Jonathan Haidt (2012) also emphasize the role of non-empathic emotional responses such as disgust in moral thinking.

The critics of empathy are half-right. It is indeed implausible that our natural empathic responses to the suffering or joy of others either explain or justify our considered moral verdicts. But there is a long sentimentalist tradition, beginning from David Hume and Adam Smith, which emphasizes the importance of *corrected* empathic responses, or what psychologists now call *emotion regulation* in the context of empathy. What I will be arguing in this paper is that *ideal-regulated empathic reactive attitudes*, such as empathic resentment or anger, may after all play a foundational role in explaining and vindicating moral attitudes and judgment. In contrast to much of the literature in this area, I will thus not address the role of empathy in altruistic motivation or morally praiseworthy behaviour, but focus on its role in moral verdicts.

I'll begin by distinguishing different forms of empathy in Section 1. What I say will be somewhat novel, since I highlight the kind of adjustments to immediate empathizing that classical sentimentalists proposed to be crucial to understanding morality. To connect the sentimentalist view with contemporary psychological debate, I briefly survey the literature on emotion regulation. What I call *regulated empathy* is a broadly affective response to the perceived situation of another that is regulated by reference to an ideal perspective. While some recent work on the role of hedonic empathy has emphasized the importance of emotional regulation in sympathy and prosocial behaviour (e.g. Eisenberg 2000), the role of regulation in reactive empathy (such as empathic anger) and moral judgment has remained largely unexamined. In Section 2, I examine how and why the classical sentimentalists believed empathy is regulated in the context of moral thinking, and how it contributes to explaining our moral judgments. Roughly, they believed that given the practical function of moral sentiments in reducing social conflict, we must try to discipline our empathetic responses in a way that guarantees others doing likewise can share them. They thus appeal to regulation by reference to an ideal perspective, such as that of an impartial spectator. I formulate the core idea in terms of a hypothesis I call Neo-Classical Explanatory Sentimentalism (NCES). It says that the best fundamental explanation of variation in moral judgment, in particular the extent to which we praise actions and endorse norms that go against our self-interest, appeals to variation in ideal-regulated reactive empathy.

In Section 3, I turn to comparing NCES to competing sentimentalist accounts and responding to objections. The accounts I consider appeal to unregulated empathy (Slote 2010), or culturally transmitted norms that resonate with emotions (Nichols 2004, Prinz 2007). I argue that NCES emerges as a strong contender for explaining why we judge as we do. In the concluding section, I briefly make the case that a regulated empathy account offers a *vindicating* rather than debunking explanation of certain emotion-based beliefs. Insofar as it is indeed ideal-regulated empathy that fundamentally underlies our moral judgments, they may well be in good order, in spite of the partiality of immediate empathy.

1. Empathy: Immediate and Regulated

In this section, my aim is to make clear what I mean by empathy and highlight the importance of regulating our empathic responses in a way that reliably avoids social and emotional conflict. Roughly, empathy makes the feelings of others our own, and interpersonal emotional conflict motivates up- or down-regulating the empathic feelings.

The term ‘empathy’ is used for a number of related phenomena. In a useful article, Daniel Batson (2009) differentiates between eight different senses. I will try to do with less, as my aim isn’t to chart everything people have thought fit to label this way. But some basic distinctions have to be made to be clear about the kind of empathy I will be talking about – especially as it does not even figure among Batson's eight types! In line with most contemporary literature, I'll use the older term 'sympathy' for *concern* for another (see Darwall 1997, Sober and Wilson 1998). As Nancy Eisenberg puts it, sympathy involves "feelings of sorrow or concern for the other" and "the other-oriented desire for the other person to feel better" (Eisenberg 1991, 129). It is one possible consequence of empathizing with another's negative feeling.

One important phenomenon in this conceptual region is *cognitive empathy* or *perspective-taking*. By cognitive empathy, I mean the capacity or process of *knowing* what another wants, believes, or feels as a result of placing oneself in her situation. There may be many ways of coming to know what others think. It is plausible that at least one of them is imagining how I would myself think in their position (self-focused cognitive empathy), or what I would think in their position if I shared their background beliefs, desires, values, and emotions (other-focused cognitive empathy). Cognitive empathizing may well be best cashed out in terms of *simulating* another's reactions (Gordon 1995, Goldman 2006). In any case, it is not what people in ordinary talk mean by empathy, since it issues in a belief about another's states, not any kind of emotional reaction.

Empathy, as ordinarily understood, is *affective empathy*. Affective empathy is feeling the way another feels, or having a congruent emotion, *because* the other feels that way.[1] Thus defined, it is a success notion: if I feel sad because I take another to feel sad, although she is in fact happy, I don't empathize with her. Nevertheless, it is fair to say that my feeling of

[1] It will not do to define empathy in *normative* terms, as Simon Baron-Cohen does: "Empathy is our ability to identify what someone else is thinking or feeling and to respond to their thoughts and feelings with an *appropriate* emotion." (2011, 16; my emphasis) This trivializes the task of arguing for the normative importance of empathy – of course we should respond to the state of mind of others with an appropriate emotion. It also vitiates Baron-Cohen's own claim of having a scientific measure for empathy, given that the appropriateness of emotion isn't a matter of science.

sadness is *broadly* empathetic, since it results from exercising the capacity for empathy. (This distinction will be somewhat important in what follows.) Within affective empathy, we can make further distinctions on the basis of the kind of affective response empathized with. One kind of affective empathy is *hedonic*: we take on the joy, happiness, pleasure, pain, or sadness of another. This is the kind of empathy that most psychological research and measures have focused on.

In my view, however, our ability to take on another's *reactive attitudes*, such as resentment and gratitude, is what is crucial for morality. As Peter Strawson famously put it, these attitudes are "natural human reactions towards the good or ill will or indifference of others towards us, as displayed in attitudes and actions" (Strawson 1962/1982, 67). They are not responses merely to the harm or benefit that results from what others do to us – after all, such consequences might be accidental or incidental. What matters for resentment, say, is instead the disregard or disrespect that the other displays towards us. Put differently, what triggers reactive attitudes is what Kant called the 'maxim' of the action. A maxim is the agent's underlying principle that specifies what she does and to what end, such as "I will slow down in order to avoid splashing the pedestrians". I'll call taking on another's reactive attitudes *reactive empathy*. In the empirical literature, it has been studied mostly in the context of *empathic anger* (Hoffman 1987, Vitaglione and Barrett 2003), which is one possible manifestation of resentment or indignation. Psychologists sometimes appear reluctant to think of anger as a moral response, as it is strongly linked to aggression. But responding to impermissible behaviours with a negative sanction is absolutely central to morality – from a practical point of view, it is its *raison d'être*.

