

Running head: MORAL FOUNDATIONS THEORY

**Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism**

Jesse Graham<sup>a</sup>, Jonathan Haidt<sup>b</sup>, Sena Koleva<sup>a</sup>, Matt Motyl<sup>c</sup>,

Ravi Iyer<sup>a</sup>, Sean P. Wojcik<sup>d</sup>, & Peter H. Ditto<sup>d</sup>

<sup>a</sup>University of Southern California, <sup>b</sup>New York University,

<sup>c</sup>University of Virginia, <sup>d</sup>University of California – Irvine

in press, *Advances in Experimental Social Psychology*

**Abstract:**

Where does morality come from? Why are moral judgments often so similar across cultures, yet sometimes so variable? Is morality one thing, or many? Moral Foundations Theory (MFT) was created to answer these questions. In this chapter we describe the origins, assumptions, and current conceptualization of the theory, and detail the empirical findings that MFT has made possible, both within social psychology and beyond. Looking toward the future, we embrace several critiques of the theory, and specify five criteria for determining what should be considered a foundation of human morality. Finally, we suggest a variety of future directions for MFT and for moral psychology.

## Table of Contents

<a href="#">1 The Origins of Moral Foundations Theory</a>	5
<a href="#">2 The Current Theory</a>	7
<a href="#">2.1 Nativism</a>	7
<a href="#">2.2 Cultural learning</a>	9
<a href="#">2.3 Intuitionism</a>	10
<a href="#">2.4 Pluralism</a>	12
<a href="#">2.4.1 The Care/harm Foundation</a>	12
<a href="#">2.4.2 The Fairness/cheating Foundation</a>	13
<a href="#">2.4.3 The Loyalty/betrayal Foundation</a>	13
<a href="#">2.4.4 The Authority/subversion Foundation</a>	13
<a href="#">2.4.5 The Sanctity/degradation Foundation</a>	14
<a href="#">3 Empirical Findings</a>	14
<a href="#">3.1 Methods and Measures</a>	14
<a href="#">3.2 Moral Foundations and Political Ideology</a>	15
<a href="#">3.2.1 Ideology in Five Dimensions</a>	16
<a href="#">3.2.2 Personality-Morality-Ideology Linkages</a>	17
<a href="#">3.2.3 Political Stereotypes and Interpersonal Judgment</a>	19
<a href="#">3.2.4 Beyond Liberal and Conservative</a>	20
<a href="#">3.3 Moral Foundations and Other Psychological Constructs</a>	22
<a href="#">3.3.1 Attitudes</a>	23
<a href="#">3.3.2 Emotion</a>	23
<a href="#">3.3.3 Moral Character</a>	24
<a href="#">3.3.4 Other Psychological Constructs</a>	25
<a href="#">3.4 Cross-Cultural Differences and Intergroup Relations</a>	25
<a href="#">3.4.1 East-West Cultural Differences</a>	26
<a href="#">3.4.2 Other Cultural Differences</a>	26
<a href="#">3.4.3 Intergroup Relations</a>	27
<a href="#">3.5 Implicit Moral Cognition</a>	28
<a href="#">3.5.1 Implicit Morality Across Ideology</a>	28
<a href="#">3.5.2 Psychophysiological Pluralism</a>	29
<a href="#">4 Future Directions</a>	30

<a href="#">4.1 Criticisms of the Theory</a> .....	30
<a href="#">4.1.1 Critiques of Nativism</a> .....	31
<a href="#">4.1.2 Critiques of Cultural Learning</a> .....	32
<a href="#">4.1.3 Critiques of Intuitionism</a> .....	32
<a href="#">4.1.4 Critiques of Pluralism per se</a> .....	33
<a href="#">4.1.5 Alternate Pluralisms</a> .....	34
<a href="#">4.2 Getting Specific: What Does it Take to be a Foundation?</a> .....	36
<a href="#">4.2.1 Criterion 1</a> .....	37
<a href="#">4.2.2 Criterion 2</a> .....	37
<a href="#">4.2.3 Criterion 3</a> .....	38
<a href="#">4.2.4 Criterion 4</a> .....	39
<a href="#">4.2.5 Criterion 5</a> .....	40
<a href="#">4.3 Looking Ahead: New Directions for Moral Foundations Research</a> .....	41
<a href="#">4.3.1 Method-Theory Co-Evolution of MFT</a> .....	41
<a href="#">4.3.2 Applying MFT to New Areas, and New Questions</a> .....	41
<a href="#">5 Conclusion</a> .....	43

### Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism

“The supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.” (Einstein, 1934, p. 165)

“I came to the conclusion that there is a plurality of ideals, as there is a plurality of cultures and of temperaments... There is not an infinity of [values]: the number of human values, of values which I can pursue while maintaining my human semblance, my human character, is finite—let us say 74, or perhaps 122, or 27, but finite, whatever it may be. And the difference this makes is that if a man pursues one of these values, I, who do not, am able to understand why he pursues it or what it would be like, in his circumstances, for me to be induced to pursue it. Hence the possibility of human understanding.” (Berlin, 2001, p. 12)

Scientists value parsimony as well as explanatory adequacy. There is, however, an inherent tension between these two values. When we try to explain an aspect of human nature or behavior using only a single construct, the gain in elegance is often purchased with a loss of descriptive completeness. We risk imitating Procrustes, the mythical blacksmith who forced his guests to fit into an iron bed exactly, whether by stretching them out or by cutting off their legs. Einstein, in our opening quote, warns against this Procrustean over-valuation of parsimony.

