Journal of Moral Education, 2013
Vol. 42, No. 3, 281–297, http://dx.doi.org/10.1080/03057240.2013.817327

Moral psychology for the twenty-first century

Jonathan Haidt*

NYU-Stern School of Business, USA

Lawrence Kohlberg slayed the two dragons of twentieth-century psychology—behaviorism and psychoanalysis. His victory was a part of the larger cognitive revolution that shaped the world in which all of us study psychology and education today. But the cognitive revolution itself was modified by later waves of change, particularly an 'affective revolution' that began in the 1980s and an 'automaticity revolution' in the 1990s. In this essay I trace the history of moral psychology within the broader intellectual trends of psychology and I explain why I came to believe that moral psychology had to change with the times. I explain the origins of my own social intuitionist model and of moral foundations theory. I offer three principles that I think should characterize moral psychology in the twenty-first century: (1) Intuitions come first, strategic reasoning second, (2) There's more to morality than harm and fairness and (3) Morality binds and blinds.

Keywords: moral psychology; moral intuition; moral foundations theory

Moral psychology for the twenty-first century

I joined the field of moral psychology in 1987, when I arrived in West Philadelphia to begin graduate training at the University of Pennsylvania. Moral psychology at that time had an epic feel to it. It was a field whose history was defined by clashes of titans more than by specific studies or discoveries. A dramatized version of the history could be told like this: Durkheim versus Piaget, Freud versus Skinner, Kohlberg versus Freud and Skinner, Kohlberg versus Gilligan. Big issues were at stake. Is morality self-constructed or received from outside? If from outside, does it come from one's parents or from society at large? Do men and women have different developmental paths and different endpoints?

Jonathan Haidt, Stern School of Business, New York University. This essay is an edited version of the 2012 Kohlberg Memorial Lecture, delivered at the annual meeting of the Association for Moral Education. Correspondence concerning this article should be addressed to Jonathan Haidt, New York University, Stern School of Business, 40 W 4th St, New York NY, 10012, USA. Email: haidt@nyu.edu

Despite their many differences, all of the titans were united by one theoretical commitment: morality is not innate. Perhaps only Skinner would have maintained that the child was a blank slate and that any moral norm one could imagine would be as easily teachable as any other. But all of the titans would have firmly rejected the biological and evolutionary approach posited by E.O. Wilson (1975) in his great work, *Sociobiology*. Wilson's most famous line for moral psychology (and infamous, for moral philosophy) was this: 'Scientists and humanists should consider together the possibility that the time has come for ethics to be removed temporarily from the hands of the philosophers and biologicized' (p. 562).

On the next page, Wilson (1975) described what it meant to 'biologicize' morality, and he made clear that he included cognitive developmentalists such as Kohlberg as allies of the philosophers:

> ethical philosophers intuit the deontological canons of morality by consulting the emotive centers of their own hypothalamic-limbic system. This is also true of the developmentalists, even when they are being their most severely objective. Only by interpreting the activity of the emotive centers as a biological adaptation can the meaning of the canons be deciphered. (p. 563)

Wilson was urging researchers to stop reading Rawls and Kant for a while, stop talking about the lofty peaks of moral development, and search instead for the foundations of morality. Those foundations were likely to be found, he said, in a small set of affectively laden responses produced by the 'emotive centers' of the brain, which themselves should be studied as biological adaptations crafted by natural selection working on intensely social species such as ourselves and other primates.

But Wilson's charge came 20 years too early. In 1975, few researchers in moral psychology were interested in biologicizing anything. By the time I began to study moral psychology in 1987, I got the feeling that sociobiology was radioactive. It was dismissed as reductionist and it was tainted as a gateway theory leading to racism and sexism. After all, if evolution shaped human minds (not just bodies), then race and gender differences could more plausibly be innate. But that conclusion was morally unacceptable to most psychologists and, therefore, so was sociobiology (see the angry response to Wilson by leftist intellectuals in Allen and Beckwith [1975]). Moral psychology in 1987 seemed untouched by evolutionary thinking. The academic field it linked to most frequently was moral philosophy.

Moral psychology in 1987 was a branch of developmental psychology. There was research on topics related to morality in social psychology, such as altruism or empathy, but if you said that you studied moral psychology in the 1980s, then you probably studied the development of moral reasoning. You didn't need to agree with Kohlberg on any particular claim, but you lived and worked on land that Kohlberg had cleared. In defeating the twin dragons of behaviorism and psychoanalysis, Kohlberg (1969) had won a major battle in the cognitive revolution. He had made it permissible for researchers to look inside the 'black box' of the mind and study moral reasoning. He made it possible to assume that such

reasoning was not an epiphenomenon (as a Skinnerian might say), nor was it merely rationalization (as a Freudian might say). Rather, reasoning was where the action was. If you understood how a child reasoned about a moral dilemma, and you could track how that reasoning changed over time, then you could infer the underlying mental structures and their developmental course. I have referred to this general approach as *rationalism*, defined as the belief that 'reasoning is the most important and reliable way to obtain moral knowledge' (Haidt, 2012, p. 7).

But all that began to change while I was in graduate school. There were at least six intellectual trends—six waves that came from different directions, but that washed ashore within a decade and altered the landscape. By the late-1990s, rationalism no longer seemed so plausible.

