

LEARNING EMOTIONS AND ETHICS¹

Patricia Greenspan

University of Maryland, College Park

Abstract

Innate emotional bases of ethics have been proposed by authors in evolutionary psychology, following Darwin and his sources in eighteenth-century moral philosophy. Philosophers often tend to view such theories as irrelevant to, or even as tending to undermine, the project of moral philosophy. But the importance of emotions to early moral learning gives them a role to play in determining the content of morality. I argue, first, that research on neural circuits indicates that the basic elements or components of emotions need not be limited to what psychologists think of as basic *emotions*. But in that case, innate mechanisms of social transfer of emotion, such as infants' tendency to facial imitation, gaze-following, and emotional contagion or empathy, provide a source of plasticity in developing the basic elements that lets emotions incorporate cultural influence from early on. This leaves room later for cognitive components of adult human emotions and hence for the further role of language in conveying cultural influence. We can thus see how moral judgment might depend on innate emotional capacities that are both modifiable by culture and capable of registering objective values. I use Rawls's treatment of the development of moral sentiments to illustrate the kind of supportive role emotions can play in a principle-based approach – though my own approach involves modifications I go on to indicate.

Evolutionary psychologists and others thinking in scientific terms about the sources of morality sometimes hark back to eighteenth-century British moral philosophy, which works from a conception of innate human nature. Darwin himself (see 1981, ch. 3) followed eighteenth-century models such as Hume 1978 and Adam Smith 1982 in basing morality on the social emotions. Recent arguments for an innate basis of ethics often take for granted some sort of explanation in terms of cross-cultural commonalities in emotion (see, e.g., Wilson 1993, Pinker 1997).

However, contemporary moral philosophers would question whether attempts to capture the origins of our moral responses can reveal appropriate bases for moral judgment. Explanation is one thing, justification another, and a central point of philosophic inquiry is to get beyond our emotional reactions to results of rational reflection. Even Hume held that genuine moral sentiments emerge only after we eliminate the distortions of natural sympathy – to the point where moral judgments on his emotion-based account need not involve feeling.

Some of the contemporary literature specifically on emotions feeds into a tendency to see them as irrelevant to moral judgment by linking the claim of innateness to a view of emotions as feelings or physiological responses lacking any complex cognitive content. I think it should be possible, though, to find a reasonable spot in between the polarities in recent debate that gives the full range of emotional sources of ethics its due and thereby allows a role for emotions in developed moral judgment. The view I favor recognizes innate emotional sources of ethics but lets emotions incorporate cognitive elements of a sort susceptible to cultural influence but compatible with objectivist ethics.

At one point not too long ago work on emotions in psychology and the social sciences was divided into opposing camps on the issue of innate emotions. Theories of basic emotion in psychology (see, e.g., Ekman 1971, Izard 1977) were opposed by a “social constructivist” position favored by anthropologists (cf. Averill 1980, Lutz 1986), which in extreme form holds that all emotion types are variable inventions of culture. But there is room for a less extreme version of both this and the “nativist” approach, which need only hold that some original subset of emotions

precedes social influence.

In philosophy a version of the social constructivist position appears in Armon-Jones 1991. By contrast, Griffiths 1997 defends basic emotions as evolutionary "affect-programs," clusters of responses to selected classes of stimuli – physiological and behavioral as well as affective responses – that might be said to be programmed into us by evolution. They are found in all cultures, though without the sort of fixity or unchangeability by either culture or further development that the term "innate" is often taken to imply.

Griffiths acknowledges responses resembling emotions that are set up entirely by culture, but he thinks of these as "social pretenses" of emotion that should be dealt with in a separate category (see esp. pp. 140ff.; but cf. Greenspan 2004). He also raises questions about whether the full range of states that seem offhand to be genuine emotions can be explained in terms of the basic set – whether we can "get there from here," as it were, with "there" taken as covering what he calls "the higher cognitive emotions." Though not himself concerned with moral emotions, Griffiths's discussion of attempts to "get there from here" by appeal to evolution deals with authors such as economist Frank (1988) who do assign a central role to moral (and other social) emotions.

This paper explores how we might at least move *toward* moral emotions, and thence to moral judgment, from a fairly minimal innate basis in emotion – along with whatever learning mechanisms and other conceptual equipment we also possess innately. Despite its ambiguities, which lead Griffiths to reject the term, I use "innate" to mean roughly what Griffiths has in mind when he speaks of evolutionary "programs": unlearned responses or response tendencies, emerging (sometimes well after birth) as a consequence of genetic endowment – but also subject to serious cultural modification, in ways that are particularly significant in the moral sphere.

Although I think it is plausible to suppose that there is an innate basis of morality, understood in such terms, this should not be equated, as it often is, with the eighteenth-century notion of a universal "moral sense." Among other things, there is more learning in this area than meets the eye – particularly an inexplicit (and initially undifferentiated) kind of moral learning that rests on

educating the emotions. I later provide some examples of how an innate emotional basis of ethics might be modified in essential ways by adult interaction with pre-linguistic children.

At various points in my argument I also indicate how an emotional basis may be further altered and expanded through language-encoded cognitive elements of emotion at more advanced stages. I use some neuroscientific evidence to suggest that we can break down basic emotions into even more basic components *of* emotion that can combine with cognitive components to generate the full set of developed human emotions. On the sort of account I favor, emotions can thereby incorporate an element of evaluative thought that is capable of taking on sophisticated content.

My discussion will draw attention to a factor that often seems to be overlooked in the literature on innate emotions: besides innate emotion *types*, we also have some general *mechanisms* whereby infants and children pick up reactions from others. These include the tendency of an infant to follow a caretaker's gaze (to look where the caretaker looks) and to imitate facial expressions or other behavior, including behavior that evokes emotions. An everyday illustration will suffice to show how these learning mechanisms allow for the social transfer and modification of emotions. So even if the mechanisms are themselves innate, as I assume in what follows, they serve to provide support for emotional learning.

The initial aim of my discussion here is to bring together diverse lines of inquiry into emotions and the origins of ethics in support of a moderate view on innateness. Section 1 attempts a rough overview of the earlier literature in psychology on basic emotions, along with some more recent neuroscientific results indicating how basic emotions might be broken down further, with a source of cultural plasticity supplied by innate learning mechanisms. I also indicate how such materials might give rise to specifically moral emotions, with room for increasing cognitive complexity and a further source of cultural variability, as emotions come to incorporate linguistic influence. Then, in Section 2, I focus more narrowly on my own area, moral philosophy, to defend moral learning in conjunction with emotion as an influence on moral judgment that need not

undermine objectivity. In particular, I use the treatment of moral development in Rawls 1998 to supply the structure for an alternative to the Humean sentimentalist picture of the derivation of ethics from emotion.

