

Sentimentalism

Antti Kauppinen (a.kauppinen@gmail.com)

Trinity College Dublin (and the University of Amsterdam/NWO research project *Good Guides or False Pilots? The Role of Emotions in Moral Judgment*)

For the *Blackwell International Encyclopedia of Ethics*, ed. Hugh LaFollette

Final version, February 18, 2012

Word count: 5896

Our ethical stance is manifest in our attitudes and reactions, in what inspires or outrages us, in what we admire or detest, in whom we praise or blame. For sentimentalists, these human reactions are what is fundamental about ethics. There is no need to assume a moral reality that transcends the natural world or find a Moral Law binding on all rational creatures. Ethics is all about practice. It emerges when, under pressure of social conflict, we come to regulate our emotional responses from a point of view that goes beyond self-interest, and serves to foster mutually beneficial cooperation by guiding our choices. To understand ethics, sentimentalists hold, we need to investigate the practical aspects of evolved human nature as they manifest themselves in a social context. This naturalistic outlook need not mean giving up on notions like truth and objectivity, as long as we can make sense of suitable standards for fitting emotional responses.

On closer examination, sentimentalism divides into a number of logically distinct theses. For *explanatory* sentimentalists, certain emotions fundamentally explain our moral

judgments. For *judgment* sentimentalists, moral judgments themselves are constituted by emotions, or attitudes or beliefs referring to them. For *metaphysical* sentimentalists, moral facts are facts about what causes or merits emotional responses, or mere projections of sentiments. Finally, for *epistemic* sentimentalists, emotions are basic sources of moral knowledge or justification. Thoroughgoing sentimentalists endorse versions of all four theses, but it is possible to accept only some. This entry will discuss all these views in their historical and contemporary incarnations, starting where the pioneers of sentimentalism did, namely with the question of what explains why we consider certain things virtuous or vicious. This is the context in which many themes central to sentimentalism in general are introduced: the practicality of moral thought, the limitations of reason, methodological naturalism, and the appeal to emotions that are in some sense corrected rather than mere knee-jerk reactions.

1. Explanatory Sentimentalism

No one denies that emotions influence moral judgments, for better or for worse. Some ancient philosophers, most prominently Aristotle, emphasized the role of emotions in virtuous thought and action, but considered them supplementary to reason. For genuine explanatory sentimentalists, emotions play a *fundamental* role in the process of moral judging, being necessary either for making individual judgments or for having a capacity to do so. The right kind of emotional origin suffices to account for the nature and content of our moral judgments; it is a non-contingent fact that neither pure reason nor mere understanding of moral propositions will suffice.

Explanation, along with an associated moral epistemology, was arguably the primary concern of the classical 18th century sentimentalists. For example, Hume's (see HUME, DAVID) first question is whether we "distinguish betwixt vice and virtue" by way of ideas

(reason) or impressions (sentiments) (1740/1978: 456). For Smith (see SMITH, ADAM), the main theoretical question in ethics is “[H]ow and by what means does it come to pass, that the mind prefers one tenour of conduct to another, denominates the one right and the other wrong[?]” (1759/2002: 314). In the 20th century, these questions were comparatively neglected, but they have again gained a prominent place in contemporary sentimental work.

1.1 The Rejection of Rationalism

The roots of sentimentalism are in the British empiricist response to the psychological egoists Hobbes and Mandeville, who explained moral judgment in terms of enlightened self-interest, and their Platonist rationalist rivals, according to whom reason or understanding informs us of the eternal and immutable relations or fitnesses in which moral facts consist (see INTUITIONISM, MORAL). Francis Hutcheson (see HUTCHESON, FRANCIS) pointed out that egoist views can’t explain why we approve of what is not in our interest, such as actions of historical characters or the courage of our enemies, and why we fail to approve of what is in our interest, if it springs from a selfish motive (Hutcheson 1728: 136). Against intuitionism, he argued that practical reason only tells us to choose the action that maximizes the satisfaction of our preferences, given the probabilities of different outcomes. It cannot tell us which ultimate ends to adopt, nor lead us to approve actions that are fit for promoting any particular end.

Hume developed this critique of rationalism further. According to his *Practicality Argument*, our moral evaluations influence our actions and reactions to the actions of others (“Morals excite passions, and produce or prevent actions”), while reason lacks such power (“Reason can never immediately prevent or produce any action by contradicting or approving of it”); “Moral distinctions, therefore, are not the offspring of reason.” (Hume 1740/1978: 457). Here is one natural reading of this influential argument in contemporary terms:

1. Moral judgments are intrinsically motivating. (Moral Judgment Internalism) (see MOTIVATIONAL INTERNALISM)
2. Reason(ing) alone does not motivate. (Inertia of Reason Thesis)
3. So, moral judgments cannot be conclusions of reason. (Anti-Rationalism)

The first premise says that moral judgments themselves move us to act. It is now standardly taken as an a priori claim, but for the arch-empiricist Hume, it is probably just an a posteriori observation: people's moral views, on their own, tend to make a difference to how they act and react.