In this article we ask: How many “irreducible basic elements” are needed to represent, understand, and explain the breadth of the moral domain? We use the term *monist* to describe scholars who assert that the answer is: *one*. This *one* is usually identified as justice or fairness, as Lawrence Kohlberg asserted: “Virtue is ultimately one, not many, and it is always the same ideal form regardless of climate or culture... The name of this ideal form is justice” (Kohlberg 1971, p. 232; see also Baumard, André, & Sperber, in press). The other common candidate for being the one foundation of morality is sensitivity to harm (e.g., Gray, Young, & Waytz, 2012), or else related notions of generalized human welfare or happiness (e.g., Harris, 2010). Monists generally try to show that all manifestations of morality are derived from an underlying psychological architecture for implementing the one basic value or virtue that they propose.

Other theorists—whom we’ll call *pluralists*—assert that the answer is: *more than one*. William James’s (1909/1987) extended critique of monism and absolutism, *A Pluralistic Universe*, identifies the perceived messiness of pluralism as a major source of intellectual resistance to it:

Whether materialistically or spiritualistically minded, philosophers have always aimed at cleaning up the litter with which the world apparently is filled. They have substituted economical and orderly conceptions for the first sensible tangle; and whether these were morally elevated or only intellectually neat, they were at any rate always aesthetically pure and definite, and aimed at ascribing to the world something clean and intellectual in the way of inner structure. As compared with all these rationalizing pictures, the pluralistic empiricism which I profess offers but a sorry appearance. It is a turbid, muddled, gothic sort of an affair, without a sweeping outline and with little pictorial nobility. (p. 650)

Aristotle was an early moral pluralist, dismissed by Kohlberg (1971) for promoting a “bag of virtues.” Carol Gilligan (1982) was a pluralist when she argued that the “ethic of care” was not

derived from (or reducible to) the ethic of justice. Isaiah Berlin said, in our opening quotation, that there is a finite but potentially large number of moral ideals that are within the repertoire of human beings, and that an appreciation of the full repertoire opens the door to mutual understanding.

We are unabashed pluralists, and in this chapter we'll try to convince you that you should be, too. In the first two parts of this chapter we present a pluralist theory of moral psychology – Moral Foundations Theory (MFT). In part three we'll provide an overview of empirical results that we and others have obtained using a variety of measures developed to test the theory. We will show that the pluralism of MFT has led to discoveries that had long been missed by monists. In part four we'll discuss criticisms of the theory and future research directions that are motivated in part by those criticisms. We'll also propose specific criteria that researchers can use to decide what counts as a foundation. Throughout the chapter we will focus on MFT's *pragmatic validity* (Graham, Nosek, Haidt, Iyer, Koleva, & Ditto, 2011) – that is, its scientific usefulness for both answering existing questions about morality and allowing researchers to formulate new questions.

We grant right at the start that our particular list of moral foundations is unlikely to survive the empirical challenges of the next several years with no changes. But we think that our general approach is likely to stand the test of time. We predict that 20 years from now moral psychologists will mostly be pluralists who draw on both cultural and evolutionary psychology to examine the psychological mechanisms that lead people and groups to hold divergent moral values and beliefs.

We also emphasize, at the outset, that our project is descriptive, not normative. We are not trying to say who or what is morally right or good. We are simply trying to analyze an important aspect of human social life. Cultures vary morally, as do individuals within cultures. These differences often lead to hostility, and sometimes violence. We think it would be helpful for social psychologists, policy makers, and citizens more generally to have a language in which they can describe and understand moralities that are not their own. We think a pluralistic approach is necessary for this descriptive project. We don't know how many moral foundations there really are. There may be 74, or perhaps 122, or 27, or maybe only five, but certainly more than one. And moral psychologists who help people to recognize the inherent pluralism of moral functioning will be at the forefront of efforts to promote the kind of "human understanding" that Berlin described.

## **1 The Origins of Moral Foundations Theory**

For centuries, people looked at the map of the world and noted that the east coast of South America fits reasonably well into the west coast of Africa. The two coasts even have similar rock formations and ancient plant fossils. These many connections led several geologists to posit the theory of continental drift, which was confirmed in the early 1960s by evidence that the sea floor was spreading along mid-oceanic ridges.