1. *Constructing Moral Emotions*

The evidence for basic emotions in psychology concerns the ability to identify emotion types cross-culturally, in the first instance by facial expression (following Darwin 1965). Exactly which emotions are included in the basic set is a matter of dispute, but familiar states like fear and anger are on everyone's list and also are commonly attributed to animals, so that the view feeds into an evolutionary perspective. On the other hand, disgust is also on everyone's list, though disgust (as distinct from taste aversion) is thought to be distinctively human (see Miller 1997, esp. p. 12; cf. Rozin and Fallon 1987). The application to animals also involves appeal to other bodily signs of emotion besides facial expression, along with confirming evidence from physiology and neuroscience.

Psychologists' work on basic emotions thus suggests an evolutionary approach to emotions, but it does not connect in a clear way to evolutionary accounts of morality. Positive or altruistic social emotions such as love and gratitude do not appear on the standard lists of basic emotions. Agreed items on the list include anger, disgust, fear, happiness, sadness, and surprise. Contempt was added to Ekman's initial list in Ekman and Friesen 1986 but remains controversial. There is also disagreement about whether to include shame and guilt, which require broadening the criterion for inclusion to features of bodily posture rather than simply facial expression (cf. Izard, 1977, esp. pp. 83-92). This area is still in flux, then, but there is evidence from other fields for attributing some social emotions, or at any rate evolutionary predecessors of social emotions, to animals.

Recent work on brain circuits by neuroscientist Joseph LeDoux (1996) indicates that even standard entries on the list of basic emotions such as fear fall into two different "tracks" of neural response, some involving the cerebral cortex and some – a behavioral "freezing" response to a loud sound, say (LeDoux's measure of fear in rats) – only the amygdala. But the second, more primitive,

track might also be taken to cover elements of the social emotions. For instance, physiological psychologist Jaak Panksepp (1998, ch. 13) discusses maternal love in rats in terms of the secretion of oxytocin – not itself an emotion, but apparently an element of love and related emotions such as trust (cf. Grimes 2003).

This suggests a way of breaking down basic emotions so that cortical involvement might allow for more fundamental modification of the basic set than we get on a model of derivation of emotions that is essentially inherited from seventeenth-century philosophy. Descartes (1970, pp. 331-99), most notably, understands “the passions of the soul” as derived from a list of “primary” emotions – analogous to primary colors in that more complex cases can be explained as blends of several primaries, which are not themselves subject to a similar explanation. But the more advanced forms of emotion need not be seen as combinations of states that are already fully constituted as emotions. Instead, we might expect the more primitive neural elements of emotion to take on more of the features of full-fledged human emotion when the link is made to the cerebral cortex.

Now, even relatively primitive emotions can be said to possess a kind of evaluative content: to the extent that affective states are motivationally rewarding or aversive, they “say” something positive or negative about the objects they are directed toward, as worth seeking or avoiding (cf. Greenspan 2003). Fear, for instance, says that its object is to be avoided as a threat. But once this element of content, whether itself learned or innate, comes to be cortically mediated, it is subject to modification on the basis of cultural influence – and, for that matter, individual reflection. Sometimes it may be modified in ways that influence how a given emotion feels. Consider some of the variants of love: possessive, devoted, or dependent, wild or subdued, depending in part on how we are taught to conceptualize the love object.

For some purposes, of course, one might want to exclude the more primitive, pre-intentional states – “objectless” anxiety, for instance, as a state that persists into adult human life – from the category of genuine emotions. In my own work the aim of such line-drawing has not been to establish the uniqueness of human emotional response (cf. Sousa 1987, ch. 4), but instead to bring

out the role of emotions in the justification of action (Greenspan 1988, p. 50), a distinctively human concern. Which states we should treat as paradigm cases of emotion, the more primitive or the more developed and cognitively complex, will vary with our theoretical purposes, normative or scientific.

Might there be a theoretically principled reason for drawing the line differently for different types of emotion? Some would be inclined to deny genuine love to nonhuman animals while readily granting them fear. Apart from the moral or religious significance assigned to love, one possible reason for love's exclusion from the list of basic emotions has to do with variable cultural influences on expression. If expressions of love or other of the more developed social emotions in humans are more subject to cultural influence, then they will not form the same relatively stable clusters of response that we get with fear, anger, and similar entries on the standard list.

Unlike psychologists work on basic emotions, the treatment of emotions in terms of brain circuits would not put any special emphasis on facial expression; nor would it even imply stable clusters of bodily response perceptible to the agent and others. However, according to common evolutionary arguments, emotion types that originally evolved as mechanisms promoting behavioral readiness – "fight-or-flight" in the case of fear – later came to be selected for their communicative function. The central role of facial expression in humans essentially serves to concentrate this secondary function of emotions into an area that facilitates social transfer. In the first instance, transfer of emotions from caretakers is accomplished by patterns of shared attention, or "gaze-following," along with the tendency of the human infant to imitate facial expression (see, e.g., the discussion in Bloom 2000, pp. 62-4; for relevant work more specifically on emotions see Witherington, Campos, and Hertenstein 2002 and the evidence cited in Hobson [this volume]; cf. also Sousa 1987, p. 183). These are social mechanisms below the level of culture but capable of conveying cultural influence, even if themselves innate.

The innate emotional basis of ethics thus need not be limited to a set of discrete emotion types, or basic emotions, but also would seem to include general mechanisms of emotional learning capable of modifying and expanding the basic set. Hoffman 1982 appeals particularly to the

mechanism of emotional empathy, originating in the tendency exhibited by newborns to cry at the sound of other newborns' cries, as a basis for guilt, a distinctively human emotion involving empathy with a victim of one's own harmful behavior. On this sort of account there is no implication of a culturally invariant "moral sense" as a feature of human nature.

The universal element of morality lies more importantly in its application *to* everyone, regardless of culture or some narrower set of personal affiliations. Universality in this sense – as in Hume 1978 on the correction of sympathetic emotions to eliminate distortions of personal standpoint – is a question of the content or scope of morality. This does seem to require cognitive advancement beyond animals, specifically the development of language, as needed to generate the full range of objects of moral response. But it also would seem to be facilitated by cross-culturally recognizable basic emotions, just insofar as they enable us to learn moral responses *from* anyone, regardless of culture. I think this is the real point of basic emotions in relation to ethics: they essentially allow for extension of a particular group's social contract.

At the same time, though, social influence introduces variations in the circumstances thought to justify a given emotion – what counts as an appropriate object of guilt or shame, say, in a given culture. For that matter, morality enlists competing emotional mechanisms – empathy or sympathy, but also the retributive tendencies encoded in resentment, indignation, and the like – that might be mixed or balanced differently in different cultures or even by different personality types. So there is ample room for divergent moralities or patterns of moral behavior or moral codes – for another point of contrast with the "moral sense" tradition.

