Moral psychology is both old and new. Old because moral thought has long been a central focus of theology and philosophy. Indeed, many of the theories that we explore today were proposed first by scholars such as Aristotle, Kant, and Hume. Young because the scientific study of morality—and, specifically, the study of what goes on in a person’s head when making a moral judgment—has been a topic of serious inquiry only over the last couple of decades. Even now, it is just barely mainstream.

This chapter is itself a combination of the old and the new. I am going to consider two broad questions that would have been entirely familiar to philosophers such as Aristotle, but are also the topic of considerable contemporary research and theorizing: (1) What is our natural human moral endowment? (2) To what extent are moral judgments the products of our emotions? I will have the most to say about the first question, and will review a body of empirical work that bears on it; much of this research is still in progress. The answer to the second question will be briefer and more tentative, and will draw in part upon this empirical work.

Many leading researchers, including other contributors to this volume, have their own answers to these questions, and at every point in this chapter, I will do my best to make contact
with their work, being explicit about areas of overlap and disagreement. To anticipate, my conclusions are as follows: Humans possess an innate and universal system for moral evaluation. And emotions play a significant role in moral judgment—but so does deliberative reason.

What is our Natural Human Moral Endowment?

The Question of Moral Universals

I will start by putting the question somewhat bluntly: How much of moral understanding is innate? It is important to clarify the terms here; one traditional meaning of “innate” is “present at birth,” but this is not how I am using the word—or how anyone else uses it these days. Rather, by “innate” I mean “not learned”—and by “learned,” I mean, roughly, “gets into the head by means of the extraction of information from the environment.”

This is not meant to be an airtight definition (for discussions of some of the subtleties here, see Ariew, 1999; Bloom, 2000). But it captures the right intuitive contrast. Nobody thinks that the rules of baseball are innate, and this is because it is obvious that they do not start off inside anyone’s head; they become known only through exposure to the relevant information in the world. That is, everyone who knows the rules of baseball does so because he or she has been told about them or read about them or figured them out by watching a baseball game. Indeed, there are cultures without baseball, and children raised in such cultures do not have any baseball knowledge. In contrast, there is considerable evidence that human babies possess certain systems of knowledge that have not been extracted from the environment, such as a foundational understanding of physical objects and numbers (e.g., Baillargeon, Spelke, and Wasserman, 1985; Wynn, 1992, 1998). This knowledge is universal, showing up in all human groups. It is likely to
be innate.

This helps frame the question regarding moral beliefs and moral feelings. If they are learned, then one would expect them to emerge as a result of exposure to the right sort of information. They should not be present in children’s minds before this information is accessible to them. And they might vary substantially across cultures. Conversely, if they are innate, one might expect cross-cultural universals. Some understanding of morality should emerge without any exposure to relevant information, and thus may show up in young children and even babies.

Many philosophers and psychologists would defend the view that morality is learned. Perhaps the most thorough defense of this view is the one developed by Jesse Prinz (2004, in press). I want to focus here on his argument that there are no substantive moral universals.

He starts by exploring a specific proposal, which is that we have a universal prohibition against harm, a moral response to the suffering of others (for proposals along these lines, see Blair, 1995; Turiel, 2002). He then provides several examples that suggest that this claim is mistaken. It is not merely that people harm one another all the time, it is that people are often morally untroubled by harming others. Indeed, in many cases, they believe that harming others is morally praiseworthy.

Prinz makes his case through exotic examples from faraway cultures, such as the brutal customs of the Ilongot of Luzon. But his claim is a plausible one, and is supported by more familiar cases. Most people in industrial societies eat meat that was created through imposing terrible conditions on animals, causing them great pain, but few are morally troubled by it, even if they are fully aware of these conditions. Most people believe it is acceptable to harm or even kill another adult in self-defense, or in the course of defending those we love, or in times of war,
or even to protect one’s property or dignity. Some would go further and view such an act as praiseworthy. A police officer who killed a terrorist about to set off a bomb would be seen as a hero. Indeed, this would be so for at least some of us even if the terrorist’s death was not instrumental in stopping the bomb: When such scenes are shown in movies, audiences cheer. Most people think it is fine to cause mild pain or injury during sport. Most people believe that it is morally right to harm criminals, such as by taking away their property, putting them in cages, or even killing them. And most parents think it is morally obligatory to harm their own children, with the harm being physical, as in spanking, or emotional, as when punishing a child by taking away certain treats. Such examples can easily be multiplied.

