The New Synthesis in Moral Psychology

Jonathan Haidt

People are selfish, yet morally motivated. Morality is universal, yet culturally variable. Such apparent contradictions are dissolving as research from many disciplines converges on a few shared principles, including the importance of moral intuitions, the socially functional (rather than truth-seeking) nature of moral thinking, and the coevolution of moral minds with cultural practices and institutions that create diverse moral communities. I propose a fourth principle to guide future research: Morality is about more than harm and fairness. More research is needed on the collective and religious parts of the moral domain, such as loyalty, authority, and spiritual purity.

If you ever become a contestant on an unusually erudite quiz show, and you are asked to explain human behavior in two seconds or less, you might want to say “self-interest.” After all, economic models that assume only a motive for self-interest perform reasonably well. However, if you have time to give a more nuanced answer, you should also discuss the moral motives addressed in Table 1. Try answering those questions now. If your total for column B is higher than your total for column A, then congratulations, you are Homo moralis, not Homo economicus. You have social motivations beyond direct self-interest, and the latest research in moral psychology can help explain why.

In 1975, E. O. Wilson (1) predicted that ethics would soon be incorporated into the “new synthesis” of sociobiology. Two psychological theories of his day were ethical behaviorism (values are learned by reinforcement) and the cognitive-developmental theory of Lawrence Kohlberg (social experiences help children construct an increasingly adequate understanding of justice). Wilson believed that these two theories would soon merge with research on the hypothalamic-limbic system, which he thought supported the moral emotions, to provide a comprehensive account of the origins and mechanisms of morality.

As it turned out, Wilson got the ingredients wrong. Ethical behaviorism faded with behaviorism. Kohlberg’s approach did grow to dominate moral psychology for the next 15 years, but because Kohlberg focused on conscious verbal reasoning, Kohlbergian psychology forged its interdisciplinary links with philosophy and education, rather than with biology as Wilson had hoped. And finally, the hypothalamus was found to play little role in moral judgment.

Despite these errors in detail, Wilson got the big picture right. The synthesis began in the 1990s with a new set of ingredients, and it has transformed the study of morality today. Wilson was also right that the key link between the social and natural sciences was the study of emotion and the “emotive centers” of the brain. A quantitative analysis of the publication database in psychology shows that research on morality and emotion grew steadily in the 1980s and 1990s (relative to other topics), and then grew very rapidly in the past 5 years (fig. S1).

In this Review, I suggest that the key factor that catalyzed the new synthesis was the “affective revolution” of the 1980s—the increase in research on emotion that followed the “cognitive revolution” of the 1960s and 1970s. I describe three principles, each more than 100 years old, that were revived during the affective revolution. Each principle links together insights from several fields, particularly social psychology, neuroscience, and evolutionary theory. I conclude with a fourth principle that I believe will be the next step in the synthesis.

Principle 1: Intuitive Primacy (but Not Dictatorship)

Kohlberg thought of children as budding moral philosophers, and he studied their reasoning as they struggled with moral dilemmas (e.g., Should a man steal a drug to save his wife’s life?). But in recent years, the importance of moral reasoning has been questioned as social psychologists have increasingly embraced a version of the “affective primacy” principle, articulated in the 1890s by Wilhelm Wundt and greatly expanded in 1980 by Robert Zajonc (2). Zajonc reviewed evidence that the human mind is composed of an ancient, automatic, and very fast affective system and a phylogenetically newer, slower, and motivationally weaker cognitive system. Zajonc’s basic point was that brains are always and automatically evaluating everything they perceive, and that higher-level human thinking is preceded, permedated, and influenced by affective reactions (simple feelings of like and dislike) which push us gently (or not so gently) toward approach or avoidance.

Evolutionary approaches to morality generally suggest affective primacy. Most propose that the building blocks of human morality are emotional (3, 4) (e.g., sympathy in response to suffering, anger at nonreciprocators, affection for kin and allies) and that some early forms of these building blocks were already in place before the hominid line split off from that of Pan 5 to 7 million years ago (5). Language and the ability to engage in conscious moral reasoning came much later, perhaps only in the past 100 thousand years, so it is implausible that the neural mechanisms that control human judgment and behavior were suddenly rewired to hand control of the organism over to this new deliberative faculty.