For the crucial Inertia of Reason premise, Hume offers two supporting arguments. Reason, for Hume, is the faculty of *reasoning* or inferring. The *Anti-Representationalist Argument* says that passions don't represent how things are, so they can't be true or stand in logical relations, which means reasoning gets no grip on them. The *Elimination Argument* says that reasoning consists in demonstrating necessary relations between ideas, as in logic or mathematics, or in inferring probable causal relations on the basis of experience. But neither discovering logical truths nor discovering that an action would probably produce a certain outcome will *by itself* move us to act, unless we want the end to be produced. Nor will instrumental reasoning necessarily move us. If we do wish to bring something about, we will *probably* acquire a desire to take what causal reasoning reveals as a necessary means, but if we don't, that is not a failure of reason. Many commentators consequently argue that Hume doesn't really have a conception of *practical* reasoning at all (see PRACTICAL REASONING). The Elimination Argument has also been accused of straightforward question-begging: after all, Kantian rationalists say that there is a third type of reasoning, namely (non-instrumental) practical reasoning, that *does* motivate (Millgram 1995). However, the burden is on the rationalist to come up with an account that displays why their favoured

kind of motivating process deserves to be called *reasoning*, and not something else, and that it delivers determinate conclusions about ends.

How should the *conclusion* of the Practicality Argument be understood? Some have taken it to establish that moral *judgments* are not *beliefs* (see the next section). But Hume always talks about the inertia of reason, not of belief. As long as there are beliefs that result from something other than reasoning, they might be motivating, and he explicitly says that hedonic beliefs are. But this relation between beliefs and motivation is causal, not rational, so reasoning alone does not motivate. However, this still leaves the conclusion puzzling. Hume has argued that moral judgment (a psychological state) motivates, while reasoning (a psychological process) doesn't. It doesn't follow that reasoning alone couldn't produce a moral judgment, which then (non-rationally) motivates. Hume needs a further, undefended assumption to the effect that non-motivating processes cannot produce motivating states. Perhaps, as Rachel Cohon (2008) has recently argued, he only aims to establish that moral discrimination and reasoning are distinct processes, the former being sentiment-driven and inherently motivating.

Contemporary explanatory sentimentalists often appeal to considerations related to the Practicality Argument. But they have added some new anti-rationalist arguments, often based on experimental research. The *Argument from Emotional Deficits* begins with the observation that if reason is sufficient for moral judgment, absence of normal emotions won't lead to deficits in judgment. The crucial case is psychopaths, who seem to be just as rational as the rest of us, but lack normal emotions. Do they make moral judgments? Shaun Nichols (2004) frames the issue in terms of the distinction between moral and conventional rules. Even small children distinguish between violations of rules that prohibit hurting others and rules that prohibit dressing in a certain way. Paradigmatic moral violations are considered to be more serious and wrong regardless of the say-so of de facto authorities, time, and place. Drawing

on experimental research, Nichols argues that people come to treat some norms as non-conventional, because the actions they prohibit give rise to empathic distress – we feel bad when others are hurt, but not when someone wears a hat indoors. Psychopaths do not empathize, so Nichols’s account predicts that they treat all norms as conventional. That does indeed seem to be the case – psychopaths consider hitting someone is wrong only in the sense in which wearing jeans to a job interview is wrong. This suggests rationality is not sufficient for moral competence.

Empirical moral psychologists used to focus on people’s verbal justifications for their judgment, concluding that they manifest increasing rationality. The *Argument from Rationalization*, however, maintains that people’s judgments are generally not explained by the reasons that they give, but by quick and automatic affective reactions. In a well-known experiment, Jonathan Haidt and colleagues gave subjects a scenario involving consensual, harmless sibling incest. Most people judged the behaviour of the siblings to be wrong, citing various harms that by stipulation weren’t involved. When this was pointed out, people were dumbfounded, but nevertheless held to their judgment. Haidt’s (2001) explanation is that what is really responsible for the judgment is a negative affect such as disgust at the thought of incest, but when asked, people *confabulate*, coming up with a likely story.

This picture is derived from a more general model of the mind as divided into two distinct systems. The *intuitive system* is fast, effortless, automatic and unintentional, and only its products but not processes are accessible to consciousness. Many of its elements are task-specific evolutionary adaptations. The *reasoning system*, by contrast, is slow, requires effort and attention, and involves at least some consciously accessible and controllable steps and verbalization. For Haidt, reason for the most part enters the scene only under social pressure, after the judgment has already been made. It should be noted, however, that the only explanation that the dumbfounding evidence rules out is conscious reasoning – unconscious

reasoning, intellectual intuition, or automatic computation by a moral faculty (see Mikhail 2011) are all compatible with it.