*

What we seem to have, then, as innate bases of ethics are, first, a set of primitive states or elements of emotion, some but not all of which we are willing to count as full-fledged emotions, at any rate at earlier stages of development or in nonhuman animals, and (2) a set of mechanisms for emotional learning, or social transfer of emotions. What we do not have is either the seventeenth-century notion of emotional primaries whose combination yields the full range of human emotions

or the eighteenth-century idea of an innate "moral sense" grounded in human nature.

Note that my discussion so far concerns innate *emotional* bases of ethics, albeit not limited to innate moral emotions – emotions with a specifically moral content, such as shame or guilt. Though animals might be said to have rudimentary social emotions – or even the rudiments of moral emotions, as in philosopher Allan Gibbard's explanation of guilt in terms of the submissive posture in animals (see Gibbard 1990, ch. 7) – full-fledged moral emotions are usually placed solidly in the "higher cognitive" category. Hauser 2000, for instance, denies nonhuman animals genuinely moral emotions (and genuine empathy) on the grounds that they lack self-awareness (p. 224): awareness of themselves (and also of others) as bearers of the intentional states to which moral behavior in humans responds, such as beliefs, desires, and needs (see pp. 250ff.; but cf. Waal, e.g. 1989).

It is still possible, of course, that moral and other cognitively complex reactions are innate in humans – in a sense compatible with emerging well after infancy and requiring certain conditions to emerge, on the model of the capacity to learn a language. But this cannot be established just by appeal to psychologists' evidence for basic emotions. While brain studies supply further evidence (see, e.g., Rolls 1999), some current authors instead put forth arguments for innateness based on Chomskian "poverty of stimulus" arguments in linguistics, appealing to the paucity of explicit moral teaching to account for cross-cultural commonalities in moral judgment (see Dwyer 1999; cf. Hauser 2006).

However, we should note that moral teaching takes many forms apart from explicit instruction. We educate the moral emotions in children, for instance, by reacting to characters in the stories we tell them in a way that encourages shared emotional response. Chomskian nativists appeal to results on the order of psychologist Elliot Turiel's (1983) finding of a distinction made cross-culturally by preschool children between moral and merely conventional rule infractions (but cf. Haidt et al 1993 for anthropologists' criticism of Turiel's research). However, whether we express indignation or rather amusement at a certain character's exploits in a story can convey the idea that harm or a slight to another is of serious moral concern in a way that mere flouting of conventional

rules is not.

There is also an abundant supply of implicit moral instruction that only much later gets pulled apart from the other purposes of our use of emotional display to train children. Just saying "no" in an emphatic or warning tone counts as moral instruction in this not-yet-differentiated sense, which does not distinguish wrong from discourteous or dangerous action. Moreover, even if "poverty of stimulus" arguments based on Turiel's results favor innate origins of the *concept* of morality, this should not be confused with an innate basis of *morality*, in the sense of moral motivation and behavior – which can precede thought *about* morality, or reflective moral judgment, deploying the concept.

Some of the ways we teach moral motivation via emotion in advance of the development of moral thought – and also without reliance on specifically moral emotions – can be seen by looking at our play rituals with pre-linguistic children. A child thrusts one of its toys at us and we respond by elaborately portraying surprise and gratitude. The point is not just to teach the phrase "thank you" or conventions of politeness, but rather to convey and heighten joy at social exchange. We give the toy back, are handed it again with an expectant look, and repeat the emotional display. Our feelings, or pretended feelings, are contagious – not as gratitude per se, at this stage, but rather as a combination of simpler positive responses to giving and taking: excitement, delight, and amusement. The elements of affectionate interest and attention involved in care-giving and general play begin to get transferred by these means to other forms of social exchange and sharing.

This is a kind of early moral learning, though it has not yet been put to an explicitly moral use. When we include such interactions, though, the stimulus for moral learning can be seen to be quite rich, even if susceptible to further development by way of its pairing with language. At this stage we are not (or not merely) teaching language, but rather are using language as one vehicle of emotional expression in an effort to shape social emotions in children.

A more specifically moral example involves encouraging sympathy: a child kicks its mother and the mother exaggerates her expression of pain and hurt feelings, conveying the sadness and

resentment of a victim that becomes moral guilt. We also may discourage certain emotions, or in the first instance overt expression of them – by not making too prolonged a fuss, say, about a child's own minor hurts and pains. Clearly all of this is subject to the influence of cultural or other local norms, even if it presupposes some innate emotions, along with innate mechanisms of social transfer of emotions, possibly along with some innate conceptual categories or structures of moral thought, of the sort that Chomskian nativists have in mind. Ekman 1971 used the term "display rules" to explain Japanese subjects' denial of negative emotions in the presence of authority figures; the term was meant to represent such rules as extrinsic to the content of emotions, but on an account like the one I favor they can come to be incorporated into emotional content. Whose pain counts, how you should feel about causing it, and whether and when you may express or acknowledge those feelings, even to yourself, are questions that can be answered differently at an early enough stage of social learning to influence what people feel.

There also, at least arguably, are other concerns of morality besides harm or pain that can be emphasized and interpreted differently by different cultures. Disgust is one of the entries on the standard list of basic emotions, and it is put to use particularly by moralities that depend on a concept of moral sin (see, e.g., Shweder et al 1997). But it also is one of the emotions expressed in our own, primarily guilt-based, morality in response to behavior seen as "beneath" a scrupulous moral agent. While the emotion itself is found in all cultures, apparently there is not much cultural uniformity in its specific objects – in which food-sources evoke disgust, and which acts seem repugnant.

A different sort of example in support of cultural shaping of emotions is provided by feelings of respect or deference. These are essential to the workings of social hierarchy – and arguably, in a different way, to social equality – but it is not obvious how they could be derived from entries on the standard list of basic emotions. They seem to lack any tight connection to fear, for instance. Perhaps they might be understood as originating in reactions *to* certain basic emotions in others – submissively accepting the anger of caretakers when one disobeys them, say, as an instance of Gibbard's understanding of guilt (1990, ch. 7; cf. Blackburn 1998, ch. 1, sec. 3). But coming to

distinguish among sources of anger, assigning rightful authority only to some of them, involves augmenting the set of basic emotions with further materials provided by cognition and social interaction.