There are different mechanisms whereby the feelings of others are transmitted to us. Some are cognitively undemanding and can be found in other species (see e.g. de Waal 2008), and others involve inference or association (Hoffman 2000). However, what may be the paradigm kind of affective empathy involves cognitive empathy as well. In what we might call *combined empathy*, we come to have an emotion congruent with another's situation *as a result of* imaginatively placing ourselves in another's situation. In the following, I will use the term 'empathy' as shorthand for combined empathy. This is simply for reasons of convenience: the other forms of empathy I've mentioned are quite rightly so labelled. It is nevertheless very important to keep them apart. For example, psychopaths may have cognitive empathy, but little or no affective empathy. They can figure out what others think, but are untouched by it. Autists, it seems, have affective empathy, but little or no cognitive empathy – they have to attribute feelings to others by explicit theory, but are then capable of sharing them.[2]

(Combined) empathy is *self-focused* when I imagine being myself in your situation, and as a result have a feeling suited for your situation rather than mine. In this case, my feeling may differ from what you actually feel (so my feeling may be only broadly empathetic). Empathy is *other-focused* when I imagine being in your situation *as* you – with your goals, beliefs, and character – and feel the way you do.[3] In terms of simulation theory, what happens in other-focused empathy is not just that I run my own psychological systems with inputs taken from your situation, but I also adjust the settings of those systems themselves. Consequently, the outputs may be different. Maybe I'm thin-skinned, so that hearing a racial epithet would wound me in your situation. But I know you're not, so I simulate as a thick-skinned person, and don't feel vicariously hurt by what was said to you.

Empathic responses can come apart from another's feelings when the latter are based on false belief. Suppose you believe that your partner cheated on you last night. I know this to be false, having spent the entire evening gambling with him. You may be angry, and I will

[2] This crucial distinction is lost when Baron-Cohen (2011) lumps psychopaths and autists together as having "zero degree" of empathy.

[3] For the distinction between self-focused and other-focused empathy, see e.g. Gordon 1995 and Oxley 2012.

empathize with this feeling if I place myself in your shoes with your beliefs as inputs – I'll be vicariously angry. Yet it makes little sense for me to be angry with your partner, knowing he's innocent. Or: you thought your partner was innocently gambling with me, but I know he was cheating on you. Surely it is a kind of empathic response on my part to be angry with your partner (cf. Hoffman 2011). Let's say my empathy is *truth-adjusted* when I have an emotional response as a result of simulating being in your situation with true beliefs (as far as I know) as inputs. It is clear from the cases that truth-adjusted empathy may result in emotions that are not congruent with your actual emotions, but it is still a form of empathy, broadly speaking. It is not what is now typically called sympathy, since it need involve no concern for you, or desire to make you feel better – after all, you may feel quite good about your partner in your blissful ignorance.[4]

Regulated Empathy

Empathizing with others presents particular challenges in *conflict situations*, which frequently call for some form of emotion regulation. Consider the following scenario:

> *Rich Man, Poor Man*
>
> A beggar is sitting on the street with all his possessions. As a man in a fine suit approaches, he holds up his cup and says politely "Sir, would you have a coin for a cup of coffee?" The rich man says "Not for you, my friend" and brushes past. "Jerk," mutters the beggar resentfully.

If I empathize with the beggar, I will predictably resent the rich man's behaviour to some extent. But if I corner him later at the casino to complain, he might well object that he's worked very hard for what he's got and has no time to stop for every beggar. It may seem hard-hearted, but refraining from pitying the poor encourages them to practice self-reliance in the land of opportunity. If I were in his shoes, I would not give to beggars either, he might

[4] In my view, Sober and Wilson (1998, 234-5) thus confuse things when they use the label 'sympathy' for what I've called truth-adjusted empathy.

say. If I empathize with the rich man, I might even resent the poor man for living off the hard work of others, and certainly won't disapprove of the rich man's behaviour. So empathizing with both results in conflicting sentiments towards the same action.

Emotional conflict is an everyday fact of life. In such cases, it is impossible for a third party to take on both opposing feelings as her own, except at the cost of internalizing the conflict. She is bound to *regulate* her empathic responses somehow. How might one do this? Emotion regulation (or more broadly, emotion-related regulation) has become a major topic in psychology in the last decades, along with self-regulation in general (see Vohs and Baumeister (eds.) 2010). Emotion regulation involves "goal directed processes functioning to influence the intensity, duration, and type of emotion experienced" (Guyrak, Gross, and Etkin 2011, 401). There are various questions to ask about such processes. In the interest of making headway towards understanding how empathic emotions might be regulated, I will focus on two main questions about emotion regulation in general: How are emotions regulated, and why?

The first question concerns the *strategies* of regulation. There are various ways to categorize them. James Gross's (1998) well-known proposal is to distinguish between different stages of an emotional episode: the eliciting situation, its attended features, their appraisal, and the behavioural, experiential, and physiological response tendencies. Emotions can then be regulated, first, by *situation selection* or *situation modification* with a view to avoiding or generating emotional responses in oneself. Some do not count such situation management as a species of emotion regulation, strictly speaking. But the next stage, *attentional deployment*, is by consensus an important strategy. Well-known variants include distraction (focusing on something other than the emotion-elicitor), concentration, and rumination (attending to one's feelings as well as their causes and consequences) (Gross 1998, 284). Another cognitive strategy is *reappraisal* of the emotion-eliciting features.

Ochsner and Gross (2008) distinguish between two main variants: reinterpreting the stimuli (for example, imagining an arousing image is a fake) and distancing from stimuli “by adopting a detached, third-person perspective” (ibid., 154). The second kind of strategy will be important in what follows. For Gross, the final stage is *response modulation*, which includes direct suppression of emotion, modifying its expression with a view to changing the underlying emotion, or causing physiological changes with this purpose, for example by using drugs or alcohol. These strategies are not all equally fruitful. For example, Gross and John (2003) found that habitual use of reappraisal strategy was related to high well-being and interpersonal functioning (such as having close relationships and emotional and practical social support), while habitual suppression had opposite consequences.

An important distinction that cross-cuts between these strategies is between *explicit* and *implicit* emotion regulation. Explicit emotion regulation involves conscious effort to change one’s emotional state and thus requires some level of awareness of one’s state and insight into what might change it. Recently, many have argued that emotion regulation can be an automatic, System 1 process as well as a conscious effort (Bargh and Williams 2007; Guyrak, Gross, and Etkin 2011). Such *implicit emotion regulation* occurs without effort or conscious awareness. It can be the result of habit. As Guyrak, Gross, and Etkin point out, “frequent use of a given explicit strategy can quickly render the initiation of the strategy more implicit during regulation, thus making it more implicit over time” (2011, 405).

The second general question about emotion regulation concerns the reasons for doing so. I’ll distinguish between intrapersonal and interpersonal reasons. *Intrapersonal* reasons include, most centrally, avoiding hedonic discomfort. They motivate us to reduce the intensity and frequency of unpleasant emotion, and increase positive emotions. There may also be reasons deriving from maintaining integrity, which may have the opposite effect in some contexts (Koole 2011). Unpleasant emotions may also be up-regulated because they are useful to achieve a goal – for example, anger can make success in a competitive situation more likely (Tamir 2009; Tamir and Ford 2012). *Interpersonal* reasons have to do with facilitating interaction with others. Our emotions make a difference to what we do and what it's like to be with us, and thus influence our relationships with others. Psychologists who emphasize interpersonal reasons argue that the need for emotion regulation arises from conflicting goals that different people have (Campos et al. 2011).

What is the role of emotion regulation in empathy? In the past, this question has been addresses in the context of explaining helping response. Nancy Eisenberg has long argued that emotion regulation is one determinant of whether empathic arousal results in sympathy or in personal distress. The core idea is that unregulated empathic response may be so strong that the person focuses on relieving her own situation rather than on the other's problems. This depends in part on the individual's general level of emotionality (intensity and frequency of emotional episodes) (Eisenberg and Fabes 1992). This hypothesis has received modest support from various empirical studies (Eisenberg 2000).