1.2 Sentimental Explanation

In his pioneering work, Lord Shaftesbury (see SHAFTESBURY, THIRD EARL OF) argued that when we become aware of affections behind actions, there arises “another kind of Affection towards those very Affections themselves, which have been already felt, and are now become the Subject of a new Liking or Dislike” (Shaftesbury 1699/2001: 16). Positive second-order affection, aroused by motives tending towards social harmony, constitutes the sentiment of moral approval, which determines moral judgment, unless gradual habituation into a false religion interferes with this natural process. Hutcheson argued that these second-order affects are best thought of as deliverances of a *sense* because they are fundamentally non-voluntary ideas: we can no more choose to approve of a motive than we can choose to make honey taste bitter. This “Determination of our Minds to receive amiable or disagreeable Ideas of Actions” (Hutcheson 1725/2004: 100) regardless of consideration of our own advantage is an innate and primitive faculty that cannot be explained in terms of simpler abilities, but only by appeal to a benevolent creator.

By contrast, Hume does not appeal to a moral sense as an explanatory primitive, but rather aims to lay bare its mechanisms in a naturalistic fashion. The key mechanism is *sympathy*, which is not a feeling but rather the capacity and tendency to take on the feelings of others (now called empathy) (see EMPATHY). For Hume, sympathy is typically involuntary, arising out of association between the behaviour of others and our own feelings when we act alike. The source of approval or disapproval of motives is sympathy with the pain and pleasure of those affected by the action, in particular those in the “narrow circle” of the evaluated person (such as friends and colleagues).

A challenge for this account is that natural sympathy varies where judgment doesn't: even if the motive is the same, it is harder to sympathize with strangers and unsuccessful attempts. Hume responds to this by noting that judgments are based on 'corrected' sympathetic responses. In order to prevent "continual contradictions" with each other and ourselves over time, we must "fix on some steady and general points of view" (Hume 1740/1978: 581) in our imagination, and guide our judgments by feelings from such "common point of view" (589) that is "without reference to our particular interest" (472). We must also consider the general tendency of a trait to produce sympathetic pleasure rather than success in a particular case. The pleasures and pains resulting from corrected sympathy have a distinctive nature in virtue of which they give rise to the moral sentiments of love or hatred toward the agent. These in turn motivate reward and punishment.

For Smith, in turn, sympathetic sentiments result from placing ourselves in the shoes of others in imagination. We take the psychological states of others to be appropriate when we can imaginatively 'enter into' them, that is, find that we would respond the same way in their situation. This perceived accord gives pleasure, which constitutes approval. When it comes to *moral* merit of an agent (deserving blame or praise), our judgment is driven by imaginatively sharing the (sometimes hypothetical or 'illusive') *resentment* or *gratitude* of those directly affected, which in turn involves failing to sympathize with the agent.

Smith extends the same model to judgments concerning our *own* motives and actions. We originally come to approve or disapprove of them by sympathizing with the reactions of actual others to our actions. But experience teaches that actual others are often ignorant of our true motives or biased in their reactions, and that 'natural misrepresentations of self-love' bias our own. Thus as mature moral agents, we judge ourselves on the basis of the responses of an *imaginary* impartial spectator, an 'internal representative of mankind' with full knowledge of

our motives, and are not satisfied with ourselves if we perform actions we would despise other people for performing.

Most contemporary sentimentalists follow Hume and reject *moral innatism*, the view that evolution has provided us with an innate moral faculty or principles, and argue that morality can be explained in terms of non-dedicated psychological capacities, such as rule-following, empathy, and emotion (Nichols 2005). Again, new arguments often draw on empirical research. The *Argument from the Emotional Brain* is based on neuroimaging studies, which suggest that areas of the brain associated with emotion are active when making moral judgments (Moll et al. 2002). Joshua Greene claims that emotions selectively explain deontological judgments, such as the judgment that it is wrong to push down one person in order to save five (Greene 2008). People who make utilitarian judgments (it is okay to sacrifice one to save five), in contrast, show activity in areas associated with conscious reflection, such as the prefrontal cortex. Further, Greene claims that if people's deliberative capacities are taxed during the experiment, utilitarian judgments take more time, but deontological judgments remain fast, suggesting that deliberation plays a role only in the former. This Dual Process Theory is thus only half sentimentalist. One important problem with such neuroscientific arguments is that our current knowledge of the brain is very incomplete. Colin Klein (2011) contends that present data in fact provide no evidence that emotions are selectively involved in deontological judgment, given the multiple roles that brain areas can play. Further, the data leaves open whether emotions cause judgments or the other way around (in which case explanatory sentimentalism is false), or whether emotions *constitute* judgments (which would be a form of judgment sentimentalism).

The *Argument from Emotional Manipulation* draws on many experiments showing that changing people's emotions changes their judgments. For example, Wheatley and Haidt (2005) hypnotized subjects to feel disgust at the mention of random words, such as 'often'.

Subsequently, these subjects rated entirely innocent actions as morally suspicious, if the charged words were used in describing them, suggesting that the irrational affective response was driving their judgment. However, rationalists are unlikely to be impressed by the fact that emotions can influence moral judgment – after all, they’ve traditionally seen emotions as a *distorting* influence. Mere causal influence doesn’t show that the automatic emotional reactions Haidt focuses on are in any sense *fundamental* to moral thought. One reason to think they aren’t is that while quick flashes of affect may account for the making of particular moral judgments, they are badly suited to explain why moral judgment in general has the *special features* it does, such as felt subjective and intersubjective authority and independence from the commands of social superiors. Indeed, Haidt (with Björklund, 2008) talks revealingly about liking and disliking in the same breath as moral judgments, as if there were no difference to explain.