Similarly, for decades, social scientists noted that many of the practices widely described by anthropologists fit reasonably well with the two processes that were revolutionizing evolutionary biology: kin selection and reciprocal altruism. When discussing altruism, Dawkins (1976) made occasional reference to the findings of anthropologists to illustrate Hamilton's (1964) theory of kin selection, while Trivers (1971) reviewed anthropological evidence illustrating reciprocity among hunter-gatherers. So the idea that human morality is derived from or constrained by multiple innate mental systems, each shaped by a different evolutionary

process, is neither new nor radical. It is accepted by nearly all who write about the evolutionary origins of morality (e.g., de Waal, 1996; Ridley, 1996; Joyce, 2006; Wright, 1994). The main question up for debate is: how many mental systems are there?

Kohlberg (1969) founded the modern field of moral psychology with his declaration that the answer was one. He developed a grand theory that unified moral psychology as the study of the progressive development of the child's understanding of justice. Building on the work of Piaget (1932/1965), Kohlberg proposed that moral development in all cultures is driven forward by the process of role-taking: as children get more practice at taking each other's perspectives, they learn to transcend their own position and appreciate when and why an action, practice, or custom is fair or unfair. If they come to respect authority or value group loyalty along the way (stage 4), this is an unfortunate way-station at which children over-value conformity and tradition. But if children get more opportunities to role-take they will progress to the post-conventional level (stages 5 & 6), at which authority and loyalty might sometimes be justified, but only to the extent that they promote justice.

The deficiencies of Kohlberg's moral monism were immediately apparent to some of his critics. Gilligan (1982) argued that the morality of girls and women did not follow Kohlberg's one true path, but developed along *two* paths: an ethic of justice, but also an ethic of care that could not be derived from the former. Kohlberg eventually acknowledged that she was right (Kohlberg, Levine, & Hower, 1983). Moral psychologists in the cognitive developmental tradition have generally been comfortable with this dualism: justice *and* care. In fact, the cover of the *Handbook of Moral Development* (Killen & Smetana, 2006) shows two images: the scales of justice, and African statues of a mother and child.

Turiel (1983) allowed for both foundations in his widely-cited definition of the moral domain as referring to "prescriptive judgments of justice, rights, and welfare pertaining to how people ought to relate to each other." (Justice and rights are the Kohlbergian foundation; the concern for "welfare" can encompass Gilligan's "care.") Kohlberg, Gilligan, and Turiel were all united in their belief that morality is about how *individuals* ought to relate to, protect, and respect other individuals.

But what if, in some cultures, even the most advanced moral thinkers value groups, institutions, traditions, and gods? What should we say about local rules for how to be a good group member, or how to worship? If these rules are not closely linked to concerns about justice or care, then should we distinguish them from true moral rules, as Turiel did when he labeled such rules as "social conventions?" Shweder (1990) argued that the cognitive-developmental tradition was studying only a subset of moral concerns, the ones that are most highly elaborated in secular Western societies. Shweder argued for a much more extensive form of pluralism based on his research in Bhubaneswar, India (Shweder, Much, Mahapatra, & Park, 1997). He proposed that around the world, people talk in one or more of *three* moral languages: the ethic of autonomy (relying on concepts such as harm, rights, and justice, which protect autonomous individuals), the ethic of community (relying on concepts such as duty, respect, and loyalty, which preserve institutions and social order), and the ethic of divinity (relying on concepts such as purity, sanctity, and sin, which protect the divinity inherent in each person against the degradation of hedonistic selfishness.)

So now we're up to three. Or maybe it's four? Alan Fiske (1991) proposed that moral judgment relies upon the same four "relational models" that are used to think about and enact social relationships: Communal Sharing, Authority Ranking, Equality Matching, and Market Pricing (see also Rai & Fiske, 2011).

Having worked with both Fiske and Shweder, Haidt wanted to integrate the two theories into a unified framework for studying morality across cultures. But despite many points of contact, the three ethics and four relational models couldn't be neatly merged or reconciled. They are solutions to different problems: categorizing explicit moral discourse (for Shweder) and analyzing interpersonal relationships (for Fiske). After working with the two theories throughout the 1990s – the decade in which evolutionary psychology was reborn (Barkow, Cosmides, & Tooby, 1992) – Haidt sought to construct a theory specifically designed to bridge evolutionary and anthropological approaches to moral judgment. He worked with Craig Joseph, who was studying cultural variation in virtue concepts (Joseph, 2002).

The first step was to broaden the inquiry beyond the theories of Fiske and Shweder to bring in additional theories about how morality varies across cultures. Schwartz and Bilsky's (1990) theory of values offered the most prominent approach in social psychology. Haidt and Joseph also sought out theorists who took an evolutionary approach, trying to specify universals of human moral nature. Brown (1991) offered a list of human universals including many aspects of moral psychology, and de Waal (1996) offered a list of the “building blocks” of human morality that can be seen in other primates.