Prinz is right, then, to conclude that there is no general prohibition against causing harm. But there is a reasonable nativist response here. Any plausible hard-wired prohibition is going to be more subtle than “Do not harm.” This is because any innate moral notion is likely to have evolved through natural selection, and any tenable evolutionary account would predict that harm can be morally neutral or even morally positive under certain delineated circumstances. For instance, no nativist should deny that our moral intuitions are calibrated to the genetic and social distance from whomever is harmed. It is no surprise, then, that we are more outraged by harm to ourselves and those we love than harm to our neighbors, or that we care less about strangers in faraway lands and often not at all about non-human animals. Furthermore, no nativist would deny that our moral reaction to harm must be balanced against other moral and non-moral priorities. This is why it is morally fine to scream at your child — purposefully upsetting her, causing her harm — if she walks into the street without looking for cars. Such examples, then, as well as others raised by Prinz, are consistent with the notion that we have an evolved moral
Prinz considers this objection, and responds with the argument that such nuanced thoughts and actions—the prohibition of “just some harms”—are most parsimoniously explained without nativism, through general considerations of rationality and concerns about social stability. He asks, “How would society work if you could punch your neighbor regularly?” If all we have left is the platitude “Harm when and only when the pros outweigh the cons,” then we need not appeal to hard-wired universals. Prinz goes on to make parallel arguments for the moral feelings associated with reciprocity and sharing. There is considerable cross-cultural variation in the extent to which people feel obliged to share, trade, and cooperate with others (see Henrich et al., 2004), and, insofar as there are cross-cultural generalizations, they can be understood as cultural constructions, invented and sustained because of the work they do for members of a society.

This is an intriguing view, and is surely true to some extent. Moral intuitions do emerge in part as a response to societal conditions. One well-known example is that of so-called “Cultures of Honor.” Historically, these are societies in which resources are vulnerable and people cannot rely on external authority to protect these resources and mete out justice. A reputation for toughness matters a lot in such situations, and hence people in such societies are, more than most others, disapproving toward acts of disrespect and forgiving of acts of retribution (see Nisbett & Cohen, 1996, and Cohen & Leung, 2010, for reviews and empirical support). Similarly, Abarbanell and Hauser (2010) find some unusual patterns of moral judgment in their study of a Mayan community (the individuals in this community apparently lack an “omission bias”)—and argue that this moral property is due to some unusual features of that small and
isolated community. Culture does shape morality, then, sometimes in a rational manner.

In the end, though, it seems unlikely that all moral responses are cultural constructions that work for the good of society. After all, one of the most salient facts about our everyday moral responses is that they often have awful consequences. Consider our intuitions about punishment, which underlie the desire for vengeance. This desire may have been of adaptive value when we lived in small groups, but it is wasteful, at best, in the world in which we now live (see Greene and Cohen, 2004). People have died, needlessly, in attempts to settle a score, often reacting with violence to an injustice that is, in an objective sense, utterly trivial. Or consider our strongly tribal impulses, which lead us to morally value those who are close to us, particularly those related by blood, far more than anyone else. These impulses make evolutionary sense, but often cause savage discord in contemporary nation-states. Such responses seem Stone Age, and we would be better off, right now, without them.

In the end, this is an empirical issue; we can explore just how much variation there really is, and the extent to which it is best explained as a response to social and economic forces (see, e.g., Henrich et al., 2005). We can also explore whether certain moral notions are present in non-human primates (de Waal, 1996; Santos, Hauser, Chen, Chen, & Chuang, 2003; Silk, 1987). After all, if capuchin monkeys have some of the same moral intuitions as people, it is an excellent bet that these intuitions are not cultural inventions, since capuchins do not have culture – at least not the sort that humans have. (The inference only works one way, though: If capuchins do not have such intuitions, it does not necessarily support the cultural theory. It might be that they lack the right innate endowment.)

A different data source comes from my own field of developmental psychology. If moral
judgments are invented by societies, then they should not be present in babies; they need to be learned. In contrast, if they are innate, they should not require exposure to certain relevant sources of information, and might be present even in pre-linguistic babies.

Psychologists have long been interested in moral development, much of this because of Lawrence Kohlberg’s classic proposals about stage theories of moral reasoning (e.g., Kohlberg, 1981). Ongoing research programs include studies of the emerging capacity to distinguish between moral and conventional violations (e.g., Smetana, 2006; Turiel, 1998), to understand fairness and justice (e.g., Damon, 1998), to understand the role of intentionality in moral responsibility (e.g., Leslie et al., 2006), and to respond with moral disapproval to certain disgusting acts (e.g., Haidt et al., 1993).