Social-psychological research strongly supports Zajonc’s claims about the speed and ubiquity of affective reactions (6). However, many have objected to the contrast of “affect” and “cognition,” which seems to imply that affective reactions don’t involve information processing or computation of any kind. Zajonc did not say that, but to avoid ambiguity I have drawn on the work of Bargh (7) to argue that the most useful contrast for moral psychology is between two kinds of cognition: moral intuition and moral reasoning (8). Moral intuition refers to fast, automatic, and (usually) affect-laden processes in which an evaluative feeling of good-bad or like-dislike (about the actions or character of a person) appears in consciousness without any awareness of having gone through steps of search, weighing evidence, or inferring a conclusion. Moral reasoning, in contrast, is a controlled and “cooler” (less affective) process; it is conscious mental activity that consists of transforming information about people and their actions in order to reach a moral judgment or decision.

My attempt to illustrate the new synthesis in moral psychology is the Social Intuitionist Model (8), which begins with the intuitive primacy principle. When we think about sticking a pin into a child’s hand, or we hear a story about a person slapping her father, most of us have an automatic intuitive reaction that includes a flash of negative affect. We often engage in conscious verbal reasoning too, but this controlled process can occur only after the first automatic process has run, and it is often influenced by the initial moral intuition. Moral reasoning, when it occurs, is usually a post-hoc process in which we search for evidence to support our initial intuitive reaction.

Evidence that this sequence of events is the standard or default sequence comes from studies indicating that (i) people have nearly instant implicit reactions to scenes or stories of moral violations (9); (ii) affective reactions are usually good predictors of moral judgments and behaviors (10, 11); (iii) manipulating emotional reactions, such as through hypnosis, can alter moral judgments (12); and (iv) people can sometimes be “morally dumbfounded”—they can know intuitively that something is wrong, even when they cannot explain why (8, 13). Furthermore, studies of everyday reasoning (14) demonstrate that people generally begin reasoning by setting out to confirm their initial hypothesis. They rarely seek disconfirming evidence, and are quite good at finding support for whatever they want to believe (15).
The importance of affect-laden intuitions is a central theme of neuroscientific work on morality. Damasio (16) found that patients who had sustained damage to certain areas of the prefrontal cortex retained their “cognitive” abilities by most measures, including IQ and explicit knowledge of right and wrong, but they showed massive emotional deficits, and these deficits crippled their judgment and decision-making. They lost the ability to feel the normal flashes of affect that the rest of us feel when we simply hear the words “slap your father.” They lost the ability to use their bodies—or, at least, to integrate input from brain areas that map bodily reactions—to feel what they would actually feel if they were in a given situation. Later studies of moral judgment have confirmed the importance of areas of the medial prefrontal cortex, including ventromedial prefrontal cortex and the medial frontal gyrus (17, 18). These areas appear to be crucial for integrating affect (including expectations of reward and punishment) into decisions and plans. Other areas that show up frequently in functional magnetic resonance imaging studies include the amygdala and the frontal insula (9, 11, 16). These areas seem to be involved in sounding a kind of alarm, and for then “tilting the pinball machine,” as it were, to push subsequent processing in a particular direction.

Affective reactions push, but they do not absolutely force. We can all think of times when we deliberated about a decision and went against our first (often selfish) impulse, or when we changed our minds about a person. Greene et al. (19) caught the brain in action overriding its initial intuitive response. They created a class of difficult dilemmas, for example: You would smother your own baby if it was the only way to keep her from crying and giving away your hiding place to the enemy soldiers looking for you, who would then kill the whole group of you hiding in the basement? Subjects were slow to respond to cases like these and, along the way, exhibited increased activity in the anterior cingulate cortex, a brain region that responds to internal conflict. Some subjects said “yes” to cases like these, and they exhibited increased activity in the dorsolateral prefrontal cortex, suggesting that they were doing additional processing and overriding their initial flash of horror.