2. Judgment Sentimentalism

For judgment sentimentalists, moral judgments are not caused but rather *constituted* by emotions or beliefs concerning emotions. The crucial question is whether judgments purport to represent how things are (see MORAL JUDGMENT). Hume’s remarks on this point are ambiguous and possibly contradictory. Some of the things he says support the non-cognitivist interpretation, on which to think that someone is virtuous just *is* to feel a sentiment of approbation towards her (see NON-COGNITIVISM). This is not a matter of *describing* how one feels, but of *expressing*, making public the sentiment one has, and perhaps inviting others to share it. On the whole, however, Hume is probably better interpreted as a cognitivist for whom moral judgments are beliefs about how traits and motives do or are disposed to make us feel when contemplated from the common point of view (see COGNITIVISM). As Smith puts it, “The very words, right, wrong, fit, improper, graceful, unbecoming, mean only what

pleases or displeases [moral] faculties.” (Smith 1759/2002: 192) Both kinds of view have been further developed by contemporary sentimentalists, often in the context of discussing the semantics of sentences expressing the judgments.

To start with non-cognitivism, the *Argument from Judgment Internalism* derives from the Practicality Argument, but where the latter concludes that the process of moral *judging* isn't one of *reasoning*, the former concludes that moral *judgment* isn't a *belief*:

1. Moral judgments are intrinsically motivating. (Moral Judgment Internalism)
2. Only non-cognitive psychological states with a world-to-mind direction of fit are intrinsically motivating; beliefs, which have a mind-to-world direction of fit, do not alone motivate (see MOTIVATION, HUMEAN THEORY OF).
3. So, moral judgments are (at least in part) constituted by non-cognitive psychological states.

Contemporary non-cognitivists differ on just what kind of states are involved in moral judgment. Emotivists like Ayer (1936) believed moral thoughts consisted in a *sui generis* moral feeling. For Blackburn (1998), judgments consist in higher-order attitudes of approval and disapproval. For Gibbard (2003), they are contingency plans to feel guilt and anger in certain situations. The basic challenge for these *expressivist* theories is to explain the *descriptive surface* of moral thought and language: If moral judgments don't represent how things are, how can they be (objectively) true? What accounts for the logical validity of moral arguments? Expressivists have developed sophisticated theories to account for these features in terms of the non-representational states they take judgments to express, but it is fair to say that no solution commands wide assent.

Jesse Prinz's (2007) recent view fuses non-cognitivist and cognitivist elements. The key to this is his theory of emotions, according to which they are perceptions of patterns of

bodily changes that represent what they are “set up to be set off by” – what it is their biological or culturally elaborated *function* to detect. For Prinz, a moral *sentiment* or (internalized) rule is a disposition to respond to certain actions with a range of self- and other-directed blame- or praise-constituting emotions, such as guilt, contempt, anger, and gratitude. A moral *judgment* consists in the emotion that results from activating a sentiment, such as anger at stealing or shame for fleeing. In context, the anger represents stealing as being such as to cause disapprobation in the judge – that is, as morally wrong in the relativist sense Prinz embraces. This means the judgment can be true or false, while simultaneously motivating to punish the agent. For Prinz, this is the most parsimonious explanation of the many strands of evidence linking emotion to moral judgment.

Cognitivist sentimentalists, by contrast, believe that moral judgments are not sentiments, but beliefs about what merits moral sentiments, or would cause them in suitable circumstances (see below). It is not entirely implausible that beliefs with such content have an intimate link to motivation, either by themselves or by way of a necessary connection to desires (see MORAL MOTIVATION). But why would belief that some improved version of me, or a disinterested spectator, *would* feel approbation toward an action actually move me, *as I am*, to act that way? If I do actually inhabit the common point of view and experience the sentiment, I will be motivated, but what if I don't? Mackie (1980) suggests that only a kind of secondary motivation to measure up to an ideal, buttressed by sensitivity to the actual reactions of others, could account for the motivational effect of such beliefs. But this means giving up on the internalist idea that judgment itself can move us.

More radically, *sensibility theorists* blur the very distinction between beliefs and desires, rejecting the Humean Theory of Motivation. Inspired by Aristotle, John McDowell argues that for a virtuous person to see an action as called for, or to perceive a reason for action, is to be in a state that is both cognitive and motivational (McDowell 1979). For a

generous person, someone's need appears as calling for assistance, which itself moves her to act. A selfish person who lacks the motivation will also necessarily perceive (or conceive) the need differently (see PERCEPTION, MORAL). A consequence of this is that the process of coming to appreciate practical reasons will involve shaping the agent's motivational sensitivities and may itself be non-rational (McDowell 1995).