The first trend was the *affective revolution* that began in the early-1980s (see Fischer and Tangney [1995] for a review). It was a correction or addition to the cognitive revolution. The affective revolution opened, in my view, when Robjert Zajonc (1980) revived Wilhelm Wundt's doctrine of affective primacy (but see also the ongoing work on empathy, such as Hoffman, 1982). Zajonc argued for a dual process model in which 'affect' (or feeling, or hot cognition) has primacy over 'thinking' (or reasoning, or cold cognition). It has primacy in four ways: it comes first evolutionarily, it comes first developmentally, it comes first (within milliseconds) in any real-time judgment and it is more powerful than 'thinking', which depends on language. Affective reactions structure and constrict the mental space within which subsequent thinking occurs. In 1984, Jerome Kagan called for greater attention to affect in the study of moral development: 'beneath the extraordinary variety of surface behaviors and consciously articulated ideals, there is a set of emotional states that form the bases for a limited number of universal moral categories that transcend time and locality' (p. 118). Zajonc and Kagan were both, in effect, supporting Wilson's call for greater attention to quick, affectively laden processes.

The second trend was the rebirth of *cultural psychology* in the mid-1980s, led by the anthropologist Richard Shweder (1990; Shweder & Levine, 1984). One of Shweder's main research topics was morality and in 1987 he published an essay that seemed to me, at the time, to be yet another round of the clash of the titans. In 1987, Eliot Turiel and his students were the most productive group of researchers generating empirical research on the development of moral judgment. They developed an interview technique that was much easier to administer than Kohlberg's and they used it to show that children as young as 4 or 5 years old distinguished between violations of social conventions (which could be changed by authorities) and violations of moral rules, such as the prohibition on violence (hitting is still wrong, even if the teacher says that kids can hit each other [see Nucci & Turiel, 1978]).

In an important edited volume (Kagan & Lamb, 1987), Shweder, Mahapatra, and Miller (1987) presented evidence that Turiel's moral-conventional distinction was not found robustly in India. The early emergence of this distinction was crucial for Turiel's (1983) domain theory, in which children actively construct their

own understanding of morality as they encounter a variety of social rules and regularities and come to realize that rules that prevent harm are different from other kinds of rules. Shweder proposed instead a 'social communication' theory of moral development, in which adults and other 'local guardians of the moral order' assist children in learning morality, in large part by expressing emotions in response to violations of rules.

The third trend was the *automaticity revolution*, led by the social psychologist John Bargh (1994), who proposed that mental processes can be placed along a continuum from fully automatic to fully controlled. Automatic processes are what animal minds have been doing for 500 million years: they are fast, effortless and efficient responses to patterns of input from perceptual systems. Controlled processes, in contrast, only became possible in the last million years, when human beings acquired language. They are slower and require some mental effort. Bargh's research demonstrated that a great deal of human behavior——including higher cognition such as decision making and evaluative judgment——was handled primarily or entirely by automatic processes. Bargh also stated his findings in terms consonant with Wilson's charge:

> it may be, especially for evaluations and judgments of novel people and objects, that what we think we are doing while consciously deliberating in actuality has no effect on the outcome of the judgment, as it has already been made through relatively immediate, automatic means. (Bargh & Chartrand, 1999, p. 475)

The fourth trend was research in *neuroscience*, made possible by advances in imaging technology in the 1980s and 1990s. In 1994, Antonio Damasio published *Descartes' error: Emotion, reason, and the human brain*. Damasio found that patients who sustained damage to the ventromedial pre-frontal cortex lost the ability to integrate affect into their decision making. They retained their knowledge of right and wrong, but were deprived of the affective flashes that everyone else uses to structure, constrain and guide reasoning. When emotion is removed, the result is not hyper-rational behavior, it is a disastrous inability to narrow down the choices and then choose among them. As Damasio (1994) put it: 'Nature appears to have built the apparatus of rationality not just on top of the apparatus of biological regulation, but also from it and with it' (p. 128). Once again, Wilson's charge was supported.

The fifth trend was *primatology*, particularly the work of Frans de Waal (1996). De Waal rejected the prevailing fear of anthropomorphizing apes. He saw us as close cousins, evolutionarily, and he saw evidence that most or all of the 'building blocks' of human morality were present in other apes as well. These building blocks were largely social-emotional capacities, such as the emotions that bind individuals into friendships and reciprocal relationships (see also Trivers, 1971) and that make individuals care about others and give succorance to those who are suffering, even when those others are not kin.

The sixth trend was the rebirth of sociobiology, relabeled as *evolutionary psychology* (Barkow, Cosmides, & Tooby, 1992). Social psychologists had never been

committed to a blank slate view of the mind and they gradually embraced evolutionary psychology as part of their toolkit for understanding how social environments shape cognition and behavior. Moral emotions came to be studied much the way that Wilson had urged——as products of natural selection that conferred advantages on individuals trying to navigate complex social terrains (see, for example, Keltner & Buswell, [1997] on embarrassment and Rozin, Haidt & McCauley, [1993] on moral disgust).