But there has been little research thus far into moral evaluation in infants or toddlers, and so—unlike domains such as naïve physics, naïve psychology, and number—the more direct evidence that could help support or falsify specific developmental claims of the origin of moral judgments is thin on the ground. It has been only in the last five years or so that the relevant experiments have been done, and this is what I will discuss below.

The Question of Moral Babies

It has long been known that babies grow anxious when others are in distress. Charles Darwin (1872/1913, p.358) tells this story of his first son, William: “When this child was about four months old, I made in his presence many odd noises and strange grimaces, and tried to look savage; but the noises, if not too loud, as well as the grimaces, were all taken as good jokes; and I attributed this at the time to their being preceded or accompanied by smiles.” But then William was fooled by his nurse: “When [he was] a few days over six months old, his nurse pretended to
cry, and I saw that his face instantly assumed a melancholy expression, with the corners of his mouth strongly depressed.”

Babies cry when they hear other babies cry (Simmer, 1971) and, once they have enough physical competence, they soothe others in distress. This has been observed both spontaneously and in controlled studies where an adult (typically, the child’s mother) pretends to be pain (Sagi & Hoffman, 1976; Zahn-Waxler et al., 1992). This intentional soothing of others is not uniquely human: de Waal (2001) observes that chimpanzees—but not monkeys—“will approach a victim of attack, put an arm around her and gently pat her back, or groom her” (p. 326).

More recent studies have found elaborate altruistic behavior by toddlers, behaviors in which they give up their time and energy to help strangers, without being prodded or rewarded. If an adult reaches for something beyond his reach, for instance, the toddler will often just hand it over to him, even if the adult doesn’t ask or even look at the toddler (Warneken & Tomasello, 2006, 2009; see also Wynn, 2009).

How do these sensitivities and inclinations relate to mature morality? The problem here is that there is no consensus about what is distinct about the domain of morality even for adults—no accepted definition or list of criteria (Nado et al., 2006). But, presumably, a mature moral psychology entails the capacity to evaluate different acts and different people; this evaluation ties into systems of praise and blame, into emotions such as anger, gratitude, and guilt, and, ultimately, to broader issues of fairness and justice. What we find so far in babies falls far short of this. Perhaps their responses reflect nothing more than a set of reactions and motivations—the pain of others is upsetting to them; they are compelled to help under certain circumstances. It is conceivable that such reactions are no different in kind from purely non-moral and non-social
reactions and motivations, such as being frightened at a loud noise or wanting to void a full bladder.

At the same time, though, it is hard to conceive of a moral system that does not have such responses as a starting point: A creature that lacked them might never attain a moral system. The point here is owed to David Hume; pure reason gets you only so far. To have a genuinely moral system, things have to matter: "‘Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger.” What we see in these emerging emotions is the development of mattering. The woes of others are painful to the children, and when they are capable, they are motivated to act so as to make these woes go away.

Do babies also possess the ability to evaluate the acts of others? This is the topic of considerable ongoing research (see Wynn, in press, for review). I can illustrate the logic of many of these studies with the following case: Suppose an individual A is trying unsuccessfully to climb a steep hill. This event continues in one of two ways. In one, individual B then appears on the scene and repeatedly pushes A down the hill until A is fully at the bottom. In the other, individual C appears on the scene and repeatedly nudges A up the hill until A reaches the top.

This paradigm is partially based on a classic study by Premack and Premack (1997), who showed babies a series of computer-animated displays in which one intentional agent acted either positively or negatively towards another, and they used a looking-time measure to see how the babies grouped the actions together. The babies’ pattern of looking times suggested that helping someone push through a gap was grouped along with caressing the person, and preventing someone from going through a gap was grouped along with hitting the person. Premack and Premack suggested that infants appreciated that the former two actions, though perceptually very
different from one another, were positive and the second two acts were negative.

This finding motivated a series of studies in which we explored babies’ intuitions about scenes with helping and scenes with hindering. Our initial studies asked about babies’ expectations of how individual A would respond to the hinderer (B) versus to the helper (C). As an adult, if you were to witness such scenes, you would likely expect A to later tend to avoid B and to approach C. Using a looking-time method, we found that 12-month-olds and even 9-month-olds had these same intuitions (Hamlin, Wynn & Bloom, 2007; Kuhlmeier, Wynn, & Bloom, 2003). (See Figure 1.)