Justin D'Arms and Daniel Jacobson (2000) argue that what unifies contemporary non-cognitivist and cognitivist sentimentalists is that for both, judging that something is wrong or admirable (say) is not just a matter of having an emotional reaction, but of thinking that a response of blame or admiration is *appropriate* in some circumstances. For non-cognitivists like Gibbard (1990), thinking that a response such as anger or guilt is appropriate (or rational) is a matter of accepting a norm that permits or requires it, while for cognitivists it is some kind of belief about appropriateness. D'Arms and Jacobson argue that both kinds of account face the *Conflation Problem*: an attitude may be appropriate for strategic or other intuitively irrelevant reasons even if its object lacks the property the judgment predicates. For example, it may be appropriate for anyone to admire a vain rich person (at least one with mind-reading capacities) because of extrinsic rewards, but that doesn't mean she is *admirable* (see WRONG KIND OF REASONS PROBLEM). This is important, because unless sentimentalists find a way to distinguish between warranted and unwarranted feelings without conflation, they won't have earned the right to talk of truth and objectivity. D'Arms (2005) suggests that the answer can be found in how the specific emotions present their objects – for example, admiration presents its target as having accomplished something significant, so only considerations relevant to significant accomplishment can make admiration a *fitting* response. Insofar as responses won't be fitting for the wrong kinds of reasons, the Conflation Problem is avoided.

3. Metaphysical Sentimentalism

While explanatory and judgment sentimentalism are theses in moral psychology, metaphysical sentimentalism is a view about the nature of moral facts or properties. Metaphysical sentimentalists reject the claim that moral facts or properties are mind-independent features of the world. Against non-naturalist mind-independent moral realism, metaphysical sentimentalists deploy the same *Argument from Queerness* as error theorists do: were moral properties to be *sui generis* mind-independent properties, they would be unlike everything else we have reason to believe in, and it would be a mystery how we can come to know them (see ERROR THEORY). Non-cognitivist judgment sentimentalists further agree with error theorists that we tend to *project* values onto the world, treating our own reactions to things as if they were their mind-independent features (Blackburn 1984). In this spirit, Hume talks about taste “gilding or staining all natural objects with the colours, borrowed from internal sentiment” (Hume 1751/1975: 294). Projectivism can be seen as a negative metaphysical thesis, a form of ontological nihilism: what we call moral properties are not genuine properties of things any more than images projected on a wall are genuine properties of the wall.

Non-skeptical cognitivist judgment sentimentalists, however, argue that there is a third metaphysical option besides mind-independent realism and ontological nihilism: the properties that moral terms predicate of things could be among the *mind-dependent* features of the world, in the same broad ontological category as being funny, red, or disgusting (see RESPONSE-DEPENDENT THEORIES). Such properties would not exist in the absence of human beings and their subjective reactions, but are nevertheless on most views genuine qualities of objects, being independent of any particular subjective response. Consequently, we can have true or false beliefs concerning them in the most straightforward sense, and they may enter into explanations of other phenomena.

Mind-dependence views fall into two main groups. According to *dispositionalism*, moral properties are dispositions of actions or traits to *cause* the relevant responses in us, perhaps in certain circumstances, such as when we are well-informed and disinterested spectators (Westermarck 1906, Prinz 2007). Hume explicitly compares vice and virtue to “sounds, colours, heat, and cold” (Hume 1740/1978: 469), that is, Lockean secondary qualities – powers of things to cause subjective responses in us. (On Michael Slote’s variant, empathic reactions of warmth or chill to agents’ motives serve rather to *fix the reference* of moral terms, so that moral properties are whatever (response-independent) properties actually cause such empathic responses (Slote 2010).) There is no doubt that there are actions that have, say, the property of being such as to make most people angry. There is nothing queer or non-natural about such features. The question is whether they are *moral* properties, that is, the properties we attribute when we say that actions are wrong (say). One important challenge is that while an action’s being wrong seems to give anyone a categorical reason not to do it, it is not clear whether being such as to make most people angry (or any similar disposition) gives such a reason.

According to an alternative kind of metaphysical sentimentalism, the counterpart of sensibility theory about moral judgment, we should care about something being right or wrong, because moral properties are features that *merit* the relevant responses (McDowell 1985; Wiggins 1987). Just as not everything that amuses us is funny, not everything that elicits blame from us is blameworthy. The secondary quality analogy holds only so far. Another disanalogy is that we can only identify the relevant reactions, such as amusement or moral disapproval, by reference to the properties that give rise to them. This “no-priority view” is openly circular: the funny is that which merits amusement, and amusement is the response merited by the funny (McDowell 1987: 158). For sensibility theorists, such circularity is no bar to informative elucidation.

One challenge to these models comes from naturalist mind-independent moral realism. Why couldn't moral properties be the natural properties, such as maximizing happiness, that underlie the disposition to give rise to moral emotions? Sentimentalists respond with the *Argument from Shapelessness*: the only thing that morally right or wrong things have in common is that they arouse the relevant responses in us (McDowell 1979). Jackson, Pettit, and Smith (2000), however, claim that this would render moral predicates unlearnable, and insist that there must be a pattern on the level of natural properties.