All six of these waves washed ashore while I was in graduate school, or soon afterwards. My dissertation research had been inspired directly by Shweder's challenge to Turiel. I had what I can only describe as a strong intuition that Shweder was right, but yet I agreed with Turiel's rebuttal essay (Turiel, Hildebrandt, & Wainryb, 1991), in which he charged that Shweder et al. (1987) had used a method with a variety of flaws that made it hard to interpret their results. Foremost among these flaws was a failure to determine whether Shweder's Indian participants perceived harm in the vignettes——harm that Americans did not perceive——because Indians and Americans have different 'informational assumptions' about how the world works.

For example, when Hindus say that it is universally wrong for widows to eat fish, their judgments are uninterpretable by non-Hindus until we know that Hindus believe that souls reincarnate after death and can follow the affairs of the living. We must also know that fish is a 'hot' food that might stimulate a widow's sexual appetites and cause her to behave in ways that bring pain to the soul of her deceased husband. So Shweder's finding that Indians moralized many practices that Americans judged to be social conventions did not necessarily show that Hindu Indians lacked the moral-conventional distinction.

I therefore set out to replicate Shweder's study using a new set of vignettes I had created to be emotionally disturbing yet completely harmless. For example, a family eats its pet dog, after the dog had been killed by a car in front of their house. A woman can't find a rag with which to clean her toilet, so she cuts up her old American (or Brazilian) flag and uses the scraps to clean the toilet.

I asked adult and adolescent participants, of high and low social class, in Philadelphia, and in two Brazilian cities (Recife and Porto Alegre). I included a question about the presence of harm so that I could isolate those participants who specifically stated that there was no harm done in each story. And I found that Shweder's claims replicated cleanly. Turiel's domain theory worked perfectly in my upper-class Philadelphia samples, where adults and children made a neat distinction between moral and conventional stories, and where actions that were seen to be harmless were usually judged to be matters of social convention. But as you move away from that group, either down in social class or toward more 'third-world' cultural and economic conditions, domain theory began to break down and cultural variations in the moral domain began to grow.

I didn't have the term WEIRD back then (Western, Educated, Industrialized, Rich and Democratic——see Henrich, Heine and Norenzayan [2010]), but my dissertation was a clear demonstration that people who grow up in WEIRD cultures

should not be taken as representative of human psychology. The last line of my dissertation has stood the test of time: 'Psychological processes such as moral judgment may be variable across social classes and national borders. An adequate model of moral judgment must capture this variation' (Haidt, Koller, & Dias, 1993, p. 627). The rest of my career has been an effort to specify that model.

The righteous mind (Haidt, 2012) is the culmination of that effort. In that book, I propose three basic principles of moral psychology that work together to explain what morality is, where it comes from, why it varies across cultures and how it develops in children. I'll structure the remainder of this essay using those three principles. I didn't invent any of them, but I'll explain how and why I came to believe them and rely upon them. I'll explain why I believe they offer guidance for the future of moral psychology in the twenty-first century.

Principle 1: Intuition comes first, strategic reasoning second

In 1995, I got my first faculty position, at the University of Virginia, in the social psychology program. I continued to study the role of emotional responses in moral judgment and I continued to read the works that constituted the six waves of new thinking that I described above. My empirical productivity was low in those early years, in part because I had what I now believe is the wrong model of moral judgment. From my dissertation research I knew that emotions were powerful inputs to moral reasoning and so I created what I thought was the most logical model: a dual-process model in which emotion and reasoning are each separate inputs to moral judgment, which can push against each other, as shown in Figure 1.

I then did what any social psychologist would do: I tried to strengthen and weaken both inputs and show that I could push around the final judgment. But while it was very easy to alter judgments by strengthening or weakening emotional inputs, I was not able to alter judgments by strengthening the reasoning process (e.g., giving people more time to think) or by weakening it (e.g., by asking people to hold a seven digit number in their working memory).

Given that my empirical publication rate was low, my tenure committee warned me that I was in danger and suggested that I go with what they perceived to be my strengths: scholarship and theory. Since I have always loved reading books more than running experiments, I agreed, and began to devote most of my time to writing a review paper that I eventually published with the title 'The emotional dog and its rational tail' (Haidt, 2001).

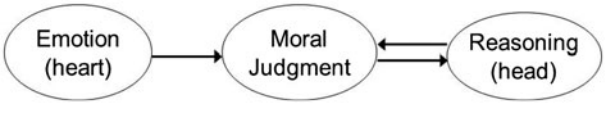
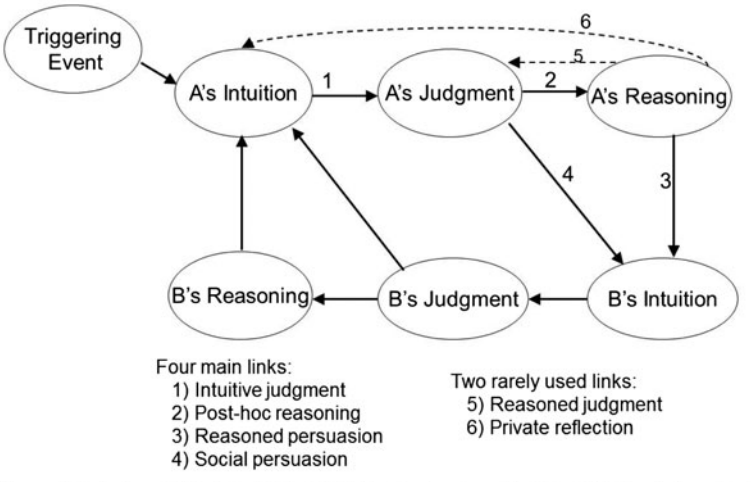


Figure 1. A dual process model of moral judgment in which emotion and reasoning are equal inputs, although moral judgment can also produce post-hoc reasoning.