While the data concerning the relation between emotion regulation and hedonic empathy are important in the context of understanding morally praiseworthy motivation, they do not address my present question, which is the regulation of reactive empathic responses to *conflict situations*. They are precisely the sort of circumstances in which moral judgment is typically called for, and thus crucial for understanding the role of empathy in moral judgment. There are both intrapersonal and interpersonal reasons to regulate our empathic responses to conflict situations. The intrapersonal reasons derive from the discomfort of experiencing conflicting emotions. The interpersonal reasons derive from the potentially destructive social conflict that results from conflicting reactive attitudes and resulting behaviour.

As far as intrapersonal conflict goes, any regulatory strategy may be effective. It may be easy enough to empathize with the near and the dear and the similar, and put the suffering of others out of mind. In Rich Man, Poor Man, I may focus on the cocktails I'm drinking with the rich man or join in his rationalizations to ignore the plight of the beggar. I will no longer feel torn about the situation. Crucially, however, such strategies for managing the intrapersonal conflict only make the *interpersonal* conflict worse. For an extreme example, consider the Israeli-Palestinian conflict. On both sides, many people respond with extreme empathy to the plight of their own while ignoring or rationalizing away ("they brought it on themselves" etc.) the suffering of those on the other side. As a result, inter-community conflict is heightened, with consequences known to all.

It is a core insight of the classical sentimentalist tradition that there are ways to regulate our empathic responses in a way that robustly reduces interpersonal conflict. Roughly, we modify our empathic responses so that they could be non-accidentally shared by anyone doing likewise. In practice, this means *counteracting our natural empathic biases*. For example, we must *down-regulate* our empathic reaction to the treatment of those who are personally important to us or similar to us, and *up-regulate* our empathic reaction to the treatment of those who are distant or different. When we do so, we respond to what is done to someone from what Hume termed 'the common point of view', a perspective anyone could come to share (see below). The classical sentimentalist accounts of this process appeal to regulation by reference to an *ideal*. Roughly, we down- and up-regulate in the most socially adaptive way when we reappraise the situation while abstracting away from our particular interests, relationships, and expectations – we look at it from the perspective of a *sympathetic impartial spectator*. This will dampen some empathic responses and strengthen others. The resulting reactive attitudes will then be such that anyone can come to share them, if they're willing to be equally reasonable. This means they are *justifiable* to others in the sense that they can't reasonably object to them.

I'll call this kind of empathy *ideal-regulated* empathy. It is a broadly affective response to another's perceived situation that is regulated by reference to an ideal perspective. My hypothesis, briefly, is that the most effective strategy for such regulation involves both refocusing attention to what things look like from the perspective of each of those affected by an action, and reappraising the meaning of the action in the light of expectations that any social actor as such could be expected to share. Refocusing attention impartially predictably increases empathic responses congruent with the situation of those different from or opposed to us. The reappraisal involves detachment from personal ideals, social identifications, and goals that others may not share, and predictably decreases empathic responses to those similar to us. Such ideal regulation need not be a conscious process, although it may sometimes be such – people can ask themselves how an indifferent but well-meaning bystander, or just anyone, or maybe Jesus would respond. Taking on a more or less impartial perspective before reacting with blame or praise is something that can become habitual and automatized. There is reason to think that people would converge on similar ideals, given what works best to reduce interpersonal conflict – it is not exactly alien to common sense that taking a step back from our egocentric (or ethnocentric) perspective before reacting emotionally tends to defuse tension.

Since regulation by reference to an ideal has not, to my knowledge, been identified in the psychological literature, it has not been studied either. But there are studies that seem to tap into related phenomena. For example, Eran Halperin and co-authors have been investigating emotion regulation in the context of intractable conflict, in particular the dispute between Israelis and Palestinians (Halperin, Sharvit, and Gross 2011). In one recent study (Halperin et al. in press), Israeli participants were shown a PowerPoint presentation that would predictably induce anger towards Palestinians – most plausibly, in my view, by way of empathic identification with Israeli victims of Palestinian actions. In reappraisal condition, participants were asked to respond to the slides "like scientists, objectively and analytically—to try to think about them in a cold and detached manner" (ibid., 2). This is evidently not the same as responding like an impartial spectator, but the instructions similarly call for detaching from particular identifications and values, and can thus be expected to down-regulate empathic anger for one's own and possibly up-regulate empathy for the other side (although this is less likely). In the experiment, participants in the reappraisal condition did indeed report less anger than participants in the control condition, and more support for conciliatory policies toward Palestinians (ibid.). A follow-up experiment applied the same design to a real-life situation (the Palestinian bid for UN membership) and found a similar effect, extending to five months after the manipulation.

What the Halperin studies suggest is that people are indeed capable of regulating their empathic responses by reference to an ideal (in their case detached objectivity), and that this reduces natural empathic bias, improving the odds of conflict resolution. (It is natural to assume that there is more social pressure outside experimental context to adopt such a stance in the case of ingroup conflict.) The studies did not, unfortunately, ask for the participants to morally evaluate outgroup behaviour or possible policies toward the outgroup. According to the sentimentalist tradition in metaethics, change in empathic responses predict changes in moral judgment, other things being equal. It is to this tradition that I will turn to in the next section.

2. Classical Sentimentalism and Regulated Empathy

One can be a sentimentalist in ethics about many different things – moral metaphysics, moral judgments or concepts, moral epistemology, and so on (Kauppinen 2013). *Explanatory sentimentalists* hold that our moral judgments are fundamentally explained by our emotional responses to non-moral facts. By 'fundamentally explained' I mean that even if not every individual judgment is made on an emotional basis, the belief or norm or disposition on the basis of which it is made traces back to an emotional response, which may be that of another subject. Explanatory sentimentalists disagree about the nature and ultimate explanation of the emotional response, however. What is distinctive of the *classical sentimentalism* of David Hume and Adam Smith is that they believe moral approbation and disapprobation is the outcome of empathizing with hedonic states or reactive attitudes, and more or less successfully regulating the resulting response by reference to an ideal perspective. Variation in people's moral judgments, holding the situation and factual beliefs constant, is thus fundamentally explained by variation in emotional dispositions, empathy, and regulation. (This is consistent with some variation being explained by culturally transmitted norms, which trace back to someone's emotional responses to what they believe the facts to be.)

On a sentimentalist view, sentiments of approbation and disapprobation are thus more basic than moral judgments. As sentiments, they are dispositions to feel, notice, and perceive considerations as reasons for action. In my view, *moral* sentiments are best understood as comprised of two elements: first, a *disposition to praise or blame* someone on account of an attitude, action, or act-type, and second, *an authority-independent normative expectation* that everyone share the disposition to praise or blame. The two elements can be dissociated. For example, the blame-emotion of anger need not involve or constitute moral disapproval (think of being angry with a computer). It is possible to blame someone without finding them blameworthy. The difference is in the normative expectation, which may itself be just a disposition to blame those who lack the first-order blaming response, praise those who do, and so on (for this notion of emotional ascent, see Blackburn 1998). In the moral case, this normative expectation isn't contingent on others expecting us to have it, unlike in the case of social norms. So a moral sentiment is a complex emotional disposition. It is important for a sentimentalist account that it does not *presuppose* a moral judgment or appraisal – otherwise the emotion response could not possible explain or justify the judgment. Sentimentalists thus reject judgmentalist theories of emotion (such as Nussbaum 2001).