What about babies’ own preferences for these two characters? In Hamlin et al. (2007), we tested this by showing babies a three-dimensional display in which real objects acted out the interactions, again with B hindering A and C helping A. Then we offered the babies themselves the choice between reaching for B or C. (See Figure 2.) We found that both 6- and 10-month-old infants overwhelmingly preferred the helpful individual to the hindering individual.
Now, this preference result is ambiguous. It could mean that babies are attracted to the helpful individual, or that babies are repelled by the hinderer, or both. We explored this in a further series of studies that introduced a neutral character, one that neither helps nor hinders. We found that, during choice, infants prefer a helpful character to a neutral one, and prefer a neutral character to a hindering one. This shows that both inclinations are at work—infants are both drawn to the nice guy and repelled by the mean guy.

Our preferred interpretation of these results is that they reflect a social competence, in which babies’ expectations about others and their own preferences are motivated by the perceived goodness and badness of the characters. But there is always the worry that perhaps some other non-social feature, particular to the stimuli, was triggering these responses. Of course, our experiments were designed in the hopes of avoiding such confounds, by counter-balancing the colors and shapes of the characters, matching the intensities of the helping and hindering movements, and so on. In addition we also included “non-social” controls in which the characters exhibited similar movement patterns but were not depicted as animate beings. As predicted, these did not elicit the same preferences (see Hamlin et al., 2007, for details). But still, it is useful to see whether these findings replicate more generally, across different sorts of stimuli.
This concern motivated a series of studies with 5-month-olds by Hamlin and Wynn (under review). They also involved actions that were, by adult lights, positive or negative, but in very different contexts than the hill. In the first study, an individual struggled to open a box, getting it partially open but then watching the lid fall back down. Then, on alternating trials, one puppet would grab the lid and open it all the way, and another puppet would jump on the box and slam it shut. Would babies prefer the puppet that (by adult intuitions) was the prosocial one and avoid the one that (by adult intuitions) was antisocial? (See Figure 3.)

In a second study, an individual played with a ball. On alternating trials, it would either roll the ball to a puppet who would roll it back or it would roll the ball to a puppet who would run away with it. Again, the question was whether babies would prefer to interact with the
character who was (by adult intuitions) helpful over the one who was nasty? (See Figure 4.)

In both scenarios, 5-month-olds preferred the prosocial character over the antisocial one. In other research, using a preferential looking paradigm, we replicated this finding with infants as young as 3 months of age (Hamlin, Wynn, & Bloom, in press).

Further studies explore responses more elaborate than simple preference. Part and parcel of adult morality is the notion that good acts should meet with a positive response and bad acts with a negative response. To see whether babies have the same intuition, we exposed 21-month-olds to the good guy/bad guy situations described above, and gave them the opportunity to hand over a treat to either of the characters. They overwhelmingly chose to give it to the positive character. In contrast, when asked to take a treat from one of the characters, they tended to take it from the bad guy (Hamlin et al., under review).
We then extended this paradigm by exploring reward and punishment once-removed (Hamlin et al., under review). In one condition, we showed babies a set of scenarios in which a puppet was trying to open a box; on some occasions a helpful puppet joined in and helped get the box open, and on other occasions a hindering puppet jumped on the box lid, slamming it shut. This is one of the good guy/bad guy situations tested by Hamlin and Wynn (under review), which found that babies preferred the helpful puppet. But here we took an additional step. We showed babies a second set of scenarios with these same characters as the targets of other actions. That is, we showed babies either the original helper or the original hinderer in a different situation in which it was itself treated positively by one new character and negatively by another new character. The babies were then asked to choose between the two new characters.

Did they prefer the character who rewarded the Good Guy over the one who punished the Good Guy? They did—even 8-month-olds preferred the rewarder. In itself this isn’t very surprising, given the other studies showing a preference for those who act pro-socially. That is, they would have likely shown the same preference even if they forgot, or had never encoded, the fact that the character who was rewarded and punished had behaved well in the past. What is far more interesting is the condition in which they watched the Bad Guy being rewarded or punished. Here they chose to interact with the Punisher. Infants like those who harm Bad Guys. They like punishers.