In that paper, I presented the Social Intuitionist Model (SIM), which began with the basic insight of David Hume (1739–1740/1969) that 'reason is and ought only to be the slave of the passions'. Moral reasoning is something we engage in after an automatic process (passion, emotion or, more generally, intuition) has already pointed us toward a judgment or conclusion. We engage in moral reasoning not to figure out what is really true, but to prepare for social interactions in which we might be called upon to justify our judgments to others. It is a distinctive fact about moral judgments that we can't simply point to our own preferences as sufficient justification. I can say 'I prefer chocolate to vanilla because … I just don't like vanilla.' But I cannot say 'I think we should punish Bob because … I just don't like him.' We must offer reasons to others for why they should join us in condemning Bob, and those reasons must refer to normative claims, not to anyone's preferences or self-interest. The SIM took Hume's basic insight and put it into the context of a conversation between two or more people, each of whom engages in moral reasoning to try to sway the other party or, at least, justify his or her own judgment (see Figure 2).

When I first published the model, there was plenty of evidence for it in social psychology in general (e.g., research on motivated reasoning [Kunda, 1990]), although there was little direct evidence for it from within moral psychology. But once I dropped the dual process model of Figure 1, in which reason and 'emotion' are equally powerful, and I embraced the Humean dual process model, in which 'intuition' (a broader term than emotion) comes first and reasoning comes second, my experiments began to work much better. In my own experimental work, I conducted several studies that manipulated emotional reactions, for example, by inducing a flash of hypnotic disgust (Wheatley & Haidt, 2005) or by having people



Source: Originally published in Haidt (2001) by the American Psychological Association. Reprinted with permission.

Figure 2. The social intuitionist model.

fill out moral judgment questionnaires at a desk that was dirty and disgusting (Schnall, Haidt, Clore, & Jordan, 2008). But you don't need to induce emotions to change judgments——just activating concepts of cleanliness, such as by washing one's hands (Zhong, Strejcek, & Sivanathan, 2010) or merely standing next to hand sanitizer (Helzer & Pizarro, 2011) have been shown to alter moral judgments.

Since I published the SIM, many critics have focused on the 'intutionist' part (Figure 2, links 1, 2, and 3), sometimes chopping off the social part entirely. For example, Saltzstein and Kasachkoff (2004) argued that the SIM 'appears to relegate [moral reasoning] to the status of an epiphenomenon' and that 'The one avenue of change that [Haidt] does not seem to recognize is that brought about by reasoned argument.' But if you look at the entire SIM in Figure 2, you can see why these charges are false. As I wrote in the original 'Emotional dog' article:

> The discussion thus far may have given the impression that the model dismisses reasoning as post-hoc rationalization (link 2). However it must be stressed that *four of the six links in the model are reasoning links, and three of these links (3, 5, and 6) are hypothesized to have real causal effects on moral judgment*. Link 3, the reasoned persuasion link, says that people's (ex post facto) moral reasoning *can have a causal effect … on other people's intuitions*. In the social intuitionist view moral judgment is not just a single act that occurs in a single person's mind. It is an ongoing process, often spread out over time and over multiple people. Reasons and arguments can circulate and affect people, even if individuals rarely engage in private moral reasoning for themselves. (Haidt, 2001, p. 828, emphasis added)

My claim is that each of us is flawed as an individual reasoner. We are each cursed by the 'confirmation bias' (the tendency to seek only evidence that will confirm our pre-existing beliefs [Nickerson, 1998]) and nobody has yet figured out a way to 'debias' people (Lilienfeld, Ammirati, & Landfield, 2009). Reasoning evolved to serve an 'argumentative function' (Mercier & Sperber, 2011). Yet, when you put flawed reasoners together into a social group, such as a jury, an advisory board or a legislature in which there are good social relationships among the members (unlike the US Congress), good reasoning can emerge, because each person's confirmation bias is challenged by others. Reasoning matters a great deal for moral change and progress. But it is a mistake for moral psychologists to focus on the reasoning of lone individuals reacting to a moral judgment vignette and to conclude that reasoning is where the action is.

Wilson was right. The action in moral psychology is in the 'emotive centers' (or affectively laden intuitions more broadly) and the philosophical arguments are to some extent generated post-hoc as a way of making sense of those feelings. Greene (2008) came to exactly this conclusion in his review of the implications of neuroscience research for moral psychology:

> We have strong feelings that tell us in clear and certain terms that some things simply cannot be done and that other things simply must be done. But it's not obvious how to make sense of these feelings, and so we, with the help of some especially creative philosophers, make up a rationally appealing story [about rights]. (p. 63)

Moral psychology in the twenty-first century should continue to study moral reasoning, but with the understanding that intuitions come first, strategic reasoning second.