Haidt and Joseph (2004) used the analogy of taste to guide their review of these varied works. The human tongue has five discrete taste receptors (for sweet, sour, salt, bitter, and umami). Cultures vary enormously in their cuisines, which are cultural constructions shaped by historical events, yet the world's many cuisines must ultimately please tongues equipped with just five innate and universal taste receptors. What are the best candidates for being the innate and universal “moral taste receptors” upon which the world's many cultures construct their moral cuisines? What are the concerns, perceptions, and emotional reactions that consistently turn up in moral codes around the world, and for which there are already-existing evolutionary explanations?

Haidt and Joseph identified five best candidates: Care/harm, Fairness/cheating, Loyalty/betrayal, Authority/subversion, and Sanctity/degradation.<sup>1</sup> We believe that there are more than five; for example, Haidt (2012) has suggested that Liberty/oppression should be considered a sixth foundation (see section 4.1 for other candidate foundations). We will explain the nature of these foundations in the next section, and we'll offer a list of criteria for “foundationhood” in section 4.2. But before we do, the broader theoretical underpinnings of MFT need to be explained.

## 2 The Current Theory

Moral Foundations Theory can be summarized in four claims. If any of these claims is disproved, or is generally abandoned by psychologists, then MFT would need to be abandoned, too.

### 2.1 Nativism – *There is a “first draft” of the moral mind.*

Some scholars think that evolutionary and cultural explanations of human behavior are competing approaches – one reductionist, one constructivist-- but Moral Foundations Theory was created precisely to integrate the two (see also A. P. Fiske, 1991; Richerson & Boyd, 2005). Our definition of nativism makes this clear: Innate means *organized in advance of experience*. We do

---

<sup>1</sup> Prior to 2012 we used slightly different terms: Harm/care, Fairness/reciprocity, Ingroup/loyalty, Authority/respect, and Purity/sanctity.

not take it to mean hardwired or insensitive to environmental influences, as some critics of nativism define innateness (e.g., Suhler & Churchland, 2011). Instead, we borrow Marcus's (2004) metaphor that the mind is like a book: "Nature provides a first draft, which experience then revises... 'Built-in' does not mean unmalleable; it means 'organized in advance of experience'" (pp. 34 & 40). The genes (collectively) write the first draft into neural tissue, beginning in utero but continuing throughout childhood. Experience (cultural learning) revises the draft during childhood, and even (to a lesser extent) during adulthood.

We think it is useful to conceptualize the first draft and the editing process as distinct. You can't infer the first draft from looking at a single finished volume (i.e., one adult, or one culture). But if you examine volumes from all over the world, and you find a great many specific ideas expressed in most (but not necessarily all) of the volumes, using different wording, then you would be justified in positing that there was some sort of common first draft or outline, some common starting point to which all finished volumes can be traced. Morality is innate *and* highly dependent on environmental influences.

The classic study by Mineka and Cook (1988) is useful here. Young rhesus monkeys, who showed no prior fear of snakes—including plastic snakes—watched a video of an adult monkey reacting fearfully (or not) to a plastic snake (or to plastic flowers). The monkeys learned from a single exposure to snake-fearing monkey to be afraid of the plastic snake, but a single exposure to a flower-fearing monkey did nothing. This is an example of "preparedness" (Seligman, 1971). Evolution created something "organized in advance of experience" that made it easy for monkeys – and humans (DeLoache & LoBue, 2009) – to learn to fear snakes. Evolution did not simply install a general-purpose learning mechanism which made the monkeys take on all the fears of adult role models equally.

We think the same is likely true about moral development. It's probably quite easy to teach kids to want revenge just by exposing them to role models who become angry and vengeful when treated unfairly, but it is probably much more difficult to teach children to love their enemies just by exposing them, every Sunday for 20 years, to stories about a role model who loved his enemies. We are prepared to learn vengefulness, in a way that we are *not* prepared to learn to offer our left cheek to those who smite us on our right cheek.

How can moral knowledge be innate? Evolutionary psychologists have discussed the issue at length. They argue that recurrent problems and opportunities faced by a species over long periods of time often produce domain-specific cognitive adaptations for responding rapidly and effectively (Tooby & Cosmides, 1992; Pinker, 1997). These adaptations are often called modules, which evolutionary theorists generally do not view as fully "encapsulated" entities with "fixed neural localizations" (Fodor, 1983), but as *functionally specialized mechanisms* which work together to solve recurrent adaptive problems quickly and efficiently (Barrett & Kurzban, 2006). There is not one general-purpose digestion organ, and if there ever was such an organ, its owners lost out to organisms with more efficient modular designs.