The question is then: What fundamentally explains why we praise and blame as we do? Why, in particular, do we sometimes approve of actions or character traits that are contrary to our self-interest (such as an enemy's courage), and disapprove of actions that are or would be good for us (such as the behaviour of an enemy turncoat, or stealing a child's lunch money when there's no one to catch us)? Crudely, Hume's answer is this: the more we empathize with the pain of the patient of an action, thus feeling it ourselves to an extent, the more we disapprove of the action. Insofar as we take the source of the pain to be an intentional agent, we come to some extent hate her. If I don't empathize with your pain, I won't disapprove of the person who caused it; if I do, I may disapprove even myself for causing it. The converse goes for actions the benefit others. As Hume summarizes his view, "When any quality, or character, has a tendency to the good of mankind, we are pleased with it, and approve of it; because it presents the lively idea of pleasure; which idea affects us by sympathy, and is itself a kind of pleasure." (T 580) Hume's account thus relies on what I've called hedonic empathy.

However, Hume himself was the first to observe that there is dissociation between what we naturally empathize with and what we morally approve or disapprove of. The two don't vary together. As he put it:

> We sympathize more with persons contiguous to us, than with persons remote from us: With our acquaintance, than with strangers: With our countrymen, than with

> foreigners. … But notwithstanding this variation of our sympathy, we give the same approbation to the same moral qualities in China as in England. (T 581)

This gap between what immediate empathy-based approval and moral belief is also manifest in cases of a kind of moral luck. Best intentions may go awry due to no fault of the actor. There is then no pleasure to empathize with, yet we may consider the agent virtuous.

Hume's response to these challenges is to argue, first, that we have reason, independently of any moral consideration, to regulate our empathic responses by reference to an ideal, and second, that such our degree of success in 'correcting' our sentiments for perspectival distortion serves to fundamentally explain the observed pattern of judgments. On the first point, he appeals to the *practical benefits* of disciplining our empathy in the context of moral judging. Without doing so, we're in trouble:

> Our situation, with regard both to persons and things, is in continual fluctuation; and a man, that lies at a distance from us, may, in a little time, become a familiar acquaintance. Besides, every particular man has a peculiar position with regard to others; and it is impossible we coued ever converse together on any reasonable terms, were each of us to consider characters and persons, only as they appear from his peculiar point of view. (T 581)

Part of the point of using moral language is to guide the way others feel and act towards the people we talk about. If we guide our judgments by self-interest or biased immediate empathy, I praise a kind of action or person while you blame the same kind of action or person, and tomorrow our attitudes may be reversed if our position changes, even if the action or person remains just the same. We do have a language for this kind of approval: we talk about friends and enemies, liking and disliking. But moral language suggests something different, as Hume notes (1751, 260). It manifests an expectation that others will share our praise and blame, and that our attitude hangs only on the features of its target, not our idiosyncratic and possibly fleeting responses to them. In the absence of a point of view that doesn't presuppose particular interests and ideals, we couldn't achieve any *coordination* of blame and praise. Since we are talking about moral attitudes, I would not only blame people for doing things you would praise them for, but also blame those who fail to share the first-order blame. A spiral of mutual resentment and revenge would threaten us, destroying the possibility of social trust and harmony.

So how do we avoid these problems caused by the variation in our natural sentiments “according to our situation of nearness or remoteness, with regard to the person blamed or praised, and according to the present disposition of our mind” (T 592)? Hume says that we solve them by finding a stable point of view that anyone, even rivals, can share:

> T’is impossible men cou'd ever agree in their sentiments and judgments, unless they chose some *common point of view*, from which they might survey their object, and which might cause it to appear the same to all of them. (T 591, my emphasis)

It is sentiments felt from the ‘common point of view’ that guide moral judgments that we can justify to others. Hence, there is social pressure for us to adopt such a perspective, particularly when it comes to judgments concerning in-group members. As Hume puts it, “Experience soon teaches us this method of correcting our sentiments,” (T 582) although the “passions do not always follow our corrections” (T 585). In adopting such a perspective, we regulate our uncorrected empathic responses, with more or less success, and thus arrive at more or less interpersonally justifiable judgments. Hume’s own account of this regulation involves attending to and consequently empathizing with the pleasure and pain of the typical effects of someone’s character traits on those around her, her “narrow circle” (T 602). The problem with this aspect of Hume’s view is, in a nutshell, that if we empathize with the actual hedonic experiences of the narrow circle, we’ll think Silvio Berlusconi is an excellent fellow, since he’s no doubt favoured by his cronies. But we don’t. When we morally evaluate something, we do not just look to the consequences, but also to the quality of the agent’s intentions and motives.

So I believe that Smith’s corresponding impartial spectator account of does a better job of explaining judgments of moral merit and demerit, in particular due to its emphasis on taking on the reactive attitudes rather than hedonic states of anyone affected. Generally speaking, Smith holds that any “passions of human nature, seem proper and are approved of, when the heart of every impartial spectator entirely sympathizes with them, when every indifferent by-stander entirely enters into, and goes along with them” (TMS 81). It is the reactive attitudes of resentment and gratitude that lie at the foundation of moral blame and praise. They are the sentiments that motivate punishment and reward, and are sensitive to the agent’s motives as well as consequences. They appear justified when we take any reasonable person would share them:

> He, therefore, appears to deserve reward, who, to some person or persons, is the natural object of a gratitude which every human heart is disposed to beat time to, and thereby applaud: and he, on the other hand, appears to deserve punishment, who in the same manner is to some person or persons the natural object of a resentment which the breast of every reasonable man is ready to adopt and sympathize with. (TMS 81)

But how do we come to hold that someone is the natural object of the resentment of any reasonable man? Smith’s best account occurs in the context of his discussion of self-directed moral judgment. He famously says that “We endeavour to examine our own conduct as we imagine any other fair and impartial spectator would examine it. If, upon placing ourselves in his situation, we thoroughly enter into all the passions and motives which influenced it, we approve of it” (TMS 129). So, what we do, more or less successfully, is detach ourselves from our natural perspective and look at the situation from the perspective of an imaginary impartial spectator. This is something that “habit and experience” teach us to do “so easily and readily, that we are scarce sensible that we do it” (TMS 157). If, from such a perspective, I empathize with the resentment of someone affected by an action, I take the resentment to fitting, and thus take the agent (who may be myself) to have acted wrongly. Smith places particular emphasis on correcting the “natural misrepresentations of self-love” (TMS 158), but the kind of reflective correction he believes we make will also work for the other distortions that Hume identified. He observes that for most of us, our success in such correction depends on the social context and its demands: "The propriety our moral sentiments is never so apt to be corrupted, as when the indulgent and partial spectator is at hand, while the indifferent and impartial one is at a great distance." (TMS 179) It takes the right kind of social environment for all but the most virtuous to successfully regulate their sentiments by reference to the impartial ideal.