Principle 2: There's more to morality than harm and fairness

After graduate school, I spent two years working with Richard Shweder at the University of Chicago as a post-doctoral researcher. During that time I went to Bhubaneswar, India (where Shweder did his research) to conduct interviews of my own on Indian notions of purity and pollution. I came to see that these Indian practices made little sense as practical methods for avoiding disease. Cow dung, for example, is regarded as pure and holy because it comes from the revered cow. It has a purely symbolic value. Dung is useless, at best, as a way to clean surfaces upon which rituals will be conducted. Rather, traditional Hindu practices had a loose fit with the basic psychology of disgust that I had studied at Penn with Paul Rozin (Rozin et al., 1993). Indian practices of cleansing and purification also reflected culturally-specific ideas about holiness, virtue and the proper procedure for preparing oneself for communion with God, despite the fact that each one of us has a biological nature that seems (to people in many cultures) to be incompatible with our status as children of God, or possessors of souls. My time in India profoundly altered my thinking about morality. It got me out of my own secular liberal WEIRD moral framework and prepared me for understanding moral frameworks back in the USA that were religious and conservative (see Haidt, 2012, chapter 5).

In the 'Emotional dog' article, I made the case that moral intuitions were where the action was, but I was not very specific about what those intuitions were. If I wanted to create a comprehensive theory of morality, I knew I would have to offer a list or taxonomy of mental systems, and I had to describe the adaptive pressures that shaped each one. In a short essay published in 2004, I offered my first attempt at such a list: I noted that three of the chapters of de Waal (1996) matched up closely with three of Alan Fiske's (1991) four relational models. De Waal's chapter on sympathy reveals part of the evolutionary origins of Fiske's 'communal sharing', de Waal's chapter on rank and order show us the origins of Fiske's 'authority ranking' and de Waal's chapter on 'quid pro quo' show us the origins of Fiske's 'equality matching'. De Waal and Fiske fit together better than the coasts of Brazil and West Africa. This was just the sort of consilience that Wilson (1998) had called for.

In an essay written slightly later, but also published in 2004, I tried to go beyond those three most obvious matches and look more systematically for bridges between evolutionary theory and primatology, on one side, and anthropology, on the other. Haidt and Joseph (2004) reviewed five sources that offered broad coverage of the moral domain. We reviewed two sources that aimed to describe universals (Brown, 1991; Fiske, 1991), two that aimed to describe cultural variation (Schwartz & Bilsky, 1990; Shweder, Much, Mahapatra, & Park, 1997) and de

Waal's (1996) survey of the building blocks of morality in other primates. We listed the major kinds of social situations that each author had said that people (or chimpanzees) react to with clear evaluations of positive or negative. We then tallied up the 'votes' and found that the three areas of match between Fiske and de Waal were also found in the other three sources, confirming that they were great candidates for being innate moral foundations. We also found support (from most but not all of the five sources) for positing a foundation related to purity, sanctity and the emotion of disgust, as well as for a foundation related to group boundaries and group loyalty. More recently we've suggested that Liberty/oppression is a good candidate as well (Iyer, Koleva, Graham, Ditto, & Haidt, 2012) (see Haidt 2012, chapters 7 and 8, for a comprehensive review).

The original list of foundations resulted from a somewhat ad hoc process in which Craig Joseph and I exercised some discretion and judgment as we debated how to lump together or split apart the many candidate items on our list. Several scholars have rightfully criticized us for failing to offer more objective criteria for 'foundationhood' (e.g., Suhler & Churchland, 2011). Recently, my colleagues at YourMorals.org and I have specified five such criteria:

(1) It is a common concern in third-party judgments.
(2) It provokes automatic affective evaluations.
(3) It is culturally widespread.
(4) There is evidence of innate preparedness (e.g., the response is evident in other primates or it emerges very early in infants).
(5) There is an existing evolutionary model demonstrating its adaptive advantage.

For elaboration on each of these criteria, see Graham et al. (in press; available at MoralFoundations.org). The most important point to explain here is that calling something 'innate' does not mean that it is present at birth, hardwired or universally expressed in all people or all cultures. Innate means 'structured in advance of experience' (Marcus, 2004), but experience can suppress, alter or magnify the importance of a foundation. We call them foundations to distinguish them from the finished building, the finished adult morality. Foundations are the universal psychological preparednesses (Seligman, 1971) that make it easy for children to learn some moral ideas (e.g., if someone hits you, hit him back), but hard to learn others (e.g., if someone hits you, turn the other cheek with love in your heart).

Moral Foundations Theory was originally designed for cross-cultural and cross-national work, not for political psychology. But it turns out that the moral communities of the US left and right are like separate nations in some ways. Their moral 'matrices' are built on differing configurations of the foundations and these differences go a long way toward explaining the moral anger over culture war issues. All US groups value care, fairness and liberty, although left and right often use those words in different ways (e.g., on YourMorals.org, liberals give higher

ratings to items about equality, whereas conservatives give higher ratings to items about proportionality, particularly in punishment). But many of the most enduring culture-war issues——those related to family structure, gender, bio-medical procedures (such as abortion and euthanasia) and patriotism——result from a general disagreement about the very legitimacy of the loyalty, authority and sanctity foundations.