The situation is likely to be the same for higher cognition: there is not one general-purpose thinking or reasoning organ that produces moral judgments, as Kohlberg seemed to suppose. Rather, according to the "massive modularity hypothesis" (Sperber, 1994; 2005), the mind is thought to be full of small information-processing mechanisms, which make it easy to solve – or to learn to solve – certain kinds of problems, but not other kinds.

Tooby, Cosmides and Barrett (2005) argue that the study of valuation, even more than other areas of cognition, reveals just how crucial it is to posit innate mental content, rather than positing a few innate general learning mechanisms (such as social learning). Children are born



with a preference (value) for sweetness and against bitterness. The preference for candy over broccoli is not learned by socialization and cannot be undone by role models, threats, or education about the health benefits of broccoli. Tooby et al. (2005) suggest that the same thing is true for valuation in all domains, including the moral domain. Just as the tongue and brain are designed to yield pleasure when sweetness is tasted, there are cognitive modules that yield pleasure when fair exchanges occur, and displeasure when one detects cheaters. In the moral domain, the problems to be solved are social, and the human mind evolved a variety of mechanisms that enable individuals (and perhaps groups) to solve those problems within the “moral matrices” – webs of shared meaning and evaluation – that began to form as humans became increasingly cultural creatures during the last half-million years (see Haidt, 2012, chapter 9, which draws on Richerson & Boyd, 2005, and Tomasello et al., 2005).

Moral Foundations Theory proposes that the human mind is organized in advance of experience so that it is prepared to learn values, norms, and behaviors related to a diverse set of recurrent adaptive social problems (specified below in Table 1). We think of this innate organization as being implemented by sets of related modules which work together to guide and constrain responses to each particular problem. But you don’t have to embrace modularity, or any particular view of the brain, to embrace MFT. You only need to accept that there is a first draft of the moral mind, organized in advance of experience by the adaptive pressures of our unique evolutionary history.

## ***2.2 Cultural learning – The first draft gets edited during development within a particular culture.***

A dictum of cultural psychology is that “Culture and psyche make each other up” (Shweder, 1990, p. 24). If there were no first draft of the psyche, then groups would be free to invent utopian moralities (e.g., “from each according to his ability, to each according to his need”), and they’d be able to pass them on to their children because all moral ideas would be equally learnable. This clearly is not the case (e.g., Spiro, 1956; Pinker, 2002). Conversely, if cultural learning played no formative role, then the first draft would be the final draft, and there’d be no variation across cultures.<sup>2</sup> This clearly is not the case either (e.g., Haidt, Koller, & Dias, 1993; Shweder, Mahapatra, & Miller, 1987).

The cognitive anthropologist Dan Sperber has proposed a version of modularity theory that we believe works very well for higher cognition in general, and for moral psychology in particular. Citing Marler’s (1991) research on song learning in birds, Sperber (2005) proposes that many of the modules present at or soon after birth are “learning modules.” That is, they are innate templates or “learning instincts” whose function is to generate a host of more specific modules as the child develops. They generate “the working modules of acquired cognitive competence” (p. 57). They are a way of explaining phenomena such as preparedness (Seligman, 1971).

For example, children in traditional Hindu households are frequently required to bow, often touching their heads to the floor or to the feet of revered elders and guests. Bowing is used in religious contexts as well, to show deference to the gods. By the time a Hindu girl reaches adulthood, she will have developed culturally-specific knowledge that makes her automatically initiate bowing movements when she encounters, say, a respected politician for the first time.

---

<sup>2</sup> Other than those due to individual development, e.g., some cultures might offer more opportunities for role taking, which would cause their members to be more successful in self-constructing their own moralities. This is how Kohlberg (1969) explained cultural differences in moral reasoning between Western and non-Western nations.

Note that this knowledge is not just factual knowledge – it includes feelings and motor schemas for bowing and otherwise showing deference. Sperber (2005) refers to this new knowledge –in which a pattern of appraisals is linked to a pattern of behavioral outputs—as an acquired module, generated by the original “learning module.” But one could just as well drop the modularity language at this point and simply assert that children acquire all kinds of new knowledge, concepts, and behavioral patterns as they employ their innately given moral foundations within a particular cultural context. A girl raised in a secular American household will have no such experiences in childhood, and may reach adulthood with no specialized knowledge or ability to detect hierarchy and show respect for hierarchical authorities.

Both girls started off with the same sets of universal learning modules – including the set we call the Authority/subversion foundation. But in the Hindu community, culture and psyche worked together to generate a host of more specific authority-respecting abilities (or modules, if you prefer). In the secular American community, such new abilities were not generated, and the American child is more likely to hold anti-authoritarian values as an adult. An American adult may still have inchoate feelings of respect for some elders, and might even find it hard to address some elders by first name (see Brown & Ford, 1964). But our claim is that the universal (and incomplete) first draft of the moral mind gets filled in and revised so that the child can successfully navigate the moral “matrix” he or she actually experiences.