Some might be inclined to say that feelings like respect that do not constitute relatively stereotyped sets of responses like the items on the list of basic emotions do not really count as distinct emotions. But note, for instance, that gratitude is subject to similar questions, though it is an entry on standard lists of *nonbasic* emotions that plays a crucial role in evolutionary accounts of morality, as what motivates reciprocal altruism. Both of these emotions, respect and gratitude, might be thought of as special forms of love – one a response to general dependency, the other to a particular helpful act. We would be missing much of relevance to moral and practical life if we confined attention to “affect-programs,” or even to standard emotion categories, in discussing the role of emotions.

If we do not thus restrict the range of emotions, the role of language as the ultimate vehicle of cognitive content makes the number and variety of emotions potentially limitless. An affective element – not necessarily a whole affect-program – just needs to be brought into connection with an evaluative thought or cognition. This might be a complex thought *about* emotions, as suggested by Gibbard's account of moral judgment – accepting someone else's anger toward oneself as warranted, in the case of guilt – but in the simplest cases, cognition and affect just need to register a corresponding kind of significance of the situation for the organism, on the model of threat in the case of fear. Specifically moral emotions result from having the linguistic resources to capture social norms, plus some way (whether evolutionary or social) of acquiring an aversive reaction to a norm violation.

What rough picture emerges if we combine Gibbard's interpretation of moral judgments with Hoffman's account of the guilt response as based on emotional empathy? Though it begins just as sympathetic distress, guilt picks up its characteristic moral content when objects of childhood dependency identify with a victim of the child's harmful actions, conveying a sense of social censure

and rejection in response to a misdeed. A display of corresponding self-directed affect then becomes part of the child's submissive routine for assuaging such reactions on the part of others. Where an innate concept of morality might play a role is in determining how the reaction eventually generalizes – extending beyond caretakers' concerns to a certain class of concerns of people generally, ultimately giving rise to the sense of respect for persons as such.

What the child absorbs, though, is initially just a view of others' negative reactions as warranted in response to certain behavior on its part, specified in accordance with variable social norms. Concern for a younger sibling, say, might be something for which society exacts guilt for failure or instead treats as someone else's responsibility. Once language is learned, such norms can be spelled out explicitly and modified on the basis of explicit criticism.

This is not to deny that evolutionary responses have a kind of primacy (cf. D'Arms and Jacobson 2003); nor is it just to say that innate and cultural factors interact somehow in generating moral emotions. Rather, the suggestion is that the innate factors *include* the very mechanisms whereby emotions incorporate cultural influence of the sort requisite for moral emotions. The Chomskian distinction between principles and parameters – the former innate, the latter set differently by different cultures – might be useful as a way of summing this up, were it not for a tendency to take the relevant “principles” in this area as very general moral principles (cf., e.g., Hauser 2006, pp. 44ff.). Even where moral principles are shared cross-culturally, they need not be innate, but instead might be explained as responses to common features of life in a social group.

Incidentally, guilt was not a recognized emotion category until relatively recently: the noun as used in English to name an emotion was apparently the result of linguistic error in the period of the Protestant Reformation, with its stress on emotions of religious despair and self-accusation (see my account in Greenspan 1995, ch. 4). The possibility of reconstructing emotional guilt in terms of items on the list of basic emotions – variants of sadness or anger, conveyed by the mechanism of emotional empathy – may explain why the word seems to name something that was there all along, but it is something that not all societies put to a moral use (cf., e.g., Benedict 1946 and Doi 1973 on

Japan as a “shame culture”; cf. Williams 1993 on the Homeric Greeks).

Gibbard explains the distinction between guilt and shame by pairing them with different emotions on the part of others: guilt involves submitting to others' anger, whereas shame involves accepting their contempt. This explanation may make appeal to others' basic emotions (at least if contempt is allowed on the list), but it does not obviously derive guilt and shame from specific basic emotions in the individual's own repertoire – from her own anger or contempt, considered as a particular cluster of expressive and physiological responses.

There is no requirement, of course, that the higher cognitive emotions fall into categories that correlate one-one with sources in the basic set, even if they are derived from basic emotions. But moral emotions also seem to have a basis in caretakers' and others' behavioral responses to the child – incidents of social rejection in an overall situation of acceptance and trust – that may not evoke standard emotion types but rather just generalized discomfort or unease. This acquires a cognitive content when the child comes to view others' responses, or some of them, as appropriate or warranted. The resulting picture accords with the cognitivist or appraisal-based approach to emotions in twentieth-century philosophy and psychology (see esp. Bedford 1957 and Solomon 1976 in philosophy, and Arnold 1960 and Frijda 1986 in psychology), but as limited to fully developed human emotions. Emotions as such also involve noncognitive states of feeling or affect (cf., e.g., Robinson 1995), both as their initial source and as a motivationally important component, albeit often present only in modified or attenuated form.

2. Moral Objectivity and Emotion

An account that connects cognitive and affective components of emotion in the way I have suggested might seem to bear only on psychological questions about the origins of moral thought and behavior, not more strictly philosophic questions about what moral judgments or norms we ought to accept. It might even seem to undermine the aims of philosophic justification, if a basis in emotions influences the developed content of moral norms in a way that reflects variable cultural sources. In fact, Gibbard, the philosopher whose interpretation of guilt I have used as illustration,

is a noncognitivist on the issue of how to interpret a moral judgment: rather than amounting to a cognitive state of belief, a moral judgment on his account simply expresses acceptance of a norm for assessing moral emotions.

Gibbard's theory is in the tradition of both mid-twentieth-century emotivism and a common (though now disputed) reading of Hume in the eighteenth century as a skeptic about the foundations of ethics (cf. Mackie 1995, ch. IX). In fact, it is often assumed that the only possibilities for linking ethics and emotion are the eighteenth-century alternatives of Kantian rationalism and Humean sentimentalism: either disallowing emotion any role in grounding moral judgment or understanding moral judgment entirely in terms of emotion. This reason-versus-emotion dichotomy underlies recent work in the social sciences by Haidt et al 1993 (cf. Haidt 2001) on "dumbfounding," which takes the finding that people stick to their gut feelings on moral taboos such as incest as evidence against philosophers' emphasis on rational argument in ethics.

However, I think we can extract an intermediate view from Rawls 1998, which includes an account of emotions underlying the sense of justice (see ch. VIII) in support of principles derived from considerations of rational choice. Rawls essentially interprets emotional guilt as developing in stages into a full-blown moral emotion on the basis of increasing cognitive sophistication in an environment of love and trust that promotes self-esteem. Chomskian nativists often quote from Rawls's general theoretical discussion (see pp. 41-42) an analogy he briefly draws between the complex principles of justice his view ascribes to us and the principles of grammar we employ without awareness in making ordinary judgments of grammaticalness on a Chomskian account. However, Rawls's use of the language analogy has no direct bearing on questions of innateness – its point is just that we can act on principles we could not make explicit – though he later indicates affinity for views of the moral sentiments that assign us an innate psychological proclivity toward morality rather than treating it as an alien social imposition (cf. pp. 401ff.). In any case, very little attention is paid to Rawls's account of the moral sentiments. Hauser 2006, which treats a "Rawlsian creature" as a creature with "moral instincts" insofar as it relies on unconscious principles (p. 42),

works from a contrast between principles and feelings as embodying opposite sources of moral motivation. Let me try to correct things a bit by spending some time on Rawls's account of how we learn moral emotions.