For example, Koleva et al. (2012) examined the responses from 24,000 Americans on 20 culture war issues and tried to predict subjects moral positions based on their responses on the Moral Foundations Questionnaire (Graham et al., 2011), along with a battery of demographic questions. It comes as no surprise that ideology (self-placement on the liberal-conservative dimension) was a strong predictor of judgments. For example, conservatives were more critical than liberals of same-sex marriage, abortion, open immigration and people who burn the American flag as a form of protest. Yet, even after you use ideology plus all other demographic predictors (such as age, gender and religiosity), specific moral foundations predicted substantial amounts of additional variation. For example, scores on the Care/harm foundation helped to predict opposition to testing drugs on animals, over and above all other predictors.

The biggest surprise in the data was that the Sanctity/degradation foundation emerged as the second most powerful predictor——just behind ideology. For example, which foundation do you think would best predict opposition to flag burning? Loyalty/betrayal is the obvious choice, given that the flag is the symbol of the nation. Yet in fact, Sanctity scores were just as predictive in one sub-study and slightly less predictive in another sub-study. Some people see the flag as just a piece of cloth (low sanctity), others have a general tendency to see objects as containing invisible yet morally significant properties (high sanctity), much like the priests and monks I had interviewed in India. And for all culture war issues involving sexuality or gender (such as same-sex marriage, having casual sex, having a baby outside of marriage) as well as for controversial biomedical procedures (such as cloning and stem cell research), the sanctity foundation was the best predictor among the moral foundation scores and it often exceeded ideology in its predictive power. Once again, some people see the world as being made exclusively of matter (low sanctity), others see it as full of non-material essences and properties (high sanctity). Social conservatives score higher on sanctity than do social liberals, but even within any political community, some individuals will show a more 'Indian' way of thinking (as can be seen among liberals in communities that practice yoga, eat natural foods, and demonize hydraulic fracking, even when faced with evidence that it has allowed the USA to drastically reduce its carbon dioxide output by reducing the use of coal. Sanctity is about symbolism, not efficiency. Liberals can do sanctity too).

Nearly all moral psychologists are politically liberal——I know of none who self-identify as conservative. So the moral worldview of US conservatives can seem at times quite alien and undeserving of respect. It is often explained away as a product of subconscious motives and fears (Jost, Glaser, Kruglanski, & Sulloway,

2003). But if moral psychology is to make progress in the twenty-first century, it will have to overcome its own moral homogeneity. It will have to conduct a great deal of cross-cultural research, which does not necessarily require crossing any national borders. It should commit to the principle that——descriptively speaking——there's more to morality than harm and fairness.

Principle 3: Morality binds and blinds

In college I read Richard Dawkins' (1976) book, *The selfish gene*, and the scales fell from my eyes. It was like seeing, for the first time, why human nature is the way it is. In graduate school, when I began to study morality, I took with me Dawkins' conviction that two simple processes are sufficient to explain the origin of morality: kin selection (Hamilton, 1964) and reciprocal altruism (Trivers, 1971). It was so simple, so beautiful, so … parsimonious.

For 25 years after *The selfish gene*, just about all discussions of the evolution of morality relied upon those two processes, plus a few variants (such as indirect reciprocity). I accepted Dawkins' claims about group selection too——that it was possible in theory but unworkable in practice, because, in practice, altruists who sacrifice for the good of the group will be replaced by free-riders who make no such sacrifices.

I accepted this standard formulation throughout the 1990s, even as my cross-cultural work was leading me to see an expanded moral domain in which the integrity and coherence of groups mattered a great deal. Many religious, traditional and conservative moralities seemed to be aimed at the goal of protecting groups, not individuals. It wasn't clear how to explain such groupish moralities in Dawkins' terms, other than by dismissing these moral systems as sets of parasitic memes, which is to this day Dawkins' (2006) preferred explanation for the origin and prevalence of religion.

But then I read David Sloan Wilson's (2002) book, *Darwin's cathedral*, and a second set of scales fell from my eyes. Wilson argued that human nature is a product of multi-level selection. He argued that Darwin had it exactly right. People compete within groups, while at the same time, groups compete with groups. As Darwin (1871/1998) put it:

> When two tribes of primeval man, living in the same country, came into competition, if (other circumstances being equal) the one tribe included a great number of courageous, sympathetic and faithful members, who were always ready to warn each other of danger, to aid and defend each other, *this tribe would succeed better and conquer the other* …. The advantage which disciplined soldiers have over undisciplined hordes follows chiefly from the confidence which each man feels in his comrades … *Selfish and contentious people will not cohere, and without coherence nothing can be effected. A tribe rich in the above qualities would spread and be victorious over other tribes*. (p. 134, emphasis added)

Groups that achieve coherence vanquish groups that do not. We are all descended from groups that achieved some degree of coherence and Wilson argued that religions co-evolved (as cultural products) along with religious

minds (as biological products) over the course of the tens or hundreds of millennia in which human groups have been competing with each other over scarce resources, such as territory. Wilson showed, in a series of case studies, that religions do, in fact, solve the free rider problem that Dawkins (and Williams, 1966) thought was fatal for group selection accounts. Shared gods, rituals, myths and values really to help people to trust each other and achieve together feats that they could not do on their own (Norenzayan & Shariff, 2008; Shariff, Norenzayan, & Henrich, 2009). Conservative and traditional moralities seem generally well suited to preserving moral communities of this sort, even if Turiel would dismiss rules such as 'though shalt have no other gods before me' as mere social conventions.