This is why we chose the architectural metaphor of a “foundation.” Imagine that thousands of years ago, extraterrestrial aliens built 100 identical monumental sites scattered around the globe. But instead of building entire buildings, they just built five solid stone platforms, in irregular shapes, and left each site like that. If we were to photograph those 100 sites from the air today, we’d probably be able to recognize the similarity across the sites, even though at each site people would have built diverse structures out of local materials. *The foundations are not the finished buildings*, but the foundations constrain the kinds of buildings that can be built most easily. Some societies might build a tall temple on just one foundation, and let the other foundations decay. Other societies might build a palace spanning multiple foundations, perhaps even all five. You can’t infer the exact shape and number of foundations by examining a single photograph, but if you collect photos from a few dozen sites, you can.

Similarly, *the moral foundations are not the finished moralities*, although they constrain the kinds of moral orders that can be built. Some societies build their moral order primarily on top of one or two foundations. Others use all five. You can’t see the foundations directly, and you can’t infer the exact shape and number of foundations by examining a single culture’s morality. But if you examine ethnographic, correlational, and experimental data from a few dozen societies, you can. And if you look at the earliest emergence of moral cognition in babies and toddlers, you can see some of them as well (as we’ll show in section 4.2). MFT is a theory about the universal first draft of the moral mind, and about how that draft gets revised in variable ways across cultures.

### ***2.3 Intuitionism – Intuitions come first, strategic reasoning second.***

Compared to the explicit deliberative reasoning studied by Kohlberg, moral judgments, like other evaluative judgments, tend to happen quickly (Zajonc, 1980; see review in Haidt, 2012, chapter 3). Social psychological research on moral judgment was heavily influenced by the “automaticity revolution” of the 1990s. As Bargh and Chartrand (1999, p. 462) put it: “most of a person's everyday life is determined not by their conscious intentions and deliberate choices but by mental processes that are put into motion by features of the environment that operate outside

of conscious awareness and guidance.” They noted that people engage in a great deal of conscious thought, but they questioned whether such thinking generally *causes* judgments or *follows along* after judgments have already been made. Impressed by the accuracy of social judgments based on “thin slices” of behavior (Ambady & Rosenthal, 1992), they wrote: “So 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).

Drawing on this work (including Nisbett & Wilson, 1977; Wegner & Bargh, 1998), Haidt (2001) formulated the Social Intuitionist Model (SIM) and defined moral intuition as:

the sudden appearance in consciousness, or at the fringe of consciousness, of an evaluative feeling (like–dislike, good–bad) about the character or actions of a person, without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion. (Haidt & Bjorklund, 2008, p. 188, modified from Haidt, 2001)

In other words, the SIM proposed that moral evaluations generally occur rapidly and automatically, products of relatively effortless, associative, heuristic processing that psychologists now refer to as System 1 thinking (Stanovich & West, 2000; Kahneman, 2011. See also Bastick, 1982; Bruner, 1960; & Simon, 1992, for earlier analyses of intuition that influenced the SIM). Moral evaluation, on this view, is more a product of the gut than the head, bearing a closer resemblance to aesthetic judgment than principle-based reasoning.

This is not to say that individuals never engage in deliberative moral reasoning. Rather, Haidt’s original formulation of the SIM was careful to state that this kind of effortful System 2 thinking, while seldom the genesis of our moral evaluations, was often initiated by social requirements to explain, defend and justify our intuitive moral reactions to others. This notion that moral reasoning is done primarily for socially strategic purposes rather than to discover the honest truth about who did what to whom, and by what standard that action should be evaluated, is the crucial “social” aspect of the Social Intuitionist Model. We reason to prepare for social interaction in a web of accountability concerns (Dunbar, 1996; Tetlock, 2002). We reason mostly so that we can support our judgments if called upon by others to do so. As such, our moral reasoning, like our reasoning about virtually every other aspect of our lives, is motivated (Ditto, Pizarro, & Tannenbaum, 2009; Kunda, 1990). It is shaped and directed by intuitive, often affective processes that tip the scales in support of desired conclusions. Reasoning is more like arguing than like rational, dispassionate deliberation (Mercier & Sperber, 2011), and people think and act more like intuitive lawyers than intuitive scientists (Baumeister & Newman, 1994; Ditto et al, 2009; Haidt, 2007; 2012).

The SIM is the prequel to MFT. The SIM says that most of the action in moral judgment is in rapid, automatic moral intuitions. These intuitions were shaped by development within a cultural context, and their output can be edited or channeled by subsequent reasoning and self-presentational concerns. Nonetheless, moral intuitions tend to fall into families or categories. MFT was designed to say exactly what those categories are, why we are so morally sensitive to a small set of issues (such as local instances of unfairness or disloyalty), and why these automatic moral intuitions vary across cultures. And this brings us to the fourth claim of MFT.