The most important challenge to metaphysical sentimentalism is a version of the *Euthyphro Dilemma* (Wright 1992; Blackburn 1998) (see EUTHYPHRO PROBLEM). The standard biconditional stating the mind-dependence view has the following form:

Necessarily/a priori, X is right/wrong if and only if would cause/merit response R in subjects S in conditions C.

On the simplest dispositionalist reading, it is necessarily or a priori true that something is wrong if and only if it would elicit disapproval in statistically normal subjects in normal conditions (just as, perhaps, something is red if and only if it looks red to normal subjects in normal conditions). The first horn of the dilemma is that this makes it impossible for such subjects in such conditions to be wrong, while in fact this is always possible. Statistically normal people are surely not morally infallible! Further, we use evaluative concepts to criticize, persuade, and guide each other's responses, not merely to describe them, which makes no sense unless there is a gap between actual or likely responses and responses that are in some sense correct (D'Arms and Jacobson 2000: 727). To leave open the possibility of mistakes and capture the sense in which properties are normative for response, the dispositionalist has to impose some kind of idealization on the subjects or conditions – perhaps it is only fully informed, impartial, and benevolent subjects whose reactions

determine the extension of moral properties (Firth 1952) (see IDEAL OBSERVER THEORIES). This takes the view in the vicinity of the merit reading, as we can plausibly say that something merits a response if it elicits that response from well-informed morally good or virtuous subjects – those who are impartial and kind, among other things.

This, however, leads to the other horn of the dilemma: we seem to need a response-independent standard for determining whose responses count or who is a morally good subject, since otherwise the account will be circular and trivial (something is morally good if and only if approved by the morally good, and someone is morally good if and only if she approves what is morally good). Once again, sensibility theorists don't regard this as problematic. We cannot suspend all our moral beliefs at once in determining who is a good judge, but rather make use of those that withstand scrutiny by standards internal to moral practice, leaving open the possibility that any particular belief is mistaken. In consequence, sensibility theorists must (and do) give up on any project of reducing moral properties to non-moral properties.

Dispositionalists also face the challenge of *relativism*: it may be that the same action would cause different responses in different observers. Of contemporary sentimentalists, Prinz cheerfully embraces relativism, arguing that different actions are indeed wrong for different people and cultures. We can nevertheless compare different moral frameworks in terms of non-moral values, such as contribution to human happiness. Others respond that the right kind of idealization of subjects and circumstances will guarantee that metaphysically subjective properties nevertheless remain epistemically objective – suitable subjects will all approve of the same actions, regardless of individual starting points. The related *Argument from Moral Necessity* says that things like rape and murder would be wrong even if our attitude towards them would switch to approval, so our sentimental responses cannot metaphysically determine the moral status of actions. In response, David Wiggins (1987)

proposes that we fix our *actual* sensibilities as starting points for idealization. This renders hypothetical reversals of sentiment irrelevant.

Suppose that sentimentalists can give a satisfactory account of facts about right and wrong. This may still leave open the question of why we have an *obligation* to do the right thing. Hutcheson (1725/2004) distinguished between natural (self-interested) and moral obligation, arguing that the former derives from the necessity of virtue to happiness, and the latter from reflective disapprobation of the moral sense. Hume added that once we have laid bare the sympathetic mechanisms of the moral sense, its operation will be approved of by itself, giving it a kind of authority. For Kantian critics, however, this still leaves open the question of why the moral sense should *bind* our will. Here Smith may have the best sentimentalist response to offer. Perhaps being guided by an imagined impartial spectator, as opposed to given desires, is constitutive of autonomous agency itself – Smith talks about “self-command” – and its dictates thus binding in much the same way as those of Kantian practical reason.

4. Epistemic Sentimentalism

Hume famously argued that we cannot infer an ‘ought’ proposition from an ‘is’ proposition, because the former introduces “a new relation” of an entirely different kind (Hume 1740/1758: 469). This means we cannot come to know moral truths by reasoning from known non-moral truths. This argument follows metaphysical and explanatory arguments to the effect that moral truths aren’t the sort of relations that reasoning can discover. The consequence is that ethics is *epistemically autonomous*. In this, sentimentalists agree with intuitionists, but they also deny that moral propositions are self-evident, so we cannot come to know their truth merely by understanding them either. Instead, we come to know moral truths by way of *emotional experience*.

The basic positive sentimentalist argument, too, draws on sentimentalist explanation and metaphysics. If to be wrong is just to be such as to give rise to a negative emotional reaction in suitable conditions, we can reliably discover which things are wrong by placing ourselves in such conditions – which may require the use of theoretical reasoning to learn the pertinent facts – and letting our response guide our judgment (Lewis 1989). Kauppinen (forthcoming) argues that emotional responses felt from a common point of view can plausibly be *identified* with moral intuitions: they are non-belief states that arise involuntarily (sometimes in spite of contrary beliefs), present their object as having a moral property, and incline us to form the corresponding belief (see INTUITIONS, MORAL). They thus function much like intuitions in other areas, providing non-inferential epistemic justification for belief. This is consistent with other emotional states, such as the ‘flashes of affect’ that Haidt (2001) talks about having a *merely causal*, justification-undermining role.