Once I began taking the psychology of religion seriously, I began to see it in operation everywhere. Sports teams and college fraternities have rites and rituals and make heavy use of synchronous movements (also known as 'muscular bonding') to create trust and cohesion, allowing them to reap benefits of cooperation. Political movements elevate leaders——especially martyred leaders——to a godlike status, and then circle around them, brooking no dissent, showering wrath on traitors and apostates. I even see it happening in the academy, where the disappearance of conservatives (Duarte et al., 2013; Rothman, Lichter, & Nevitte, 2005) has allowed many fields to take on a distinctively left-leaning political orientation. Once a group is bound together by shared values (such as tolerance and diversity) and shared demons (such as racism), people in the group find it hard to perceive or accept empirical findings that contradict their shared moral 'matrix'.

It's easy for academics to see this happening on the religious right. If we heard that the biology department at Liberty University (an evangelical school in Virginia) forbade any mention of evolution, we'd all have a good laugh. How can you do biology while banning the most powerful explanatory theory in biology? Yet much the same thing happens in social psychology. One of the largest causes of stereotypes and racism is the fact that most stereotypes correspond to a substantial degree with observable reality (Jussim, 2012). People are intuitive Bayesians——they cannot help but detect statistical differences in behavior between social groups. Yet research on stereotype accuracy is rarely cited in social psychology textbooks, it is rarely mentioned by academics as a cause of discrimination in the real world and it is actively tabooed in academic discussions. How can you do social psychology while banning one of the most powerful explanations of social behavior?

Morality binds and blinds. This process may be helpful for groups that require a great deal of cohesion, such as military units, but it is devastating for communities whose purpose is the pursuit of truth. Moral psychology in the twenty-first century should do what it can to increase its moral and political diversity and, thereby, decrease its internal cohesion and its vulnerability to political correctness. (For more on ideological diversity, and how to increase it, see Duarte et al., 2013.)

Conclusion

Moral psychology has changed radically in the last 25 years. In this essay I offered three principles for characterizing three major new trends in the field:

(1) Intuitions come first, strategic reasoning second.
(2) There's more to morality than harm and fairness.
(3) Morality binds and blinds.

In other words, compared to the state of the art 25 years ago, morality is now seen to be more intuitive, the moral domain is seen to be broader and the function of morality is seen to be more social, less about finding truth.

Another way to describe the change is to say that E. O. Wilson's prophesy has come true. Moral psychology has been taken out of the hands of philosophers and biologicized. The center of gravity is now the study of automatic, intuitive and affective processes, which, as Wilson suggested, can be understood as a diverse set of evolutionary adaptations. Moral reasoning plays many important roles, but those roles are now seen as being more interpersonal (done to prepare for social interaction), rather than intrapersonal (done to find the truth or work out one's intrapsychic conflicts).

But biologicizing morality has not meant the triumph of reductionism. Morality is not 'just' a biological phenomenon, to be studied primarily with functional magnetic resonance imaging techniques. Rather, the new emphasis on automatic and intuitive processes has made the study of morality into one of the most active convergence zones in the academy. Like sex, language or eating, morality is an aspect of humanity that is both deeply biological and profoundly cultural. It can't possibly be studied by one discipline alone. At scientific conferences on morality nowadays, one finds not just neuroscientists and evolutionists but also social and developmental psychologists, primatologists, economists, philosophers and historians, all of whom know and cite each other's work. Morality cries out to be studied not just by reductionists, but by emergentists, too, eager to understand the origins of moral order, social structure and cultural variation.

Moral psychology in the twenty-first century is in a golden age——just in time for a century sure to be full of global threats, commons dilemmas, cultural clashes and injustice.

References

Allen, E., & Beckwith, B. (1975). Against 'sociobiology'. *New York Review of Books, 22*, 43–44.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, efficiency, intention and control in social cognition. In J. R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*, 462–479.

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford: New York.

Brown, D. E. (1991). *Human universals*. Philadelphia, PA: Temple University Press.

Damasio, A. (1994). *Descartes' error: Emotion, reason and the human brain*. New York: Putnam.

Darwin, C. (1998/1871). *The descent of man and selection in relation to sex*. Amherst, NY: Prometheus Books.

Duarte, J. L., Crawford, J. T., Stern, S., Haidt, J., Jussim, L., & Tetlock, P. E. (2013). *Ideological diversity will improve psychological science.* Unpublished manuscript, Arizona State University.

Dawkins, R. (1976). *The selfish gene*. London: Oxford University Press.

Dawkins, R. (2006). *The God delusion*. Boston, MA: Houghton Mifflin.

de Waal, F. B. M. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.