#### ***2.4 Pluralism – There were many recurrent social challenges, so there are many moral foundations.***

Evolutionary thinking encourages pluralism. As Cosmides and Tooby (1994, p. 91) put it: “Evolutionary biology suggests that there is no principled reason for parsimony to be a design criterion for the mind.” Evolution has often been described as a tinkerer, cobbling together solutions to challenges out of whatever materials are available (Marcus, 2008). Evolutionary thinking also encourages functionalism. Thinking is for doing (S. T. Fiske, 1992; James, 1890/1950), and so innate mental structures, such as the moral foundations, are likely<sup>3</sup> to be a responses to adaptive challenges that faced our ancestors for a very long time.

Table 1 lays out our current thinking. The first row lists five longstanding adaptive challenges that faced our ancestors for millions of years, creating conditions that favored the reproductive success of individuals who could solve the problems more effectively. For each challenge, the most effective modules were the ones that detected the relevant patterns in the social world and responded to them with the optimal motivational profile. Sperber (1994) refers to the set of all objects that a module was “designed”<sup>4</sup> to detect as the *proper domain* for that module. He contrasts the proper domain with the *actual domain*, which is the set of all objects that nowadays happens to trigger the module. But because these two terms are sometimes hard for readers to remember, we’ll use the equivalent terms offered by Haidt (2012): the *original triggers* and the *current triggers*.

We’ll explain the first column – the Care/harm foundation, in some detail, to show how to read the table. We’ll then explain the other four foundations more briefly. We want to reiterate that we do not believe these are the only foundations of morality. These are just the five we began with – the five for which we think the current evidence is best. In section 4.2 we’ll give criteria that can be used to evaluate other candidate foundations.

**2.4.1 The Care/harm foundation.** All mammals face the adaptive challenge of caring for vulnerable offspring for an extended period of time. Human children are unusually dependent, and for an unusually long time. It is hard to imagine that in the book of human nature, the chapter on mothering is completely blank—not structured in advance of experience—leaving it up to new mothers to learn from their culture, or from trial and error, what to do when their baby shows signs of hunger or injury. Rather, mammalian life has always been a competition in which females whose intuitive reactions to their children were optimized to detect signs of suffering, distress, or neediness raised more children to adulthood than did their less sensitive sisters. Whatever functional systems made it easy and automatic to connect perceptions of suffering with motivations to care, nurture, and protect are what we call the Care/harm foundation.

The original triggers of the Care/harm foundation are visual and auditory signs of suffering, distress, or neediness expressed by one’s own child. But the perceptual modules that detect neoteny can be activated by other children, by baby animals (which often share the proportions of children), by stuffed animals and cartoon characters that are deliberately crafted to have the proportions of children, and by stories told in newspapers about the suffering of people (even adults) far away. There are now many ways to trigger feelings of compassion for victims, an experience that is often mixed with anger toward those who cause harm.

---

<sup>3</sup> Spandrels aside (Gould & Lewontin, 1979).

<sup>4</sup> Evolution *is* a design process; it’s just not an intelligent design process. See Richerson & Boyd, 2005.

But these moral emotions are not just private experiences. In all societies, people engage in gossip – discussions about the actions of third parties that are not present, typically including moral evaluations of those parties’ actions (Dunbar, 1996). And as long as people engage in moral discourse, they develop virtue terms. They develop ways of describing the character and actions of others with reference to culturally normative ideals. They develop terms such as “kind” and “cruel” to describe people who care for or harm vulnerable others. Virtues related to the Care foundation may be highly prized and elaborated in some cultures (such as among Buddhists); less so in others (e.g., classical Sparta or Nazi Germany; Koonz, 2003).

*2.4.2 The Fairness/cheating foundation.* All social animals face recurrent opportunities to engage in non-zero-sum exchanges and relationships. Those whose minds are organized in advance of experience to be highly sensitive to evidence of cheating and cooperation, and to react with emotions that compel them to play “tit for tat” (Trivers, 1971), had an advantage over those who had to figure out their next move using their general intelligence. (See Frank, 1988, on how rational actors can’t easily solve “commitment problems,” but moral emotions can.) The original triggers of the Fairness/cheating foundation involved acts of cheating or cooperation by one’s own direct interaction partners, but the current triggers of the foundation can include interactions with inanimate objects (e.g., you put in a dollar, and the machine fails to deliver a soda), or interactions among third parties that one learns about through gossip. People who come to be known as good partners for exchange relationships are praised with virtue words such as fair, just, and trustworthy.

*2.4.3 The Loyalty/betrayal foundation.* Chimpanzee troops compete with other troops for territory (Goodall, 1986); coalitions of chimps compete with other coalitions within troops for rank and power (de Waal, 1982). But when humans developed language, weapons, and tribal markers, such intergroup competition became far more decisive for survival. Individuals whose minds were organized in advance of experience to make it easy for them to form cohesive coalitions were more likely to be part of winning teams in such competitions.<sup>5</sup> Sherif et al.’s (1961/1954) classic Robber’s Cave study activated (and then de-activated) the original triggers of the loyalty foundation. Sports fandom and brand loyalty are examples of how easily modern consumer culture has built upon the foundation and created a broad set of current triggers.