Nevertheless, it remains a fact that as human beings, we simply *cannot* adopt the common point of view in all cases, even if are strongly motivated and have unlimited time at our disposal. We quickly run up against cognitive and affective limits. As Jesse Prinz rightly emphasizes, empathy essentially targets individuals as such (Prinz 2011a). But groups and large numbers of people are also morally relevant. Moreover, it seems that sometimes the right thing to do goes against the grain of empathy. As Hume noted,

> Judges take from a poor man to give to a rich; they bestow on the dissolute the labour of the industrious; and put into the hands of the vicious the means of harming both themselves and others. (T 579)

His response to the problem was to distinguish a class of 'artificial virtues', justice among them.[5] Instead of exploring this intriguing idea, I'll propose a two-stage explanatory model inspired by Adam Smith's remarks on the role of reason in judgment. Smith's idea is that many if not most of our moral judgments result from *reasoning* from principles rather than directly from empathetic sentiments. Nevertheless, our adherence to these *principles* is ultimately explained by disciplined sentiment:

> The general maxims of morality are formed, like all other general maxims, from experience and induction. We observe in a great variety of particular cases what

[5] Roughly, the idea is that if we just focus on the individual case, we may disapprove of the enforcement of property rights. But if we consider the benefits of the convention-based scheme of property rights, invented by people seeking their enlightened self-interest, as a whole, we approve of it by virtue of empathizing with the beneficiaries.

> pleases or displeases our moral faculties, what these approve or disapprove of, and, by induction from this experience, we establish those general rules. (TMS 377)

We are better off regulating "the greater part of" our judgments by appeal to such rules because they would be "extremely uncertain and precarious if they depended altogether upon what is liable to so many variations as immediate sentiment and feeling, which the different states of health and humour are capable of altering so essentially" (ibid.).

Smith's remarks suggest the following kind of two-stage account:

> Stage 1: Moral sentiments that result from more or less successfully ideal-regulated empathy in small-scale paradigmatic instances initially explain beliefs about which act-types are wrong-making (and associate blame with those act-types, other things being equal).
>
> Stage 2: Beliefs about which act-types are instantiated combined with beliefs about which act-types are wrong-making generate beliefs about the wrongness of particular actions (and associate blame with them).

On this model, then, ideal-regulated empathy isn't needed to directly explain beliefs about particular cases. Rather, it explains beliefs about paradigm cases and *pro tanto* principles, which then generate judgments about particular cases when combined with particular facts. We don't have to try to take up the common point of view or engage in systematic reflection unless the various features of an action generate opposing responses. It is also important that we can acquire beliefs about principles from other people, such as parents or other role models. But all the same, these chains of cultural transmission come to an end somewhere – someone or some people must have formed the beliefs to begin with. And the hypothesis is that when it comes to judgments that are likely to be widely accepted, their beliefs will have resulted from more or less successful ideal-regulated empathizing.

3. Comparing Sentimentalist Explanations

In the preceding sections, I have articulated and tentatively endorsed the classical sentimentalist explanation of our moral approval and disapproval in terms of more or less ideal-regulated empathic responses. Call this type of explanatory account that casts classical sentimentalism in contemporary terms *neo-classical sentimentalism*. The main thesis may be formulated somewhat more precisely as follows:

> *Neo-Classical Explanatory Sentimentalism*
>
> The best fundamental explanation of variation in core moral judgments along the dimension of interpersonal acceptability is variation in empathy and exercising emotion regulation by reference to an ideal perspective.

This account predicts that people who are wholly deficient in empathy (such as psychopaths) will only be capable of making interpersonally acceptable moral judgments parasitically on others. It is unsurprising if they are consequently unable to distinguish between moral and social norms, which are also transmitted by others (see Blair 1995). At the other end of the scale, people who are maximally empathic and capable of ideal regulation are predicted to make judgments that are widely acceptable and withstand reflective scrutiny. They may be considered moral exemplars (sages, saints) by others, and may thus be sources of culturally transmitted norms. Finally, I add the qualification *core* moral judgments, but which I mean judgments about what we owe to each other (Scanlon 1998). There are other moral judgments that NCES doesn’t purport to explain, such as condemnation of harmless behaviours that offend a religious sensibility.

To evaluate the thesis that ideal-regulated empathy is the *best* explanation of something, we need to compare it with the most plausible alternatives. I will not attempt a comprehensive survey in this paper. Sentimentalist accounts in general have two things going for them: *simplicity* and *parsimony*. They typically postulate only very simple and straightforward psychological capacities that have a non-moral distal evolutionary explanation. They are compatible with an austere picture of practical reason that is precisely modelled by rational choice theory. Sentimentalists argue that neither reason nor mere understanding tells us to care about the interests of others or respect. Since discussing these arguments is beyond the scope of this paper, I simply proceed on the assumption that they are along the right lines, and that emotions of one kind or another are fundamental in explaining moral thought.

There are, however, many competing sentimentalist explanatory accounts. Some involve appeal to unregulated empathy, while others dismiss or sideline it. In this section, I will compare NCES to two leading contemporary views. First, Michael Slote's account gives *unregulated empathy* pride of place in moral thinking. Second, Jesse Prinz's and Shaun Nichols's accounts appeal to the emotional resonance of *culturally transmitted* norms, and Prinz in particular is sceptical of empathy's role.

a) Slote's Unregulated Empathy View

The most notable recent empathy enthusiast in philosophy has no doubt been Michael Slote. He argues that empathy plays a constitutive role in moral attitudes, and makes both explanatory and justificatory claims on behalf of empathy. The kind of empathy he sees as the "cement of the moral universe" (2010, 13), is what I've called immediate empathy. To begin with Slote's main moral psychological thesis, he claims that empathic feelings constitute moral approval and disapproval:

> [I]f agents' actions reflect empathic concern for (the well being or wishes of) others, empathic beings will feel warmly or tenderly toward them, and such warmth and tenderness empathically reflect the empathic warmth or tenderness of the agents. … [S]uch empathy with empathy … also constitutes moral approval, and possibly admiration as well, for agents and/or their actions. (Slote 2010, 34-35)

Some people’s actions exhibit empathy toward others. This empathy is a warm feeling. When we empathize with the agent, we come to share this warm feeling. And this empathetic warm feeling constitutes moral approval. In contrast, unempathic actions manifest a coldness towards others. Since moral approval and disapproval “enter into the making of moral judgments”, empathy also “enters into our understanding of moral claims” (2010, 53). Slote believes this accounts for why moral judgments have motivating force for us. Roughly, the reason is that the underlying empathic response motivates us to do things that we judge to be morally right (2010, 54).

Given this account of what moral approval is, it’s not surprising that Slote believes empathy can *explain* our intuitions and judgments. More precisely, “differences in (the strength of) our empathic reactions (or tendencies to react) to various situations correspond pretty well to differences in the (normative) moral evaluations we tend to make about those situations” (Slote 2010, 21). For example, we naturally flinch more strongly from causing harm to someone than from allowing the same harm to happen. This, for Slote, basically accounts for the commonsense deontological distinction between the seriousness of the wrong of doing harm rather than allowing it to happen. Slote defends the following normative thesis tying moral status of actions with empathy or its absence:

> actions are morally wrong and contrary to moral obligation if, and only if, they reflect or exhibit or express an absence (or lack) of fully developed empathic concern for (or caring about) others on the part of the agent. (2007, 31)

So even though Slote fully acknowledges that (natural) empathy, even when “fully developed”, is influenced by factors like similarity and distance, he believes that it offers a criterion of rightness.