Fischer, K. W., & Tangney, J. P. (1995). Self-conscious emotions and the affect revolution: Framework and overview. In J.P. Tangney & K.W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment and pride* (pp. 3–24). New York, NY: Guilford.

Fiske, A. P. (1991). *Structures of social life*. New York: Free Press.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. H. (in press). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the Moral Domain. *Journal of Personality and Social Psychology, 101*, 366–385.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology*The neuroscience of morality (3, pp. 35–79). Cambridge, MA: MIT Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York: Pantheon.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, Fall*, 55–66.

Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*, 613–628.

Hamilton, W. D. (1964). The genetical evolution of social behavior, parts 1 and 2. *Journal of Theoretical Biology, 7*, 1–52.

Helzer, E. G., & Pizarro, D. A. (2011). Dirty Liberals! Reminders of physical cleanliness influence moral and political attitudes. *Psychological Science, 22*, 517–522.

Henrich, J., Heine, S., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–83.

Hoffman, M. L. (1982). Affect and moral development. In D. Ciccetti & P. Hesse (Eds.), *New directions for child development*Emotional development (16, pp. 83–103). San Francisco, CA: Jossey-Bass.

Hume, D. (1969/1739–1740). *A treatise of human nature*. London: Penguin.

Iyer, R., Koleva, S. P., Graham, J., Ditto, P. H., & Haidt, J. (2012). Understanding Libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS ONE, 7*(8), e42366.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservativism as motivated social cognition. *Psychological Bulletin, 129*, 339–375.

Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. Oxford: New York.

Kagan, J. (1984). *The nature of the child*. New York: Basic Books.

Kagan, J., & Lamb, S. (Eds.). (1987). *Culture and moral development*. Chicago, IL: University of Chicago Press.

Keltner, D., & Buswell, B. N. (1997). Embarrassment: Its distinct form and appeasement functions. *Psychological Bulletin, 122*, 250–270.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago, IL: Rand McNally.

Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the threads: How five moral concerns (especially Purity) help explain culture war attitudes. *Journal of Research in Personality, 46*, 184–194.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480–498.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science, 4*, 390–398.

Marcus, G. (2004). *The birth of the mind*. New York: Basic Books.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory *Behavioral and Brain Sciences, 34*, 57–74.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.

Norenzayan, A., & Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science, 322*, 58–62.

Nucci, L., & Turiel, E. (1978). Social interactions and the development of social concepts in pre-school children. *Child Development, 49*, 400–407.

Rothman, S., Lichter, S. R., & Nevitte, N. (2005). Politics and professional advancement among college faculty. *The Forum, 3*(1). doi: 10.2202/1540-8884.1067.

Rozin, P., Haidt, J., & McCauley, C. (1993). Disgust. In M. Lewis & J. Haviland (Eds.), *Handbook of emotions* (pp. 575–594). New York: Guilford Press.

Saltzstein, H. D., & Kasachkoff, T. (2004). Haidt's moral intuitionist theory. *Review of General Psychology, 8*, 273–282.

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin, 34*, 1096–1109.

Schwartz, S. H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology, 58*, 878–891.

Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy, 2*, 307–320.

Shariff, A. F., Norenzayan, A., & Henrich, J. (2009). The birth of high gods: How the cultural evolution of supernatural policing influenced the emergence of complex, cooperative human societies, paving the way for civilization. In M. Schaller, A. Norenzayan, S. Heine, T. Yamagishi, & T. Kameda (Eds.), *Evolution, culture and the human mind* (pp. 117–136). New York: Taylor and Francis.

Shweder, R. A. (1990). Cultural psychology: What is it? In J. W. Stigler, R. A. Shweder, & G. Herdt (Eds.), *Cultural psychology: Essays on comparative human development* (pp. 1–43). New York: Cambridge University Press.

Shweder, R. A., & LeVine, R. A. (Eds.). (1984). *Culture theory: Essays on mind, self, emotion*. Cambridge: Cambridge University Press.

Shweder, R. A., Mahapatra, M., & Miller, J. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1–83). Chicago, IL: University of Chicago Press.

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The 'big three' of morality (autonomy, community and divinity), and the 'big three' explanations of suffering. In A. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119–169). New York: Routledge.

Suhler, C. L., & Churchland, P. S. (2011). Can innate, modular 'foundations' explain morality? Challenges for Haidts moral foundations theory *Journal of Cognitive Neuroscience, 23*, 2103–2116.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35–57.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.

Turiel, E., Hildebrandt, C., & Wainryb, C. (1991). Judging social issues: Difficulties, inconsistencies, and consistencies. *Monographs of the Society for Research in Child Development, 56*, 1–103.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science, 16*, 780–784.

Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton, NJ: Princeton University Press.

Wilson, D. S. (2002). *Darwin's cathedral: Evolution, religion and the nature of society*. Chicago, IL: University of Chicago Press.

Wilson, E. O. (1975). *Sociobiology*. Cambridge, MA: Harvard University Press.

Wilson, E. O. (1998). *Consilience: The unity of knowledge*. New York: Alfred A. Knopf.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35*, 151–175.

Zhong, C. B., Strejcek, B., & Sivanathan, N. (2010). A clean self can render harsh moral judgment. *Journal of Experimental Social Psychology, 46*, 859–862.