In part, my aim will be to detach the general bearing of Rawls's account on the issues under discussion here from some of the particulars of his theory of justice. While moral emotions do not play a central role in his theory of justice, he does ascribe an important role to them – or to moral sentiments, conceived as the general dispositions they manifest – in moral motivation. His developmental account comes up in defense of the stability of the “well-ordered society,” a society arranged in accordance with principles of justice. Stability amounts to the tendency of a set of social arrangements to give rise to sentiments that support rather than undermining it, such as a desire to act in accordance with its principles and a tendency to feel guilt if one violates them. Rawls essentially argues that his two principles of justice yield enough stability to make a society founded on them viable – more so than the principles proposed by competing theories such as utilitarianism.

Rawls's account of the development of moral emotions, and ultimately of the sense of justice, thus proceeds from the assumption of just institutions. Besides understanding the well-ordered society as arranged in accordance with the principles of justice, it presupposes supportive families and peer and other cooperative groups that operate fairly. He takes this to distinguish his account from work in scientific psychology, which is supposed to be value-free. Since the elements of his developmental account are often described in terms of features of his overall theory of justice, it will take some work to detach just those elements of it that are relevant here; for a detailed scholarly treatment of Rawls's moral psychology and its role in the theory – a theory he ultimately saw as political rather than moral – see Baldwin 2008.

Guilt comes up in Rawls's account because infractions of principles of justice are bound to occur even in a well-ordered society. The emotion serves as a force tending to bring behavior back into line with the principles. A version of guilt is said to correspond to each of his three developmental stages, or in his terms “moralities”: the morality of authority, of association, and of

principles. At each stage a corresponding type of guilt arises from what amounts to a variant of love, assuming an appropriate element of reciprocity in feeling and behavior. Its initial basis is love and trust directed toward childhood caretakers whose behavior exhibits similar feelings toward us and thereby tends to instill in us a sense of our own self-worth, so that we essentially come to imitate their reactions to our misdeeds. Later, in the various groups we come to be involved in – from the family, through associations formed in play and education, to the larger society – guilt gets linked to fellow feeling toward others seen as well-disposed toward us and as doing their part in a cooperative arrangement. Finally, the sense of justice effectively adds to this a kind of gratitude toward the principles of justice themselves, as underlying social arrangements seen to benefit us and those we care for, though a given application of them may instead advance the interests of others outside our circle of personal affiliations.

Rawls speaks of commonsense morality, and of moral guilt and other moral feelings, in his account of the second stage, which turns on the development of various cognitive skills, especially the ability to take the perspectives of others and to appreciate their contributions to an overall cooperative enterprise. Though he suggests at one point that we do not have guilt “in the strict sense” (p. 415) until the third stage, when we see ourselves in relation to principles, I take him to mean that, understood strictly, guilt is based on an abstract judgment that one has committed a moral wrong, irrespective of any personal attachment to those who suffer it. He distinguishes guilt from shame in that guilt involves the idea of right, whereas shame rests on the broader notion of good (pp. 422-23), which extends beyond moral good.

With the interpretation of Rawls’s principles of justice as political rather than moral, which emerged more clearly in his later work (see esp. Rawls 1993), we also apparently need to recognize as part of the third stage of moral development a commitment to general *moral* principles, distinct from those governing basic social arrangements, but more abstract and systematically organized than the norms of commonsense morality. Here is where one sort or element of an innate concept of morality might be said to kick in – morality in contrast, not just to mere convention, but also to

merely personal considerations.

While Rawls's political principles allow for various different moral views as reasonable alternatives, his own views favor a Kantian ethic of respect for persons. However, what I want to take from him here is limited to two main structural features of the role he assigns emotions in defense of a normative view, whether moral or political. First, the role is evidently justificatory, even if secondary to the argument already presented for the two principles of justice. Insofar as his developmental account is needed support the stability of a society ordered in accordance with the principles, it provides essential support for them, without which they would not be worth instituting. He at one point denies the account justificatory status (see, e.g., p. 439), but I take that to mean that he understands his argument for the principles as complete without it, as required by its reliance on institutions determined to be just by some independent measure.

Second, and crucially, Rawls sees the role he assigns to emotion in support of normative principles as compatible with the assignment of objective status to them. The claim of objectivity comes out explicitly in his later writings on Kantian constructivism in metaethics (see esp. Rawls 1980), but as interpreted in a sense distinct from factual truth, with Rawls's principles of justice seen as limited to a certain cultural setting, namely democratic pluralist society. The contractual basis of Rawls's theory – in the rational choice of principles in an “original position” characterized by individuals' ignorance of their particularizing features – is understood as a reasonable procedure under these circumstances for constructing principles of justice. The principles therefore count as objective, without any claim of correspondence to an independent order of moral facts. It is their basis in a reasonable decision procedure that confers objectivity.

*

Let me now pull away sharply from the particulars of Rawls's theory, while making use of its basic structural features, in order to ask what it is that emotions might add to the backing for a moral code. When Rawls himself turns to arguing that his principles would promote greater stability than competitors, he does not extend his comments on guilt and other moral emotions but focuses

instead on self-esteem, as a product of the same developmental process. However, I want to suggest that the role of emotions in early moral teaching itself promotes something analogous to stability – understood, not as a feature of political systems set up in accordance with one or another philosophical theory, but instead as a feature of commonsense moral codes, diverging across different cultures.

In Greenspan 1995 (ch. 6) I treat this question under the heading of “viability”: whether a given moral code is up to the task of sustaining a society based on it. As with stability, all that is really needed for viability is to pass a certain threshold, not to outdo competitors, though in the contemporary situation of global communication and mobility, competition for members (or for their adherence to the local code) may push some otherwise viable codes below the threshold. Cross-cultural divergence in moral norms may well be compatible, then, with moral objectivity in Rawls’s sense – and also, I would add, with truth (on a less metaphysically demanding conception than Rawls sets up for contrast to objectivity in his sense). Many variations in codes will of course just be due to differing circumstances – including historical circumstances, such as acceptance of different customs or conventions – but, additionally, there may be more than one way of arranging cooperative social life that counts as reasonable in its distribution of benefits and burdens. So objectivity on this conception is compatible with a degree of cultural relativity – though we also would expect to find certain cross-cultural constants, given uniformities in both the cooperative aims of social life and the innate basis for moral learning provided by human emotions.