To sum up, the results so far suggest that even young babies make sense of a range of interactions—helping someone up a hill versus hindering the character’s progress, helping someone open a box versus slamming it shut, passing back a ball versus absconding with it—the same way that adults do, in terms of prosocial and antisocial behavior. Babies expect whoever
was the target of such behavior to later associate with the prosocial character over the antisocial one, and they themselves prefer to reach for (or, for the younger babies, to look at) the good guy over the bad guy. They prefer to reward the good and punish the bad, and prefer others who reward the good and punish the bad.

**But Is It Moral?**

This is evidence for early social evaluation. But a critic might point out that none of this has to be moral in any interesting sense. It could just be a social preference. To see the difference, consider an adult example: Suppose you observe that Jane always hands out candies to those she meets, while Joe tries to hit people with an axe. You might reasonably infer that people would prefer to socialize with Jane over Joe, and you yourself would prefer Jane over Joe. If forced to choose, you would prefer to reward Jane and punish Joe, and would prefer characters who do the same. After all, you would like to encourage Jane’s actions and discourage Joe’s actions. But all of this might just draw upon a rudimentary understanding of others and a healthy allotment of basic self-interest. Your expectation and preference would not necessarily demonstrate that you thought that Jane was somehow good (moral) and Joe was somehow bad (immoral) or that you attributed moral value to their actions of candy-giving and axe-waving. Indeed, even a pure psychopath—a hypothesized individual with no moral sense at all, but with normal intelligence and social understanding—might have the same responses. So how can we draw conclusions about the moral lives of babies?

For adults, distinctive aspects of moral intuitions can be expressed in language. For instance, when describing Joe and Jane and what they are up to, adults would likely use terms like “right” and “wrong,” and express notions such as “should” and “shouldn’t.” We can separate
morality from preference, acknowledging that there are cases in which one prefers X to Y (people or actions) but thinks that Y is more moral than Y. You might say, “I like when X does such-and-so, but I don’t approve of it” or “It really annoys me when Y does such-and-so, but I have to admit that she is doing the right thing” or that “Nobody should do such a thing.” You might express second-order desires: “I disapprove of X, and I want to disapprove of X.” But of course, the expression of this sort of sophisticated understanding is far beyond the capacity of babies or even young children.

The question remains, then, whether babies already possess the core of what we would describe in adults as a moral sense. It would be premature to answer this with confidence, but there are four considerations that bear on the question, and, taken together, they favor an affirmative answer.

First, they respond to (what for adults are) morally relevant properties of the actions. Babies were not reacting to the characters’ prettiness, their rapidity of movement, or anything like that. They were responding to how one individual treated another, either helping that individual achieve its goals or hindering it. They were resonating to the difference that adults would describe as “nice guys” versus “mean guys,” and, indeed, when we ask 18-month-olds “Who is the nice guy?” and “Who is the mean guy?”, they respond in adult-like ways (Hamlin et al., in preparation). Keep in mind also that in all of these studies, the babies are never themselves helped or harmed by these characters; their intuitions and predictions are based on observations of third-person interactions, observations of characters that aren’t even human: animated figures, blocks of wood with eyes on them, or small puppets. Their responses and expectations, then, are not based on either self-interest or personal experience.
Second, the competences that we chose to explore were not pulled from the air. They are just what you would expect from theories of the evolution of social and moral behavior. It is plainly useful to distinguish supportive and friendly individuals from harmful and malicious ones, and respond accordingly. Indeed, some ability to evaluate others is essential for navigating the social world—perhaps just as essential as the predictive and interpretative capacities that are the topic of so much developmental research. Furthermore, it has long been noted that cooperative behavior such as group hunting and food sharing can be beneficial to the individual members of a group, but it can successfully emerge in populations of unrelated individuals only if they have the capacity to distinguish free-riders (or “cheaters”) from those willing to do their fair share (e.g., Axelrod, 1984; Cosmides, 1989; Hamilton, 1964; Trivers, 1971, 1985). It makes sense, then, for us to evolve a particular sensitivity to “bad guys,” along with a wish to punish them and to approve of others who punish them. For many scholars, this is the explanation of the foundation of morality.

Note, however, that even if this evolutionary analysis is right, it does not entail that these notions are innate. It is one of the axioms of evolutionary developmental biology that adaptations need not be hard-wired; evolution can work so that an adaptive capacity emerges in the course of normal development (see Goodman & Coughin, 2000; Jablonka & Lamb, 2007). All natural selection cares about (to put it metaphorically) is that we end up with such capacities; it makes no claim about how they get there. It is possible, then, that the foundations of moral evaluation are evolutionary adaptations, but that, contrary to our hypothesis, these foundations are learned. Innateness is a hypothesis, not a logical truth. It is just a hypothesis that has been supported.