Greene (2008) argues that neuroscientific results, which suggest that deontological intuitions derive from aversive affective responses to up close and personal violence, selectively undermine the epistemic justification of deontological theories. Greene argues that since the affective responses in question are evolutionary adaptations to life in small, tightly knit groups, they are sensitive to morally irrelevant features, such as physical distance. Occasionally Greene speaks as if the very fact (if it is such) that deontological intuitions are based on emotion rather than reasoning renders them unreliable. However, Berker (2009) argues that neither neuroscientific data, evolutionary history, nor emotional ancestry *as such* play any role in Greene’s main argument, which simply relies on the substantive moral assumption that the process leading to deontological intuitions is influenced by morally irrelevant features. That is something that deontologists are likely to dispute.

Conclusion

We have reviewed a variety of forms of sentimentalism from their origins to present. From the survey, it is easy to see the continuity between classical and contemporary variants, both with respect to the positions defended and the arguments used. Recent empirical findings have again fostered interest in the role of emotion in ethics and roused controversy that is not likely to be resolved soon.

Cross-References

COGNITIVISM; ERROR THEORY; EUTHYPHRO PROBLEM; HUME, DAVID; HUTCHESON, FRANCES; IDEAL OBSERVER THEORIES; INTUITIONISM, MORAL; INTUITIONS, MORAL; MORAL JUDGMENT; MORAL MOTIVATION; MOTIVATION, HUMEAN THEORY OF; MOTIVATIONAL INTERNALISM; NON-COGNITIVISM; PERCEPTION, MORAL; PRACTICAL REASONING; RESPONSE-DEPENDENT THEORIES; SHAFTESBURY, THIRD EARL OF; SMITH, ADAM; WRONG KIND OF REASONS PROBLEM

Sources

Ayer, Alfred Jules 1936. *Language, Truth, and Logic*. London: Gollancz.

Blackburn, Simon 1984. *Spreading the Word*. Oxford: Clarendon Press.

Blackburn, Simon 1998. *Ruling Passions: A Theory of Practical Reason*. Oxford: Clarendon Press.

Berker, Selim 2009. "The Normative Insignificance of Neuroscience," *Philosophy and Public Affairs* vol. 37 (4), pp. 293–329.

Cohon, Rachel 2008. *Hume's Morality: Feeling and Fabrication*. Oxford: Oxford University Press.

D'Arms, Justin 2005. "Two Arguments for Sentimentalism," *Philosophical Issues* vol. 15, pp. 1–21.

D'Arms, Justin and Jacobson, Daniel 2000. "Sentiment and Value," *Ethics* vol. 110 (4), pp. 722–48.

- Firth, Roderick 1952. "Ethical Absolutism and the Ideal Observer," *Philosophy and Phenomenological Research* vol. 12 (3), pp. 317–45.
- Gibbard, Allan 1990. *Wise Choices, Apt Feelings*. Oxford: Clarendon Press.
- Gibbard, Allan 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Greene, Joshua 2008. "The Secret Joke of Kant's Soul," in Walter Sinnott-Armstrong (ed.) *Moral Psychology. Vol. 3*. Cambridge, MA: MIT Press, pp. 35–80.
- Haidt, Jonathan 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* vol. 108, pp. 814–34.
- Haidt, Jonathan and Björklund, Frederik 2008. "Social Intuitionists Answer Six Questions About Moral Psychology," in Walter Sinnott-Armstrong (ed.) *Moral Psychology. Vol. 2*. Cambridge, MA: MIT Press, pp. 181–217.
- Hume, David 1740/1978. *A Treatise of Human Nature*. Ed. L. A. Selby-Bigge, 2nd, rev. ed. P. H. Nidditch. Oxford: Clarendon Press.
- Hume, David 1751/1975. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Ed. L. A. Selby-Bigge, revised by P.H. Nidditch. Oxford: Oxford University Press.
- Hutcheson, Francis 1725/2004. *An Inquiry Into the Original of Our Ideas of Beauty and Virtue in Two Treatises*. Ed. Wolfgang Leidhold. Indianapolis, IN: Liberty Fund.
- Hutcheson, Francis 1728/2002. *An Essay on the Nature and Conduct of Passions, With Illustrations on the Moral Sense*. Ed. Aaron Garrett. Indianapolis, IN: Liberty Fund.
- Jackson, Frank, Pettit, Philip, and Smith, Michael 2000. "Ethical Particularism and Patterns," in Brad Hooker and Margaret Olivia Little (eds.), *Moral Particularism*. Oxford: Oxford University Press, pp. 79–99.
- Kauppinen, Antti forthcoming. "A Humean Theory of Moral Intuition," *Canadian Journal of Philosophy*.
- Klein, Colin 2011. "The Dual Track Theory of Moral Decision-Making: a Critique of the Neuroimaging Evidence," in *Neuroethics* vol. 4, pp. 143–62.
- Lewis, David 1989. "Dispositional Theories of Value," *Proceedings of the Aristotelian Society Supplementary Volume* vol. 63, pp. 113–37.
- McDowell, John 1979, "Virtue and Reason," reprinted in McDowell 1998, pp. 50–73.
- McDowell, John 1985, "Values and Secondary Qualities," reprinted in McDowell 1998, pp. 131–50.
- McDowell, John 1987, "Projection and Truth in Ethics," reprinted in McDowell 1998, pp. 151–66.
- McDowell, John 1995, "Might There Be External Reasons?" Reprinted in McDowell 1998, pp. 95–111.
- McDowell, John 1998. *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.