What my account of emotions adds to Rawls’s basic structure is an interpretation in terms of emotion of the motivational force that authors following Hume 1978 make out as essential to morality. In a nutshell: I take the viability of a moral code to depend on how well suited it is to being learned at an early stage on the basis of our innate stock of emotions and mechanisms of emotional transfer.

In the first instance, this is meant to allow for an intuitively satisfying answer to the question in contemporary moral philosophy of whether motivational force is internal to the very meaning of

a moral judgment – what is summed up as (motivational or judgment-) internalism. Early teaching in conjunction with emotions would invest moral judgments with motivational force, not necessarily in every case, but as a function of the meaning of moral language ((see Greenspan 1995, ch. 3; cf. also Greenspan 1998 and Blackburn 1998) – so that it would make sense, but at the same time be linguistically odd as well as psychologically abnormal, for someone (an “amoralist”) to claim to be completely unmotivated by a moral judgment she accepts.

Moreover, the account of how emotions supply morality’s motivational aspect makes them out as reinforcing the status of moral judgments as reasons for action in a way that supports the viability of a moral code. It is important to the account that moral language is taught to us in childhood more or less simultaneously with moral emotions and simple moral rules, as illustrated by the cases in Section 1. Moral reasons and moral reactions are thus developmentally intertwined. Guilt, for instance, becomes a distinctively moral emotion in Rawls’s “strict” sense once innate responses like empathetic sadness get refined to incorporate a notion of wrong that is conveyed to a child initially just by inducing those simpler emotions in reaction to various rule violations. So even emotions with a distinctive moral content do not simply incorporate pre-existing moral judgments. Philosophers may tend to reconstruct the relevant relations in more systematic terms, with moral language taught first, providing the basis for a set of judgments expressed in terms of it, and certain emotions then rendered moral when annexed to those judgments to enforce moral behavior. But while this picture may afford a clearer explanation of how a philosophic theory of wrong might come to be endowed with motivational force – as in the account of “internal sanctions” of utilitarian morality in Mill 2001 (ch. 3) – it does not help explain how we actually manage to get commonsense morality off the ground. Nor is the only alternative to it a developmental picture running in reverse, with distinctively moral emotions such as guilt seen as available early on to provide an independent foundation for moral judgment (cf., e.g., Nichols 2008).

Instead, we learn moral emotions and judgments in the same breath. Even if the rough-and-ready rules we learn in childhood ultimately give way to more complex principles, the way they

develop initially in conjunction with emotions thus tends to leave us with a tendency toward emotional discomfort at rule violation. It therefore erects a barrier to the inclination in adult life to discount moral requirements as reasons for action. While an agent can legitimately discount some reasons – set them aside as motivational influences, perhaps by appeal to higher-order considerations on the order of personal priorities – reasons of the sort that ground moral requirements are not supposed to be subject to discounting insofar as they incorporate criticism from other agents’ standpoints, standpoints an individual agent lacks authority to discount (see Greenspan 2007 and Greenspan forthcoming-a). In practice, though, we sometimes do set moral reasons aside when they compete with our personal aims. But the fact that moral language typically retains a penumbra of feeling from the process by which it was learned gives us a reason against discounting that at least makes it harder to manage, given that emotional reactions are not normally malleable case-by-case on the basis of rational reflection. They are subject to alteration, of course, but on the basis of the long-term cultivation of new reactive habits that is familiar from Aristotle 2002. The upshot is that emotions function as barriers to rational discounting and are useful to us morally because they do.

Emotional learning may be seen as a way of building social norms into individual practical reasoning and motivation. I argue in Greenspan 1988 (ch. 6; cf. Frank 1988) for a sense in which emotions serve to undercut weakness of will. In prudential reasoning, they bring the future to bear on the present standpoint – reflecting envisioned consequences of action in more or less immediate emotional comfort or discomfort. In relation to ethics, they reflect the social in the individual standpoint, bringing home to an agent the consequences of her action for others. I see them as reinforcing moral reasons with further practical reasons in the agent’s interests: reasons for sustaining or minimizing emotional affect. That I might feel guilty if I did something, for instance, is a reason for me to avoid doing it, given that guilt is an unpleasant feeling I have reason to avoid undergoing. Thus understood, emotions do not function merely as blind psychological forces acting upon us (cf. Greenspan 2009). Instead, in adult life they play a kind of back-up role within practical reasoning (see Greenspan 1988, ch. 6, and Greenspan 2004; I also plan to extend and clarify this

argument in light of later work on reasons in Greenspan forthcoming-b). Of course, explicit appeal to how we feel would keep us from acting for exclusively moral reasons in the way favored by a Kantian approach. But what is normally in question in adult life is just the availability of emotions as reinforcement for moral reasons that we otherwise might be inclined to discount. At the childhood stage at which emotions launch awareness of moral reasons, moreover, purity of moral motivation is hardly a reasonable concern. Moral reasons have not yet begun to be distinguished as such at that stage, and this initial blurring of boundaries is part of what supports moral learning.

I should note that I do not see guilt and similar emotions simply, or even primarily, as serving to bring agents back into line after a norm violation. In anticipatory form – when brought to bear on the thought of a possible future norm violation – what amount to self-punitive responses to moral wrong also motivate adherence to the norms and thus tend to minimize violations. So a moral code that was not learned in conjunction with emotions would be at a serious disadvantage with respect to viability. “Wrong” and similar moral terms and concepts would carry no more motivational force than, say, “out-of-bounds,” in a game that any particular agent might be uninclined to play.

Emotions constrain the content of a moral code, then, insofar as norms based on considerations too remote from ordinary human concerns or too complex to be teachable in the way indicated would not be viable. However, the important point for the question of objectivity is that contemporaneous emotions do not supply the content of ethics on this account. Instead, certain emotions in adult life have come to incorporate moral reasons that by that stage are capable of independent formulation. The moral reason for feeling guilty – that the act one is contemplating would harm someone, say, or that it therefore would be wrong – supplies the content of the emotion. Despite having developed in conjunction with the element of emotional affect, with motivational force that depends on that history, it can be fine-tuned separately at a more advanced stage. In effect, emotions involve *two* layers of reasons, corresponding to the affective and evaluative components defended in Section 1. At the advanced stage we can pull apart the element of affect from a moral reason stemming initially from the simple precepts of caretakers, but ultimately, as Rawls allows,

capable of reflecting principles too complex to be accessible to ordinary moral consciousness. Whatever objectivity applies to moral reasons derived from these sources – possibly interpreted in realist rather than Rawlsian constructivist terms – will not be undermined by their also having a motivational basis in emotion.