There are at least three ways we can override our immediate intuitive responses. We can use conscious verbal reasoning, such as considering the costs and benefits of each course of action. We can reframe a situation and see a new angle or consequence, thereby triggering a second flash of intuition that may compete with the first. And we can talk with people who raise new arguments, which then trigger in us new flashes of intuition followed by various kinds of reasoning. The social intuitionist model includes separate paths for each of these three ways of changing one’s mind, but it says that the first two pathways are rarely used, and that most moral change happens as a result of social interaction. Other people often influence us, in part by presenting the counterevidence we rarely seek out ourselves. Some researchers believe, however, that private, conscious verbal reasoning is either the ultimate authority or at least a frequent contributor to our moral judgments and decisions (19–21). There are at present no data on how people revise their initial judgments in everyday life (outside the lab), but we can look more closely at research on reasoning in general. What role is reasoning fit to play?

**Table 1.** What’s your price? Write in the minimum amount that someone would have to pay you (anonymously and secretly) to convince you to do these 10 actions. For each one, assume there will be no social, legal, or material consequences to you afterward. *Homo economicus* would prefer the option in column B to the option in column A for action 1 and would be more or less indifferent to the other four pairs. In contrast, a person with moral motives would (on average) require a larger payment to engage in the actions in column B and would feel dirty or degraded for engaging in some of these actions for personal enrichment. These particular actions were generated to dramatize moral motives, but they also illustrate the five-foundations theory of intuitive ethics (41, 42).

<table>
<thead>
<tr>
<th>Moral category</th>
<th>Column B</th>
<th>Column A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harm/care</td>
<td>$___</td>
<td>$___</td>
</tr>
<tr>
<td>Fairness/reciprocity</td>
<td>$___</td>
<td>$___</td>
</tr>
<tr>
<td>Ingroup/loyalty</td>
<td>$___</td>
<td>$___</td>
</tr>
<tr>
<td>Authority/respect</td>
<td>$___</td>
<td>$___</td>
</tr>
<tr>
<td>Purity/sanctity</td>
<td>$___</td>
<td>$___</td>
</tr>
</tbody>
</table>

**Principle 2: (Moral) Thinking Is for (Social) Doing**

During the cognitive revolution, many psychologists adopted the metaphor that people are “intuitive scientists” who analyze the evidence of everyday experience to construct internal representations of reality. In the past 15 years, however, many researchers have rediscovered William James’ pragmatist dictum that “thinking is for doing.” According to this view, moral reasoning is not like that of an idealized scientist or judge seeking the truth, which is often useful; rather, moral reasoning is like that of a lawyer or politician seeking whatever is useful, whether or not it is true.

One thing that is always useful is an explanation of what you just did. People in all societies gossip, and the ability to track reputations and burnish one’s own is crucial in most recent accounts of the evolution of human morality (22, 23). The first rule of life in a dense web of gossip is: Be careful what you do. The second rule is: What you do matters less than what people think you did, so you’d better be able to frame your actions in a positive light. You’d better become a good “intuitive politician” (24).

From this social-functionalist perspective, it is not surprising that people are generally more accurate in their predictions of what others will do than in their (morally rosier) predictions about what they themselves will do (25), and it is not...
surprising that people so readily invent and confidently tell stories to explain their own behaviors (26). Such “confabulations” are often reported in neuroscientific work; when brain damage or surgery creates bizarre behaviors or beliefs, the patient rarely says “Gosh, why did I do that?” Rather, the patient’s “interpreter module” (27) struggles heroically to weave a story that is then offered confidently to others. Moral reasoning is often like the press secretary for a secretive administration—constantly generating the most persuasive arguments it can muster for policies whose true origins and goals are unknown (8, 28).

The third rule of life in a web of gossip is: Be prepared for other people’s attempts to deceive and manipulate you. The press secretary’s pronouncements usually contain some useful information, so we attend to them, but we don’t take them at face value. We easily switch into “intuitive prosecutor” mode (24), using our reasoning capacities to challenge people’s excuses and to seek out—or fabricate—evidence against people we don’t like. Thalia Wheatley and I (12) recently created prosecutorial moral confabulations by giving hypnotizable subjects a post-hypnotic suggestion that they would feel a flash of disgust whenever they read a previously neutral word (“take” for half the subjects; “often” for the others). We then embedded one of those two words in six short stories about moral violations (e.g., accepting bribes or eating one’s dead pet dog) and found that stories that included the disgust-enhanced word were condemned more harshly than those that had no such flash.