- Mackie, John 1980. *Hume's Moral Theory*. London: Routledge.
- Mikhail, John 2011. *Elements of Moral Cognition*. Cambridge: Cambridge University Press.
- Millgram, Elijah 1996. "Was Hume a Humean?" *Hume Studies* vol. 21 (1), pp. 75–94.
- Moll, Jorge, de Oliveira-Souza, Ricardo, Bramati, Ivanei E. and Grafman, Jordan Grafman 2002. "Functional Networks in Emotional Moral and Nonmoral Social Judgments," in *Neuroimage* vol. 16, pp. 696–703.
- Nichols, Shaun 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.
- Nichols, Shaun 2005. "Innateness and Moral Psychology," in Peter Carruthers, Stephen Laurence, and Stephen Stich (eds.), *The Innate Mind: Structure and Content*. New York: Oxford University Press, pp. 353–430.
- Prinz, Jesse 2007. *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Shaftesbury, Third Earl of (Anthony Ashley Cooper) 1699/2001. *An Inquiry Into Virtue and Merit*. In Douglas Den Uyl (ed.), *Characteristicks of Men, Manners, Opinions, Times Vol.2*. Indianapolis, IN: Liberty Fund, pp. 1–100.
- Slote, Michael 2010. *Moral Sentimentalism*. Oxford: Oxford University Press.
- Smith, Adam 1759/2002. *The Theory of Moral Sentiments*. Ed. Knut Haakonssen. Cambridge: Cambridge University Press.
- Sturgeon, Nicholas 2001. "Moral Skepticism and Moral Naturalism in Hume's *Treatise*," *Hume Studies* vol. 27 (1), 3–84.
- Wheatley, Thalia and Haidt, Jonathan 2005. "Hypnotically Induced Disgust Makes Moral Judgments More Severe," *Psychological Science* vol. 16, pp. 780–4.
- Westermarck, Edward 1906. *The Origin and Development of Moral Ideas Vol. 1*. London: MacMillan & Co.
- Wiggins, David 1987. "A Sensible Subjectivism?" In *Needs, Values, and Truth*. Oxford: Blackwell, pp. 185–214.
- Wright, Crispin 1992. *Truth and Objectivity*. Cambridge, MA: Harvard University Press.

Further Reading

- Batson, C. Daniel 1991. *The Altruism Question: Toward a Social-Psychological Answer*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- R. James Blair 1995. "A Cognitive Developmental Approach to Morality: Investigating the Psychopath," *Cognition* vol. 57, pp. 1–29.
- Darwall, Stephen 1996. *British Moralists and the Internal 'Ought': 1640–1740*. Cambridge: Cambridge University Press.

De Waal, Frans 2006. *Primates and Philosophers: How Morality Evolved*. Princeton, N.J.: Princeton University Press.

Frazer, Michael L. 2010. *The Enlightenment of Sympathy: Justice and the Moral Sentiments in the Eighteenth Century and Today*. Oxford: Oxford University Press.

Gill, Michael 2006. *The British Moralists on Human Nature and the Birth of Secular Ethics*. Cambridge: Cambridge University Press.

Hoffman, Martin 2000. *Empathy and Moral Development: Implications for Caring and Justice*. New York: Cambridge University Press.

Helm, Bennett. 2001. *Emotional Reason*. Cambridge: Cambridge University Press.

Kauppinen, Antti 2010. "What Makes a Sentiment Moral?" *Oxford Studies in Metaethics* vol. 5, pp. 225–256.

Kennett, Jeannette 2002. "Autism, Empathy, and Moral Agency," *Philosophical Quarterly* vol. 52, pp. 340–57.

Korsgaard, Christine 1986. "Skepticism About Practical Reason," *Journal of Philosophy* vol. 83 (1), pp. 5–25.

Sayre-McCord, Geoffrey 1994. "On Why Hume's General Point of View Isn't Ideal – And Shouldn't Be," *Social Philosophy and Policy* vol. 11 (1), 202–28.

Schroeder, Mark 2010. *Non-Cognitivism in Ethics*. London: Routledge.

Schopenhauer, Arthur 1841/2009. *Prize Essay on the Basis of Morals*. In Christopher Janaway (tr. and ed.), *The Two Fundamental Problems of Ethics*. Cambridge: Cambridge University Press, pp. 113–257.

Sinnott-Armstrong, Walter (ed.) 2008. *Moral Psychology. Vols. 1–3*. Cambridge, MA: MIT Press.

Zimmerman, Aaron 2010. *Moral Epistemology*. London: Routledge.