Now, besides being motivators, emotions on an evaluative account also come to have an *epistemic* function. Often our readiest means of access to the norms we accept, sometimes apparently our only means of access, is to ask ourselves how we feel about some proposed course of action. Emotions have even been said to amount to perceptions of value (see esp. Sousa 1987). But we need to understand any perceptual role in light of the ability of emotions to incorporate content from an independent source subject to philosophic refinement and grounding.

We can all think of cases where our emotions incorporate values taught to us as children but now superceded, so that they turn out to be *illusory* perceptions of value. A similar point applies to cross-cultural variation: different moral codes can be learned in conjunction with emotions, but some of them, maybe all of them, will turn out to get things wrong in some way or other, perhaps even systematically. This is compatible with a claim of objectivity that is meant to apply just to some unspecified moral code, possibly one that has not yet been realized or even conceived.

What if it were suggested, though, that in the process of working toward the right or most reasonable system of moral norms, the role of emotion would eventually be superceded? Given the tendency of emotions to resist revision, as just noted, along with other well-known pitfalls of emotional response, freeing morality of emotional influence is a tempting ideal for philosophers and others. Sometimes this ideal is summed up in the image of the “Star Trek” character “Mr. Spock” (though Mr. Spock did show signs of compassion). More realistically: we often manage to satisfy moral demands just on the basis of habit, without occurrent emotion. Moreover, reflecting critically on the moral reactions instilled in us as children can depend on setting emotions aside. For that matter, we have to be able to override various emotions (including even some moral emotions) when they undermine moral resolve. So emotions might be depicted as a ladder to adult morality that we

really ought to kick away once we get there.

Alternatively, one might grant that moral action in adult life sometimes has to rely on emotions, but only as “heuristics”: mental short-cuts, or rough aids to decision-making that sometimes misfire and yield the wrong answer (cf., e.g., Kahneman et al 1982; for a more recent approach, stressing the benefits of heuristics, cf. Gigerenzer et al 1999 and 2002). Note, however, that the heuristic role of emotion is in the first instance epistemic: emotions aid us – most of the time, though not always – in quickly forming a correct moral judgment (cf. the treatment of “snap” evaluations in Greenspan 1988, ch. 2). Whether we act on a moral judgment once we have formed it is another matter, however, and that is what my argument here concerns. On my account emotions serve in part to constitute morality as such, by endowing moral judgments with motivational force. The heuristics suggestion is helpful in capturing the role of emotions as an adaptive resource, but only if we resist the conclusion that unlimited time to access and apply an independently worked-out moral system would let us dispense with emotions in moral life.

This is not to say that it would be impossible to act morally at an advanced stage solely on the basis of intellectual apprehension of moral reasons, at any rate some of the time. I do not endorse the Humean claim that reason without emotion is necessarily inert; I mean just to question how reliable a motive that would be in a creature with other, possibly conflicting motives, and thus how viable a moral code could be that relied solely upon it.

The image of the ladder that eventually is kicked away also raises a different question: could the right moral system, once it is elaborated, turn out to be one that fails to elicit any of the emotional responses that figured in moral development? Giving the system motivational force might then seem just to require annexing it to some *new* motives – in the simplest terms, a desire to achieve some social end on the order of group flourishing or some other conception of the total good. On a certain reading of Mill 2001 (ch. 4), we need only engineer that social end into individual motivation.

We had better start early, though – and where else could we start *from* than our stock of innate emotional responses and mechanisms of social influence? The engineering would consist in

rearranging the circumstances that elicit emotional response, more or less as Mill's account suggests. Moreover, any later transitions to a motivationally effective replacement code would have to rely on continuity with our responses as so far constituted. Without that, as far as I can see, we would just be replacing morality with something else: a nonmoral system of behavioral rules aimed toward achieving a social end that might or might not mean anything to a particular agent. The replacement might improve our theoretical decisions on cases considered from a removed standpoint but still be a very bad bet for influencing individual practical reasoning. An attempt to "program in" the social end at a later stage of development could succeed only if we are at that point still able to modify our emotional responses in the way needed to back it up. But that, once again, would mean learning emotions and ethics in the same breath.

REFERENCES

- Aristotle 2002. *Nicomachean Ethics*, trans. by C. Rowe. Oxford: Oxford University Press.
- Armon-Jones, C. 1991. *Varieties of Affect*. Herfordshire, England: Harvester.
- Arnold, M. 1960. *Emotion and Personality*. New York: Academic Press.
- Averill, J. R. 1980. "A Social Constructivist View of Emotion," in R. Plutchik and H. Kellerman (eds.), *Emotion: Theory Research and Experience*, vol. 1. New York: Academic Press.
- Baldwin, T. 2008. "Rawls and Moral Psychology," in R. Shafer-Landau, *Oxford Studies in Metaethics*, vol. 3. New York: Oxford University Press.
- Bedford, E. 1957. "Emotions," *Proceedings of the Aristotelian Society*, 57, 281-304.
- Benedict, R. 1946. *The Chrysanthemum and the Sword: Patterns of Japanese Culture*. Boston: Houghton Mifflin.
- Blackburn, S. 1998. *Ruling Passions*. Oxford: Clarendon Press.
- Bloom, P. 2000. *How Children Learn the Meanings of Words*. Cambridge, Mass.: M.I.T. Press.
- D'Arms, J., and Jacobson, D. 2003. "The Significance of Recalcitrant Emotion (Or: Anti-Quasijudgmentalism)," in A. Hatzimoyisis (ed.), *Philosophy and the Emotions: Royal Institute of Philosophy Supplement 52*, pp. 127-45.
- Darwin, C. 1965. *The Expression of the Emotions in Man and Animals*. Chicago: University of Chicago Press.
- _____. 1981. *The Descent of Man, and Selection in Relation to Sex*. Princeton, N.J.: Princeton University Press.
- Descartes, R. 1970. "The Passions of the Soul," in *The Philosophical Works of Descartes*, Vol. I, trans. by E. S. Haldane and G. R. T. Ross. Cambridge: Cambridge University Press.
- Doi, T. 1973. *The Anatomy of Dependence*. Tokyo: Kodansha International.
- Dwyer, S. 1999. "Moral Competence," in K. Murasugi and R. Stainton (eds.), *Philosophy and Linguistics*. Boulder, Colorado: Westview.
- Ekman, P. 1971. "Universals and Cultural Differences in Facial Expressions of Emotion," in J. K.