Does unregulated empathy really play the sort of role in moral psychology Slote claims it does? There is good reason to doubt it. First, empathy is neither necessary nor sufficient for moral approval, either of actions or agents. It's not necessary, since it is possible for us to approve of an action that isn't empathically motivated (so that there is no agential 'warmth' for us to catch). Surely somebody incapable of empathy can do the right thing, and not just for fear of punishment either. I see no reason to deny someone could desire to be morally good without feeling empathy, or that such desire could motivate one. There is something to Kant's insistence that acting out of duty is morally praiseworthy, and in some cases we might approve of it more than acting out of empathy. I don't want you to refrain from taking my things just because you think I'd feel bad if you did so. Second, empathy isn't sufficient for approval, since it is possible for us to disapprove of an action that is empathically motivated. Think of a mother who empathizes strongly with her daughter and therefore elbows everyone else's child out of the way to get her a place in college. (Recall that Slote endorses the partiality of immediate empathy.) The same goes naturally for disapproval. Bad people, much less people doing the wrong thing, need not be cold-hearted. The Tutsi shooting the Hutu in front of his family may be full of empathy towards the victim, but nevertheless choose to follow orders.

The second problem with Slote's account of approval is that moral disapproval isn't necessarily a 'cold' feeling, nor approval 'warm' (whatever exactly these characterizations mean). Indeed, it needn't have *any* particular phenomenal quality, even if occurrent. Further, feelings of coldness or warmth may be caused by one thing or another, such as (perhaps) empathy with someone else's empathic feelings, but that doesn't mean they are *about* anything. After all, literally feeling cold isn't about anything – it's not, for example, about the north wind that causes the feeling. Nor, relatedly, is it clear that Slotean psychological weather conditions motivate us in the way that moral approval and disapproval do. If contemplating someone gives me a chill, why not contemplate something warmer rather than punish or blame the unempathetic person? For all these reasons, moral approval and disapproval are much better understood in terms of reactive attitudes we expect of each other, as discussed above.

What about Slote's empathic criterion of rightness? Focusing on only one of the many cases that he presents may suffice to bring out the problems with it. Comparing the My Lai killings and bombing civilians, he says that since face-to-face killer "demonstrates a greater lack of (normal or fully developed) empathy, than the person who kills from the air" (Slote 2007, 25), it is harder for *us* to empathize with the former, so there is a stronger moral obligation to refrain from killing face to face. I believe Slote is correct in his hypothesis about our natural empathic reactions, but wrong about the normative facts of the case. It is not the case that our obligation not to kill an innocent person is stronger when we can do so face to face than when we can do so by pressing a button far away (let's say, to update the example, by a drone strike).

To see this more clearly, we have to focus on just the relevant difference between the cases. So let's say that in *Face to Face*, a renegade soldier guns down an Afghan family in person, in full knowledge that they are not dangerous to anyone, just in order to relieve anger and frustration. In *Drone*, a renegade drone controller in Miami remotely fires a missile at an Afghan family hut (knowing of but without seeing the family inside), in full knowledge that they are not dangerous to anyone, just in order to relieve anger and frustration. If we try to empathize with the killer in each case, we may indeed catch more of a 'chill' in Face to Face, in which the soldier must surely be more callous or hardened to pull the trigger. If Slote is right, the soldier in Face to Face thus has a stronger obligation not to kill than the controller in Drone. If, on the other hand, we regulate our empathy and put ourselves to the position of the family members as they are about to be killed, or in the position of someone who cares about them, we will, I believe, have just as strong a negative reaction towards the agent in both cases. After all, though the action is different, the intended consequences and purposes are identical (and thus the underlying maxims relevantly similar).

On reflection, I believe it borders on the absurd to think that there is a difference in strength of obligation. Surely the civilians have just a strong a right not to be killed by missile as by bullet! I would object just as strongly to my son being killed by any means whatsoever, the purpose being the same, and I'd be right to do so. So Slote's view must be mistaken. Immediate empathy is no guide to moral rightness.[6]

b) Cultural Transmission Accounts

As I noted in the introduction, Jesse Prinz makes what is perhaps the most systematic case against the importance of empathy for morality. I agree with his claim that empathy isn't constitutive of moral sentiments. But what about explanation? Prinz argues that on-line empathy cannot be *causally necessary* for moral judgment or approval. He notes that there may be so many victims you can't possibly empathize with all of them. On the other hand, we may approve of an action that causes suffering (2011a, 220). Neither of these cases poses a challenge to NCES. First, the two-stage model says that we judge on the basis of more or less ideal-regulated empathy in small worlds, and then make use of principles in large-scale situations, so we need not try to empathize with all. Second, ideal-regulated empathy with, say, the anger of a victim may indeed lead us to approve of punishment that causes the perpetrator to suffer. Once we leave behind the hedonic empathy model, there is no reason to think empathy couldn't result in approval of causing suffering in some cases.

[6] There may nevertheless be some temptation to think that the soldier must at least be a worse person, being more callous or hardened. But I doubt that, too. Killing by pressing a button is cowardly for many reasons. It is said that Hitler was shielded from personally encountering the consequences of his vicious orders. I don't think his reluctance to face the suffering and terror he caused made him any better a person.

Prinz might grant these points, but offers further cases in which it seems empathy cannot explain our judgments: What if you are the victim of a crime yourself? (2011a, 220) And what about sacrificing one person to save five – if we empathize impartially, shouldn't we be willing to do so (2011b)? Yet many people have nonconsequentialist intuitions about such cases. These cases pose a *prima facie* challenge to an empathy-based explanation, but I believe that they can be accounted for. In the first-person case, direct role for empathy is indeed precluded. However, the relevant question is: what explains the difference between non-moral and moral disapproval of someone who harms me? The natural answer, according to NCES, is that I *morally* disapprove of something done to me when I would morally disapprove of the same thing done to *someone else* – when what matters is the action and its motives and consequences, not the fact that I am the victim. And this takes us back to empathy-based disapproval. Second, I argue elsewhere that empathizing with anticipated reactive attitudes predicts nonconsequentialist responses in many cases – although from a hedonic perspective it makes no difference, for reactive attitudes being hurt as a means for someone else's good is very different from being hurt as a side effect or consequences of not making use of another to save me. That's why regulated reactive empathy may lead us to disapprove of sacrificing one to save many.

Another possible objection, more in the vein of Shaun Nichols (2004), is that NCES is cognitively too demanding to account for early emergence of moral judgment in children. Emotion regulation by reference to an ideal perspective certainly goes beyond small children's capacities. But that only entails defective spontaneous (that is, unlearned) moral judgments, according to NCES, not that they couldn't make judgments at all. We should bear in mind that empathic reactive attitudes also emerge at an early age, before or contemporaneously with moral judgment. According to Martin Hoffman, "A simple example of empathic anger is that of the 17-month-old boy in the doctor's office who, on seeing another child receive an injection, responds by hitting the doctor in anger." (Hoffman 1987, 55) Less anecdotally, Kiley Hamlin and co-authors' recent studies with infants and toddlers suggest preference for helpful characters over antisocial characters, and a willingness to punish antisocial ones (Hamlin et al. 2011). I believe that these observations are plausibly manifestations of empathic (proto-) anger towards antisocial characters – more plausibly than manifestations of evaluative judgment, as the authors themselves believe. This is further supported by more recent results, according to which infants don't dislike bad treatment of individuals who are dissimilar to them (and thus are harder to empathize with) (Hamlin et al. 2013).

Finally, it is clear that when Prinz talks about empathy, he has immediate empathy in mind. He says that "[a]s I will use the term, empathy requires a kind of emotional mimicry," and explicitly rules out what I've called truth-adjusted empathy as a form of empathy (Prinz 2011b, 212). And I agree that immediate empathy is problematic. So to some extent, my disagreement with Prinz is merely verbal. But since many of Prinz's arguments are general enough to target regulated empathy, there is also substantive disagreement.