To test the limiting condition of this effect, we included one story with no wrongdoing, about Dan, a student council president, who organizes faculty-student discussions. The story included one of two versions of this sentence: “He [tries to take] [often picks] topics that appeal to both professors and students in order to stimulate discussion.” We expected that subjects who felt a flash of disgust while reading this sentence would condemn Dan (intuitive pramiacy), search for a justification (post-hoc reasoning), fail to find one, and then be forced to override their hypnotically induced gut feeling using controlled processes. Most did. But to our surprise, one third of the subjects in the hypnotic disgust condition (and none in the other) said that Dan’s action was wrong to some degree, and a few came up with the sort of post-hoc confabulations that Gazzaniga reported in some split-brain patients, such as “Dan is a popularity-seeking snob” or “It just seems like he’s up to something.” They invented reasons to make sense of their otherwise inexplicable feeling of disgust.

When we engage in moral reasoning, we are using relatively new cognitive machinery that was shaped by the adaptive pressures of life in a reputation-obsessed community. We are capable of using this machinery dispassionately, such as when we consider abstract problems with no personal ramifications. But the machinery itself was “designed” to work with affect, not free of it, and in daily life the environment usually obliges by triggering some affective response. But how did humans, and only humans, develop these gossipy communities in the first place?

**Principle 3: Morality Binds and Builds**

Nearly every treatise on the evolution of morality covers two processes: kin selection (genes for altruism can evolve if altruism is targeted at kin) and reciprocal altruism (genes for altruism can principle, I suggest, is the insight of the sociologist Emile Durkheim (30) that morality binds and builds; it constrains individuals and ties them to each other to create groups that are emergent entities with new properties.

A moral community has a set of shared norms about how members ought to behave, combined with means for imposing costs on violators and/or channeling benefits to cooperators. A big step in modeling the evolution of such communities is the extension of reciprocal altruism by “indirect reciprocity” (31) in which virtue pays by improving one’s reputation, which elicits later cooperation from others. Reputation is a powerful force for strengthening and enlarging moral communities (as users of ebay.com know). When repeated-play behavioral economics games allow players to know each others’ reputations, cooperation rates skyrocket (29). Evolutionary models show that indirect reciprocity can solve the problem of free-riders (which doomed simpler models of altruism) in moderately large groups (32), as long as people have access to information about reputations (e.g., gossip) and can then engage in low-cost punishment such as shunning.

However the process began, early humans sometimes found ways to solve the free-rider problem and to live in larger cooperative groups. In so doing, they may have stepped through a major transition in evolutionary history (33). From prokaryotes to eukaryotes, from single-celled organisms to plants and animals, and from individual animals to hives, colonies, and cooperative groups, the simple rules of Darwinian evolution never change, but the complex game of life changes when radically new kinds of players take the field. Ant colonies are a kind of super-organism whose proliferation has altered the ecology of our planet. Ant colonies compete with each other, and group selection therefore shaped ant behavior and made ants extraordinarily cooperative within their colonies. However, biologists have long resisted the idea that group selection contributed to human altruism because human groups do not restrict breeding to a single queen or breeding pair. Genes related to altruism for the good of the group are therefore vulnerable to replacement by genes related to more selfish free-riding strategies. Human group selection was essentially declared off-limits in 1966 (34).