- Cole (ed.), *Nebraska Symposium on Motivation*, vol. 19. Lincoln: University of Nebraska Press.
- Ekman, P., and Friesen, W. V. 1986. "A New Pan-Cultural Facial Expression of Emotion," *Motivation and Emotion*, 159-68.
- Frank, R. H. 1988. *Passions Within Reason: The Strategic Role of the Emotions*. New York: Norton.
- Frijda, N. H. *The Emotions*. Cambridge: Cambridge University Press.
- Gibbard, A. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, Mass.: Harvard University Press.
- Gigerenzer, G. and Selten, R. 2002. *Bounded Rationality: The Adaptive Toolbox*. Cambridge, Mass.: M.I.T. Press.
- Gigerenzer, G., Todd, P. and the ABC Research Group. 1999. *Simple Heuristics that Make us Smart*. Oxford University Press.
- Greenspan, P. S. 1988. *Emotions and Reasons: An Inquiry into Emotional Justification*. New York: Routledge, Chapman and Hall.
- _____. 1995. *Practical Guilt: Moral Dilemmas, Emotions, and Social Norms*. New York: Oxford University Press.
- _____. 1998. "Moral Responses and Moral Theory: Socially Based Externalist Ethics," *Journal of Ethics*, 2, 103-22.
- _____. 2003. "Emotions, Rationality, and Mind/Body," in A. Hatzimoysis (ed.), *Philosophy and the Emotions: Royal Institute of Philosophy Supplement 52*, pp. 113-25.
- _____. 2004. "Practical Reasoning and Emotion," in A. Mele and P. Rawlings, *Handbook of Rationality*. Oxford: Oxford University Press.
- _____. 2007. "Practical Reasons and Moral 'Ought,'" in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. II. Oxford: Oxford University Press.
- _____. 2009. Interview, in J. H. Aguilar and A. A. Buckareff (eds.), *Philosophy of Action: 5*

- Questions*, pp. 59-69. London: Automatic Press/VIP.
- _____. forthcoming-a. "Moral Reasons, Imperfect Duties, and Choice," *Social Philosophy and Policy*.
- _____. forthcoming-b. "Emotions and Moral Reasons," in C. Bagnoli (ed.), *Morality and the Emotions*. Oxford: Oxford University Press.
- Griffiths, P. E. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.
- Grimes, K. 2003. "To trust is human." *New Scientist*, 178, 32-37.
- Haidt, J. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review*, 108, 814-834.
- _____, Koller, S. H., and Dias, M. G. 1993. "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?" *Journal of Personality and Social Psychology*, 65, 613-28.
- Hauser, M. D. 2006. *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: HarperCollins.
- _____. 2000. *Wild Minds: What Animals Really Think*. New York: Henry Holt.
- Hoffman, M. L. 1982. "Development of Prosocial Motivation: Empathy and Guilt," in N. Eisenberg-Berg (ed.), *Development of Prosocial Behavior*. New York: Academic Press.
- Hume, D. 1978. *A Treatise of Human Nature*. Oxford: Clarendon Press, 1978.
- Isci, G. S., Direnc, D., and Oktem, Z. C. 2005. *Inside Outside In: Emotions, Body and Society*. Izmir, Turkey: Ege University Press.
- Izard, C. E. 1977. *Human Emotions*. New York: Plenum.
- Kahneman, D., Slovic, P. & Tversky, A. (eds.) 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Kant, I. 1993. *Grounding for the Metaphysics of Morals*. Indianapolis, Ind.: Hackett.
- LeDoux, J. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster.
- Lutz, C. 1986. "The Domains of Emotion Words on Ifaluk," in R. Harré (ed.), *The Social*

- Construction of Emotions*. London: Oxford University Press.
- Mackie, J. L. 1995. *Persons and Values: Selected Papers*, Vol. II. Oxford: Clarendon Press.
- Mill, J. S. 2001. *Utilitarianism*, 2nd ed. Indianapolis: Hackett.
- Miller, W. I. 1997. *The Anatomy of Disgust*. Cambridge, Mass.: Harvard University Press.
- Nichols, S. 2008. "Sentimentalism Naturalized," in W. Sinnott-Armstrong (ed.), *The Psychology and Biology of Morality*. Cambridge, Mass.: MIT Press.
- Panksepp, J. 1998. *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press.
- Pinker, S. 1997. *How the Mind Works*. New York: W. W. Norton.
- _____. 2002. *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.
- Rawls, J. 1980. "Kantian Constructivism in Moral Theory," *Journal of Philosophy*, 77, 515-72.
- _____. 1998. *A Theory of Justice*, rev. ed. Cambridge, Mass.: Harvard University Press.
- _____. 1993. *Political Liberalism*. New York: Columbia University Press.
- Robinson, J. 1995. "Startle," *Journal of Philosophy*, 92, 53-74.
- Rolls, E.T. 1999. *The Brain and Emotion*. Oxford: Oxford University Press.
- Rozin, P., and Fallon, A. E. 1987. "A Perspective on Disgust," *Psychological Review*, 23-41.
- Shweder, R. A., Mulch, N.C., Mahapatra, M., and Park, L. 1997. "The 'Big Three' of Morality (Autonomy, Community, and Divinity), and the 'Big Three' Explanations of Suffering," in A. Brandt and P. Rozin (eds.), *Morality and Health*. New York: Routledge.
- Smith, A. 1982. *A Theory of the Moral Sentiments*. Indianapolis: Liberty Fund.
- Solomon, R. C. 1976. *The Passions*. New York: Doubleday.
- Sousa, R. de. 1987. *The Rationality of Emotions* Cambridge, Mass.: M.I.T. Press.
- Sripada, C., and Stich, S. unpublished. "Explaining Social Norms: Psychological and Evolutionary Foundations."
- Turiel, E. 1983. *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press.

Waal, F. de 1998. *Chimpanzee Politics: Power and Sex among Apes*. Baltimore: Johns Hopkins University Press.

Williams, B. 1993. *Shame and Necessity*. Berkeley: University of California Press.

Wilson, J. Q. 1993. *The Moral Sense*. New York: The Free Press.

Witherington, D. C., Campos, J. J., and Matthew J. Hertenstein, M. J. 2002. "Principles of Emotion and its Development in Infancy," in G. Bremner and A. Fogel (eds.), *Handbook of Infant Development*. Oxford: Blackwell.

1. I owe thanks to Stephen Leighton for comments on a very early draft of this paper. A poster based largely on Section 1 was presented to the meetings of the International Society for Research on Emotion in 2002. Later versions of this interdisciplinary part of the paper were presented at meetings of the U.K. Arts and Humanities Research Board project on "Innateness and the Structure of the Mind" in 2002-3 and at Ege University's Cultural Studies Seminar in Izmir, Turkey, along with a philosophy colloquium at Bosphorus University in Istanbul in 2003; it appears along with other contributions to the CSS conference in G. S. Isci 2005, pp. 17-31. Let me also thank members of these audiences for helpful reactions. Section 2 was written for the present volume.