So, it seems that NCES survives the explanatory challenges that Prinz (rightly) raises for simple empathy-views such as Slote's account. But how does it compare with Prinz's own explanatory account? His view is a species of what I'll call the *Cultural Transmission* theory (CT). For Prinz, moral judgments consist in emotional responses. We don't need to empathize when we judge, since "moral response is linked to action-types" (ibid.). His view seems to be that we simply associate a negative response with certain act-types, so that it gets triggered by classification of something as falling under them. But how do we come to associate blame and praise with certain act-types? Prinz appeals to social conditioning: when parents punish a child for something, the child associates fear, sadness, and anguish with that type of action, and is motivated to avoid them (Prinz 2011a, 221; cf. Prinz 2007, 35–37). Alternatively, children may simply imitate the anger of their parents towards something (Prinz 2011b, 229). And why do the *parents* associate blame with theft, for example? Here culture comes into picture. Norms get culturally transmitted from generation to generation, and consequently vary from place to place, although some may be more common due to their resonance with non-transmitted affective responses. On this picture, the emotional responses that constitute judgment are not evolutionary adaptations (unlike for Haidt (2012), for example), but byproducts of exercising capacities evolved for other purposes. For example, guilt is a form of sadness caused by hurting those one cares about and being rejected by those one depends on (such as parents).

Nichols's (2004) variant of CT begins with the assumption that there are a variety of norms that individuals subscribe to. (Unfortunately, he never explains what he takes norms or norm-acceptance to consist in.) Norms prohibit and permit certain actions. When the actions they prohibit are independently emotionally upsetting, we regard the norms as being *non-conventional* – they don't just hold because of some authority says so or locally. Moral norms are a subset of these 'sentimental rules'. When we make moral judgments, we apply sentimental rules – unlike for Haidt or Prinz, no on-line emotion is needed (Nichols 2004, 25–29). This account presupposes that we have norms independently of emotional responses. Instead of asking about the origin of norms, Nichols tells an epidemiological story about why certain norms prevail. The answer is that norms that resonate with our (independent) affective reactions enjoy greater 'cultural fitness', and are thus likely to be passed on from generation to generation (Nichols 2004, ch. 7–8). Some of these affective reactions result from 'contagious distress', which explains why we typically regard (norm-forbidden) harmful actions as morally wrong. (Here there's a place for a kind of immediate empathy in Nichols's view.)

How do the CT accounts compare with NCES? It is evident that NCES has higher explanatory ambitions, since it offers an account of the *origin* and not merely the survival of norms. People originally come to have a moral norm (a normative, authority-independent expectation that everyone refrain from doing something) when they (more or less) impartially empathize with those affected by paradigm instances of an act-type. Although the 'epidemiological' framework that Nichols and Prinz employ is plausible in many cases, when it comes to moral norms, it is very implausible that people start out with a large body of random norms that are then winnowed down to those that resonate with our (culture-independent) emotions. Further, people can judge that something is morally wrong when it goes *against* the norms that have been culturally transmitted to them. This is hard for the sentimental rules account to accommodate. Finally, Nichols's version of CT also shares the weakness of immediate empathy accounts. Appealing to immediate empathetic reactions or personal distress plainly can't explain why we embrace norms that prohibit harming or cheating of those we don't naturally empathize with. The distress of the distant just isn't as contagious as the distress of the near, yet I don't adopt a rule according to which it is less bad to hurt the distant as a means to advancing one's self-interest, say. So again, NCES offers at least some explanatory advantages. I can't claim that the issue is in any way settled at present. More empirical evidence is needed.

4. Conclusion: Vindicating or Debunking Moral Judgment?

In this paper, I've developed the classical sentimentalist hypothesis that empathic sentiments more or less successfully regulated by reference to an ideal perspective fundamentally explain why we make the moral judgments we do. Suppose that this explanation is correct. What does it mean for the *justification* of the emotion-based moral judgments? On the critical side, Jesse Prinz argues that basis in natural empathy *undermines* the justification of moral judgments.

He notes that “our capacity to experience vicarious emotions varies as a function of such factors as social proximity and salience” (Prinz 2011a, 223). As a consequence, “[w]e are grotesquely partial to the near and dear” so that “we use empathy as an epistemic guide at the risk of profound moral error” (Prinz 2011a, 224). There is also experimental evidence that the here-and-now bias of empathy results in judgments we consider unfair on reflection (Batson et al. 1995). Further, when it comes to *justice* in particular, empathy can be misleading. Like some proponents of ethics of care, Prinz contrasts justice with empathy. Finally, Prinz likes to emphasize the ‘dark side’ of empathy. It is easily manipulated, selective, subject to cuteness effects, and prone to in-group biases as well as proximity and salience effects (Prinz 2011b). Indeed, it is intrinsically biased in the sense that “essentially a dyadic emotion, regulating the responses between two individuals, and its function is, arguably, to align the emotions of people in a close personal relationship” (Prinz 2011a, 229). So empathy has at best an incidental link to morality, and misleads moral judgment.

Is there anything to be said in favour of empathy when it comes to justification? Classical sentimentalists say surprisingly little about this topic, although they are not shy to employ language that implies moral knowledge. Hume may have felt that providing justification was beyond his brief as an ‘anatomist’ of morality (T 3.6). When we begin to justify moral belief, we give voice to our own convictions, abandoning the outsider’s theoretical perspective. At the very end of the *Treatise*, Hume nevertheless permits himself to say a few words from within an engaged perspective. He notes that once we see the origin of our moral judgments in (what we would call) empathy felt from the common point of view, we “must certainly be pleas’d to see moral distinctions deriv’d from so noble a source” (T 3.6). Again, our empathy-driven ‘moral sense’ “must certainly acquire new force, when reflecting on itself, it approves of those principles, from whence it is deriv’d, and finds nothing but what is great and good in its rise and origin” (ibid.)

This *coherentist* suggestion is that when we engage in a process of what would now be called *wide reflective equilibrium*, in which we try to find the best fit for our particular and general moral convictions and known psychological and sociological facts, including facts about the origin of our moral convictions (Daniels 1979), we will *reflectively endorse* those moral convictions that result from ideal-regulated empathy. On the other hand, if we come to believe that some moral belief of ours reflects partiality or the influence of mere distance, we lose confidence in it. There is thus a fundamental difference, when it comes to justification, between beliefs based on natural empathy and those based on regulated empathy. The latter kind of beliefs are not subject to the kinds of bias that Prinz points out. Consequently, when we try to get our beliefs to line up in wide reflective equilibrium, we may opt for embracing rather than rejecting judgments based on ideal-regulated empathy. Rather than debunking, we may come to see origin in regulated empathy *vindicating* our core moral beliefs.

I will finish with a final objection from Prinz. In his papers against empathy, he does, in fact, briefly consider a version of Smith's impartial spectator account (Prinz 2011b) and Hume's appeal to the common point of view. Here's what he has to say about the latter:

> As attractive as this idea is to a liberal readership, it is bad psychology. The fact is, we rarely adopt such a point of view, and empathy is probably the greatest impediment. We *can* empathize with members of the out-group but only by making their similarities salient. … But there is no way to cultivate empathy for every person in need, and the focus on affected individuals distracts us from systemic problems that can be addressed only by interventions at an entirely different scale. (Prinz 2011a, 228)

If what I've argued above is on the right track, Prinz here draws a false contrast between empathizing and adopting a common point of view. (For Hume and Smith, it would certainly be inconceivable that we could separate the two.) He is right, to be sure, that natural empathy can be one obstacle to ideal-regulated empathy, given all its biases. But the answer is not to shut out empathy, but to use it wisely – to try hard to bear in mind not only what is in front of our eyes and to rely on general principles when in doubt about one's ability to adopt an intersubjectively sharable perspective. If we are able to do so, empathy – although of a cool and challenging variety that may require looking beyond the individual in front of us – does, after all, merit some of the enthusiasm it has lately received.