Third, there is some tentative evidence that babies respond to our characters with the
same emotions that are, for adults, infused with morality—emotions such as disapproval, concern, and even anger. I will discuss this in more detail in the next section.

Finally, these findings mesh well with previous proposals about the developmental origins of morality. I will end this section by considering two of them.

1. There is an innate moral grammar, on analogy with Noam Chomsky’s proposals concerning an innate language faculty. This analogy was initially proposed by both the political philosopher John Rawls and by Chomsky himself, was developed extensively in the dissertation of John Mikhail (2000), and then extended by Marc Hauser, both with Mikhail and independently (e.g., Hauser, 2008; Hauser et al., 2007).

It is an intriguing idea. There are certain suggestive parallels between language and morality. Chomsky has long observed that our linguistic knowledge (competence) can be distinguished from how we use this knowledge in everyday life (performance). This linguistic knowledge is unconscious or tacit—every English speaker knows that something is wrong with the sentence “John seems sleeping,” but few of us can explain why it’s wrong. Similarly, moral intuitions are imperfectly linked to action; you can know the right thing to do (competence) but choose not to do it (performance). And the reasons for our moral notions are typically not accessible to us; indeed, much of the program in moral philosophy and moral psychology is trying to explain precisely why we find some acts and people morally wrong. These similarities suggest that the same methodological approaches and theoretical ideas that work well for the study of language can be productively applied to the domain of moral thought (Hauser, 2008).

There are also reasons to doubt the aptness of this analogy (see Bloom & Jarudi, 2006). Language is a generative system, where units such as words and phrases are combined through
recursive rules to generate a potential infinity of sentences. Morality does not seem to work that way. Universals of moral knowledge, should they exist, might therefore be better characterized as a small list of evolved rules, perhaps simple (such as a default prohibition against intentional harm), perhaps complex (such as some version of the doctrine of double effect), but still very different in character from linguistic knowledge.

Our developmental results favor the analogy in at least one regard, however. This is in its claim that, similar to core aspects of language, moral intuitions are not entirely learned. The research thus far has explored only the early understanding of fairly crude distinctions, such as the contrast between aiding one’s goal versus hindering one’s goal. It is an interesting program for further research to explore whether the more subtle distinctions explored by Hauser and colleagues—such as the contrast between a harmful event being a side-effect versus its being a means to an end—also show early emergence. If they do, the question would then arise as to whether these subtleties are specific to morality, along the lines of language-specific constraints proposed by Chomsky and his colleagues, or whether they emerge from more general properties of action comprehension (see Abarbanell and Hauser, 2010; Cushman, Young, and Hauser, under review).

2. Another proposal about the innate capacity emerges from the work of Jonathan Haidt and his colleagues (see Haidt, 2001, 2007). This is motivated by observations of cross-cultural variation, looking at the sorts of moral notions that different people possess (see also Shweder, et al., 1997). This theory posits five innate domains of morality: harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity.

How does it mesh with the developmental data presented above? So far, quite well. The
research I have summarized in this chapter can be seen as showing early competence in the domain of harm/care. Other recent studies suggest a rudimentary grasp of fairness/reciprocity: Babies and toddlers prefer characters who equally distribute resources to other people (Geraci and Surian, 2010) and they expect people from the same group to reciprocate kind gestures (He and Baillargeon, 2010a). There is suggestive evidence for early group-based biases: Put crudely, babies prefer their own kind, where kind can be based on physical similarity, sameness of language, or similarity of preference (e.g., Kinzler et al., 2009; Olson et al., 2006). He and Baillargeon (2010b) find that toddlers expect individuals in a group to be fairer to members of the same group than to members of a different group. Finally, in research in progress by Karen Wynn and her colleagues, they are finding that babies are prone to reward same-group members and punish individuals from other groups.

The studies summarized above bear on the first three domains. There is little data, so far, that bear on the other two domains discussed by Haidt: authority/respect and purity/sanctity. These domains emerge late in development, but it is nevertheless possible that they also have some unlearned, universal, basis (for discussion of the moral roots of disgust, see Bloom, 2004; Danovitch & Bloom, 2009).

To What Extent Are Moral Judgments the Products of Emotions?