In the following decades, however, several theorists realized that human groups engage in cultural practices that modify the circumstances under which genes are selected. Just as a modified gene for adult lactose tolerance evolved
in tandem with cultural practices of raising dairy cows, so modified genes for moral motives may have evolved in tandem with cultural practices and institutions that rewarded group-beneficial behaviors and punished selfishness. Psychological mechanisms that promote uniformity within groups and maintain differences across groups create conditions in which group selection can occur, both for cultural traits and for genes (23, 35). Even if groups vary little or not at all genetically, groups that develop norms, practices, and institutions that elicit more group-beneficial behavior can grow, attract new members, and replace less cooperative groups. Furthermore, preagicultural human groups may have engaged in warfare often enough that group selection altered gene frequencies as well as cultural practices (36). Modified genes for extreme group solidarity during times of conflict may have evolved in tandem with cultural practices that led to greater success in war.

Humans attain their extreme group solidarity by forming moral communities within which selfishness is punished and virtue rewarded. Durkheim believed that gods played a crucial role in the formation of such communities. He saw religion as “a unified system of beliefs and practices relative to sacred things, that is to say, things set apart and forbidden—beliefs and practices which unite into one single moral community called a church, all those who adhere to them” (30). D. S. Wilson (35) has argued that the coevolution of religions and religious minds created conditions in which multilevel group selection operated, transforming the older morality of small groups into a more tribal form that could unite larger populations. As with ants, group selection greatly increased cooperation within the group, but in part for the adaptive purpose of success in conflict between groups.

Whatever the origins of religiosity, nearly all religions have culturally evolved complexes of practices, stories, and norms that work together to suppress the self and connect people to something beyond the self. Newberg (37) found that religious experiences often involve decreased activity in brain areas that maintain maps of the self’s boundaries and position, consistent with widespread reports that mystical experiences involve feelings of merging with God or the universe. Studies of ritual, particularly those involving the sort of synchronized motor movements common in religious rites, indicate that such rituals serve to bind participants together in what is often reported to be an ecstatic state of union (38). Recent work on mirror neurons indicates that, whereas such neurons exist in other primates, they are much more numerous in human beings, and they serve to synchronize our feelings and movements with those of others around us (39). Whether people use their mirror neurons to feel another’s pain, enjoy a synchronized dance, or bow in unison toward Mecca, it is clear that we are prepared, neurologically, psychologically, and culturally, to link our consciousness, our emotions, and our motor movements with those of other people.

**Principle 4: Morality Is About More Than Harm and Fairness**

If I asked you to define morality, you’d probably say it has something to do with how people ought to treat each other. Nearly every research program in moral psychology has focused on one of two aspects of interpersonal treatment: (i) harm, care, and altruism (people are vulnerable and often need protection) or (ii) fairness, reciprocity, and justice (people have rights to certain resources or kinds of treatment). These two topics bear a striking match to the two evolutionary mechanisms of kin selection (which presumably made us sensitive to the suffering and needs of close kin) and reciprocal altruism (which presumably made us exquisitely sensitive to who deserves what). However, if group selection did reshape human morality, then there might be a kind of tribal overlay (23)—a coevolved set of cultural practices and moral intuitions—that are not about how to treat other individuals but about how to be a part of a group, especially a group that is competing with other groups.

In my cross-cultural research, I have found that the moral domain of educated Westerners is narrower—more focused on harm and fairness—than it is elsewhere. Extending a theory from cultural psychologist Richard Shweder (40), Jesse Graham, Craig Joseph, and I have suggested that there are five psychological foundations of morality (17, 23) that might coevolve with tribal practices and lend themselves to the study of moral behavior. These five moral foundations combine to create conditions in which group selection can operate: (i) care (people want to see others’ suffering minimized), (ii) harm (people want to see others’ harms reduced), (iii) harm and fairness (people often care more about the impact of harm on one individual than on another), (iv) power and authority (people care about how to be a part of a group, especially a group that is competing with other groups), and (v) purity and worldview (people want to maintain a group’s boundaries and position, consistent with widespread reports that mystical experiences involve feelings of merging with God or the universe). When people judge stories about how to be a part of a group, especially a group that is competing with other groups, they often use one or more of these five foundations (17).

Research on morality beyond harm and fairness is important. Much of human behavior, you might consider saying the ethos of caring (21), is developed during periods of war. And because morality may be as much a product of cultural evolution as genetic evolution, it can change substantially in a generation or two. For example, as technological advances make us more aware of the fate of people in faraway lands, our concerns expand and we increasingly want peace, decency, and cooperation to prevail in other groups, and in the human group as well.