References

Bargh, John and Williams, L. E 2007. On the nonconscious regulation of emotion. In J. Gross (Ed.), Handbook of emotion regulation (pp. 429-445). New York: Guilford.

Baron-Cohen, Simon 2011. *Zero Degrees of Empathy*. London: Penguin.

Batson, C. Daniel, Klein, Tricia R., Highberger, Lori, and Shaw, Laura L. 1995. Immorality from empathy-induced altruism: When compassion and justice conflict. *Journal of Personality and Social Psychology* 68 (6), 1042–1054.

Batson, Daniel C. 2009. These things called empathy: Eight related but distinct phenomena. In J. Decety and W. Ickes (eds.), *The Social Neuroscience of Empathy*. Cambridge, MA: MIT Press, 3–16.

Blackburn, Simon 1998. *Ruling Passions*. Oxford: Clarendon Press.

Blair 1995. A cognitive developmental approach to morality: investigating the psychopath. *Cognition* 57, 1–29.

Campos, Joseph J., Walle, Eric A., Dahl, Audun, and Main, Alexandra 2011. Reconceptualizing emotion regulation. *Emotion Review* 3 (1), 26 – 35.

Coplan, Amy and Goldie, Peter (eds.), *Empathy: Psychological and Philosophical Perspectives*. Oxford: Oxford University Press.

Daniels, Norman 1979. Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy* 76 (5), 256–282.

Darwall, Stephen 1998. Empathy, sympathy, care. *Philosophical Studies* 89 (2/3), 281–282.

Eisenberg, Nancy 1991. Values, sympathy, and individual differences: Towards a pluralism of factors influencing altruism and empathy. *Psychological Inquiry* 2 (2), 128–131.

Eisenberg, Nancy 2000. Emotion, regulation, and moral development. *Annual Review of Psychology* 51, 665–697.

Eisenberg, Nancy and Fabes, R. A. 1992. Emotion, regulation, and the development of social competence. In *Review of Personality and Social Psychology: Emotion and Social Behavior* 14: 119–50.

Ford, Brett Q. and Tamir, Maya 2012. When getting angry is smart: Emotional preferences and emotional intelligence. *Emotion* 12 (4), 685–689.

Frank, Robert 1988. *Passions Within Reason: The Strategic Role of the Emotions*. New York: WH Norton.

Goldman, Alvin 2006. *Simulating Minds*. New York: Oxford University Press.

Gordon, Robert 1995. Sympathy, simulation, and the impartial spectator. *Ethics* 105, 727–742.

Gross, James J. 1998. The emerging field of emotion regulation: An integrative review. *Review of General Psychology* 2 (3), 271–299.

Gross, James J. and John, Oliver P. 2003. Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology* 85 (2), 348–362.

Guyrak, Anett, Gross, James J. and Etkin, Amit 2011. Explicit and implicit emotion regulation: A dual-process framework. *Cognition and Emotion* 25 (3), 400–412.

Haidt, Jonathan 2012. *The Righteous Mind*. New York: Pantheon.

Halperin, Eran, Porat, Roni, Tamir, Maya, and Gross, James J. (in press). Can emotion regulation change political attitudes in intractable conflicts? From the laboratory to the field. *Psychological Science.*

Halperin, Eran, Sharvit, Keren, & Gross, James J. 2011. Emotion and emotion regulation in conflicts. In D. Bar-Tal (Ed.), *Intergroup conflicts and their resolution: Social psychological perspective*. New York, NY: Psychology Press, 83–103.

Hamlin, J. Kiley, Wynn, Karen, Bloom, Paul, and Mahajan, Neha 2011. How infants and toddlers react to antisocial others. *PNAS* 108 (50), 19931–19936.

Hamlin, J. Kiley, Mahajan, Neha, Liberman, Zoe, and Wynn, Karen 2013. Not like me = bad. Infants prefer those who harm dissimilar others. *Psychological Science* 24 (4), 589–594.

Hume, David 1739–40/1978. *A Treatise of Human Nature*. Ed. L. A. Selby-Bigge, 2nd rev. edn., P. H. Nidditch. Oxford: Clarendon Press.

Hume, David 1751/1948. *Enquiry Concerning the Principles of Morals*. In H. D. Aiken (ed.) Hume: Moral and Political Philosophy. New York: Hafner Press, 171–291.

Hoffman, Martin 1987. The contribution of empathy to justice and moral judgment. In N. Eisenberg and J. Strayer (eds.), *Empathy and Its Development*. New York: Cambridge University Press, 47–80.

Hoffman, Martin 2000. *Empathy and Moral Development*. Cambridge: Cambridge University Press.

Hoffman, Martin 2011. Empathy, justice, and the law. In A.Coplan & P.Goldie (eds.) 2011.

Kauppinen, Antti 2010. What makes a sentiment moral? *Oxford Studies in Metaethics* 5, 225–256.

Kauppinen, Antti 2013. Sentimentalism. In Hugh LaFollette (ed.), *International Encyclopedia of Ethics*. Oxford: Blackwell.

Koole, Sander 2009. The psychology of emotion regulation: An integrative review. *Cognition and Emotion* 23 (1), 4–41.

Nichols, Shaun 2004. *Sentimental Rules*. New York: Oxford University Press.

Nussbaum, Martha 2001. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.

Prinz, Jesse 2007. *The Emotional Construction of Morals*. New York: Oxford University Press.

Prinz, Jesse 2011a. Against empathy. *Southern Journal of Philosophy* 49, 214–233.

Prinz, Jesse 2011b. Is empathy necessary for morality? In Coplan and Goldie (eds.) 2011.

Scanlon, Thomas 2008. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Slote, Michael 2007. *The Ethics of Care and Empathy*. Oxford: Oxford University Press.

Slote, Michael 2010. *Moral Sentimentalism*. Oxford: Oxford University Press.

Smith, Adam 1759–90/1976. *The Theory of Moral Sentiments.* Ed. D. D. Raphael and A. L. Macfie Oxford: Oxford University Press.

Sober and Wilson 2008. *Unto Others*. *The Evolution and Psychology of Unselfish Behaviour*. Cambridge, MA: Harvard University Press.

Strawson, Peter 1962/1982. Freedom and Resentment. Reprinted in Gary Watson (ed.), *Free Will.* Oxford: Oxford University Press.

Tamir, Maya 2009. What do people want to feel and why? Pleasure and utility in emotion regulation. *Current Directions in Psychological Science* 18, 101–105.

Vitaglione Guy D. and Barrett, Mark A. 2003. Assessing a new dimension of empathy: Empathic anger as a predictor of helping and punishing desires. Motivation and Emotion 27 (4), 301–325.

Vohs, Kathleen and Baumeister, Roy F. (eds.) 2011. *Handbook of Self-Regulation*. New York: Guilford.

de Waal, Frans 2008. Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology* 59, 279–300.