Emotions Are Essential

There is a popular view about morality, developed by, among others, the philosopher Jesse Prinz (2004, in press), the neuroscientist Antonio Damasio (1995) and the social
psychologist Jonathan Haidt (2001, 2007) that lies remarkably close to the position laid out by
the 18th-century Scottish philosopher David Hume. He argued that reason, and particularly moral
reason, is nothing more than “the slave of the passions.”

A mild version of this view is uncontroversial. Few would doubt that our moral
judgments can be *influenced* by our emotions. There are numerous findings from social
psychology and social neuroscience suggesting that our judgments of good and evil are
influenced by emotional reactions such as empathy and disgust (see Haidt, 2007 for review). And
there is a consensus that Hume’s deeper point is right as well, that there are moral intuitions that
cannot themselves be justified by reason—“self-evident” truths as Thomas Jefferson put it.
Indeed, as Pizarro and Bloom (2003) note, this is the case for *all* domains of reasoning, including
deductive inference and inductive generalization.

But the Humean theory that many contemporary scholars hold is stronger than this. It
largely rejects rational deliberation as playing an important role in moral judgment. Haidt (2001)
makes this point with vigor, arguing that the reasoned arguments we use all the time concerning
why we hold certain moral positions are mostly post-hoc justifications for the gut reactions that
really matter: “Moral reasoning does not cause moral judgment; rather, moral reasoning is
usually a post hoc construction, generated after a judgment has been reached” (p. 814). To use
Haidt’s analogy, we like to think of ourselves as judges, reasoning through cases according to
deeply held principles, but in reality we are more like lawyers, making arguments for positions
that have already been established. With the notable exception of professional philosophers,
conscious deliberation plays little role in determining our moral judgments.

What does the developmental evidence reviewed in the first section have to say about
this? To a large extent, it supports it.

Recall that I discussed two foundations of early moral understanding. The first foundation just was a set of emotional reactions, such as empathy. Plainly, this fits the Humean picture. The second was a rich capacity for moral evaluation, with which babies appear to judge different characters as moral or immoral and reason and act accordingly. Interestingly, these evaluations give the impression of being emotional. As Hamlin (2010, p.138) summarizes, “Anecdotally, babies appear to be more likely to smile, clap, etc. during prosocial events, and to frown, shake their heads, and look sad or otherwise upset during antisocial events.” In addition, at least some babies are aggressive toward the antisocial character. In one of our punishment studies, a toddler took food from the antisocial puppet and then leaned forward and smacked him in the head.

This is all anecdotal so far; the next step is to have observers (who are blind to what the babies are seeing) code the babies’ facial expressions so that we can determine whether they really do exhibit a consistent emotional response to what are, for adults at least, moral and immoral acts. But the anecdotes so far do support the notion that for babies, emotions are intimately related to evaluations.

Now, as Huebner et al. (2009) emphasize, the fact that an emotion corresponds to a moral judgment does not entail that emotions cause the judgment. It is possible, as a Humean would argue, that the baby gets angry at the behavior of the antisocial character and this drives the judgment that the behavior is wrong. But it is also possible that the baby judges the behavior to be wrong and this in turn motivates the anger. The emotion might be relevant only to the motivation of moral action, or possibly it plays no role at all.
Keep in mind also that even if the emotional response is the cause of the moral judgment, the emotion itself is based on a fairly sophisticated analysis of the scene. How can the baby get angry if she does not believe that the character did something wrong? If the baby sympathizes with a victim, cheers on a rescuer, and avoids someone who rewards a bad guy, this entails that the baby somehow figured out who the victim, rescuer, and bad guy are, and this requires a rich understanding of the scene. Emotions must be smart. This is true of gut feelings as a whole—how you think about something effects how you emotionally respond to it. Finding a telephone number in the pocket of one’s spouse can engender intense jealousy in one individual but mere curiosity in another, depending on how the situation is construed. You do not normally respond with fear when you hear someone start to whistle—but you might do so if it was 3 a.m. and you had thought you were alone in the house (Pizarro & Bloom, 2003).

Still, caveats aside, the baby data are at least consistent with the intuitionist analysis. Babies are making moral (or quasi-moral) evaluations, and there is some evidence that these are associated with emotional responses. Surely their evaluations are not the product of deliberative reason. Babies are Humeans. If there was no subsequent development, then the strong Humean view would be right of humans in general. The only difference is that babies’ intuitions are the product of evolution plus some personal experience, while adults have the additional influence of culture. This affects the scope of moral response: Babies respond to hindering; adults respond to treason, necrophilia, and insider trading.