In conclusion, if the host of that erudite quiz show were to allow you 60 seconds to explain human behavior, you might consider saying the following: People are self-interested, but they also care about how they (and others) treat people, and how they (and others) participate in groups. These moral motives are implemented in large part by a variety of affect-laden intuitions that arise quickly and automatically and then influence controlled processes such as moral reasoning. Moral reasoning can correct and override moral intuition, though it is more commonly performed in the service of social goals as people navigate their gossipy worlds. Yet even though morality is partly a game of selfpromotion, people do sincerely want peace, decency, and cooperation to prevail within their groups. And because morality may be as much a product of cultural evolution as genetic evolution, it can change substantially in a generation or two. For example, as technological advances make us more aware of the fate of people in faraway lands, our concerns expand and we increasingly want peace, decency, and cooperation to prevail in other groups, and in the human group as well.

**References and Notes**

Embodying Emotion
Paula M. Niedenthal*

Recent theories of embodied cognition suggest new ways to look at how we process emotional information. The theories suggest that perceiving and thinking about emotion involve perceptual, somatosensory, and motoric reexperiencing (collectively referred to as "embodiment") of the relevant emotion in one's self. The embodiment of emotion, when induced in human participants by manipulations of facial expression and posture in the laboratory, causally affects how emotional information is processed. Congruence between the recipient's bodily expression of emotion and the sender's emotional tone of language, for instance, facilitates comprehension of the communication, whereas incongruence can impair comprehension. Taken all together, recent findings provide a scientific account of the familiar contention that "when you're smiling, the whole world smiles with you."

Here is a thought experiment: A man goes into a bar to tell a new joke. Two people are already in the bar. One is smiling and one is frowning. Who is more likely to "get" the punch line and appreciate his joke? Here is another: Two women are walking over a bridge. One is afraid of heights, so her heart pounds and her hands tremble. The other is not afraid at all. On the other side of the bridge, they encounter a man. Which of the two women is the man of her dreams?

You probably guessed that the first person of the pair described in each problem was the right answer. Now consider the following experimental findings:

1) While adopting either a conventional working posture or one of two so-called ergonomic postures, in which the back was straight and the shoulders were held high and back or in which the shoulders and head were slumped, experimental participants learned that they had succeeded on an achievement test completed earlier. Those who received the good news in the slumped posture felt less proud and reported being in a worse mood than participants in the upright or working posture (1).

2) Images that typically evoke emotionally "positive" and "negative" responses were presented on a computer screen. Experimental participants were asked to indicate when a picture appeared by quickly moving a lever. Some participants were instructed to push a lever away from their body, whereas others were told to pull toward their body. Participants who pushed the lever away responded to negative images faster than to positive images, whereas participants who pulled the lever toward themselves responded faster to positive images (2).

3) Under the guise of studying the quality of different headphones, participants were induced either to nod in agreement or to shake their heads in disagreement. While they were "testing" their headphones with one of these two movements, the experimenter placed a pen on the table in front of them. Later, a different experimenter offered the participants the pen that had been placed on the table earlier or a novel pen. Individuals who were nodding their heads preferred the old pen, whereas participants who had been shaking their heads preferred the new one (3).

All of these studies show that there is a reciprocal relationship between the bodily expression of emotion and the way in which emotional information is attended to and interpreted (Fig. 1). Charles Darwin himself defined attitude as a collection of motor behaviors (especially posture) that conveys an organism's emotional response toward an object (4). Thus, it would not have come as any surprise to him that the human body is involved in the acquisition and use of attitudes and preferences. Indeed, one speculates that Darwin would be satisfied to learn that research reveals that (i) when individuals adopt emotion-specific postures, they report experiencing the associated emotions; (ii) when individuals adopt facial expressions or make emotional gestures, their preferences and attitudes are influenced; and (iii) when individuals' motor movements are inhibited, interference in the experience of emotion and processing of emotional information is observed (5). The causal relationship between embodying emotions, feeling emotional states,