**Rationality Is Essential**

I will suggest here that there is a substantial difference in the moral reasoning of babies and adults. Unlike babies, children and adults have the capacity for rational deliberation. This
underlies one of the most significant aspects of human nature, the existence of moral progress—or, to put it less contentiously, of directional moral change (Bloom, 2004, 2010, in prep.)

The sort of change I am talking about has been explored by many scholars (Pinker, 2002, 2009; Singer, 2007; Wright, 2001). Consider Peter Singer’s discussion of the expansion of the moral circle. There is no doubt that the average person’s sympathies have grown substantially over our history. There was a time when human sympathies never extended beyond the immediate group. Now some of us donate money and even blood to people with whom we have no contact and little in common. Contemporary readers of this chapter have profoundly different beliefs about the rights of women, racial minorities, homosexuals, and non-human animals than readers in the late 1800s. In 1960, most Americans thought interracial marriage was wrong; now very few do. A hundred years before this, many Americans thought slavery was morally justified; now very few do.

I am happy to describe this as “progress” but am aware that this assumes a moral realism that many readers would reject. Actually, the relevant phenomenon can be characterized in more neutral terms. It is that the change is directional. It is not that the circle has remained constant or that it randomly grows and shrinks. It started off small, and over a long history, has grown. This is a critical way in which the history of morality is sharply different from domains such as fashion in clothing and taste in food, domains that exhibit no direction. Indeed, in its cumulative nature, morality looks a lot more like science.

What drives this directional moral change? There are many possible factors, some of them fully consistent with a Humean theory of human morality. One main driver, for instance, is contact: When we associate with other people and share common goals with them, we extend
them our affection (ref). In recent times, increased political and economic interdependence has ensured that we associate with many more people than our grandparents and even our parents ever did, and this trend has accelerated through the virtual contact enabled by television and the Internet. Robert Wright (2001), in particular, has outlined the argument that the cold-blooded forces of increased populations and interactions with others have led to greater niceness. As he puts it, "Among the many reasons I don't think we should bomb the Japanese is that they built my minivan" (cited by Pinker, 1994). Since the extent of interdependence tends to grow over human history, moral change takes on a certain positive direction, all without any sort of rationality or reason.

But this explanation is incomplete. It does not explain the shifts in opinions on issues such as slavery and animal rights. Contact and inter-dependency are not enough: Owning slaves did not turn people into abolitionists, and running a factory farm does not increase sympathy for cows, pigs and chickens. And these non-rational processes do not explain certain genuinely new moral notions, such as the immorality of torture as a form of punishment or the morality of democracy.

I propose that some people generate—though the interplay of reason and emotion and imagination—novel moral insights. This process is similar to what goes on when we generate other sorts of ideas, including philosophical and scientific ones. As a core example of this, Singer (1981) argues that the great insight about morality is the notion that it should be built from an objective position. Put crudely, the idea here is that nobody is special, which is an insight enshrined in the Golden Rule, the “impartial spectator” of Adam Smith, and the “original position” of John Rawls. We have the capacity to generate such ideas, and they really do matter,
shaping the societies in which we humans live.

This is a claim about where ideas originally come from, not how they are usually acquired. And so I agree with the Humean position that people are often swayed more by emotional appeals than by rational arguments. Indeed, it is likely that certain moral views, such as the wrongness of slavery, become commonplace through processes that are, for the most part, non-deliberative. Most of us were never persuaded that slavery is wrong; we acquired this belief in much the same way that we learned what language to speak or what clothes to wear. But still, as psychologists, we also need to explain how it was that someone once came to the moral insight that slavery is wrong, and, just as impressive, was able to convey this moral insight in such a way that it persuaded others. Explaining this process requires abandoning Hume.

In sum, the existence of moral progress shows that humans have the capacity for moral generativity, some capacity for coming to novel moral views. Psychologists often miss this, in large part because we rarely look for it. In moral psychology, as in other domains of psychology, there is a nearly exclusive focus on how people respond to various situations, and so there are countless studies in which various populations (mostly undergraduates, but also children, psychopaths, and the like) are exposed to artificial moral dilemmas. (This is surely true of my own research.) But it is unusual for us to study how people naturally arrive at their moral judgments, including those that might clash with those of the communities around them. We should do more of this, because at present the nature of this critically important process is almost entirely mysterious.
References


Hamlin, J.K., & Wynn, K. (under review). Five- and 9-month-old infants prefer prosocial to antisocial others.