*2.4.4 The Authority/subversion foundation.* Many primates, including chimpanzees and bonobos, live in dominance hierarchies, and those whose minds are structured in advance of experience to navigate such hierarchies effectively and forge beneficial relationships upwards and downwards have an advantage over those who fail to perceive or react appropriately in these complex social interactions (de Waal, 1982; A. Fiske, 1991). The various modules that comprise the Authority/subversion foundation are often at work when people interact with and grant legitimacy to modern institutions such as law courts and police departments, and to bosses and leaders of many kinds. Traits such as obedience and deference are virtues in some subcultures – such as among social conservatives in the USA – but can be seen as neutral or even as vices in others – such as among social liberals (Haidt & Graham, 2009; Stenner, 2005; Frimer et al., in press).

---

<sup>5</sup> There is an intense debate as to whether this competition of groups versus groups counts as group-level selection, and whether group-level selection shaped human nature. On the pro side, see Haidt (2012) ch. 9. On the con side, see Pinker, 2012.

*2.4.5 The Sanctity/degradation foundation.* Hominid history includes several turns that exposed our ancestors to greater risks from pathogens and parasites, for example: leaving the trees behind and living on the ground; living in larger and denser groups; and shifting to a more omnivorous diet, including more meat, some of which was scavenged. The emotion of disgust is widely thought to be an adaptation to that powerful adaptive challenge (Oaten, Stevenson & Case, 2009; Rozin, Haidt, & McCauley, 2008). Individuals whose minds were structured in advance of experience to develop a more effective “behavioral immune system” (Schaller & Park, 2011) likely had an advantage over individuals who had to make each decision based purely on the sensory properties of potential foods, friends, and mates. Disgust and the behavioral immune system have come to undergird a variety of moral reactions, e.g., to immigrants and sexual deviants (Faulkner, Schaller, Park, & Duncan, 2004; Navarrete & Fessler, 2006; Rozin, Haidt & McCauley, 2008). People who treat their bodies as temples are praised in some cultures for the virtues of temperance and chastity.

In sum, MFT is a nativist, cultural-developmental, intuitionist, and pluralist approach to the study of morality. We expect—and welcome—disagreements about our particular list of foundations. But we think that our general approach to the study of morality is well justified and is consistent with recent developments in many fields (e.g., neuroscience and developmental psychology, as we’ll show in section 4). We think it will stand the test of time.

As for the specific list of foundations, we believe the best method for improving it is to go back and forth between theory and measurement. In the next section we’ll show how our initial five foundations have been measured and used in psychological studies.

### **3 Empirical Findings**

In this paper we argue for the pragmatic validity of MFT, and of moral pluralism in general. Debates over our theoretical commitments—such as nativism and pluralism—can go on for centuries, but if a theory produces a steady stream of novel and useful findings, that’s good evidence for its value. MFT has produced such a stream of findings, from researchers both within and outside of social psychology. Through its theoretical constructs, and the methods developed to measure them, MFT has enabled empirical advances that were not possible using monistic approaches. In this section we review some of those findings, covering work on political ideology, relations between foundational concerns and other psychological constructs, cross-cultural differences, intergroup relations, and implicit processes in moral cognition.

#### **3.1 Methods and Measures**

In a provocative article titled “There is nothing so theoretical as a good method,” Anthony Greenwald (2012) argued that while theory development can bring about new methods, method development is just as crucial (if not more so) for the advancement of psychological theory. While MFT’s origins were in anthropology and evolutionary theory, its development has been inextricably linked with the creation and validation of psychological methods by which to test its claims (and, when necessary, revise them accordingly). In fact, we see MFT’s current and future development being one of *method-theory co-evolution*, with theoretical constructs inspiring the creation of new ways to measure them, and data from the measurements guiding development of the theory.

Although a detailed description of all methods and measures created to test MFT's constructs is beyond the scope of this article, researchers interested in what tools are available can find brief descriptions and references in Table 2. As the table indicates, four kinds of MFT measures have been developed: 1. *Self-report surveys* – Although MFT is fundamentally about moral intuitions, these have been the most widely used by far, mostly to describe individual and cultural differences in endorsed moral concerns. (We note that according to most definitions of intuition, including the one we gave in section 2.3, intuitions are available to consciousness and explicit reporting; it's the mechanisms that give rise to the intuition that are inaccessible.) 2. *Implicit measures* – Reaction-time and other methods of implicit social cognition have been modified to bypass self-report and capture reactions to foundation-related words, sentences, and pictures (see section 3.5.1 below). 3. *Psychophysiological and neuroscience methods* – These are also intended to bypass self-report, and measure nonconscious and affective reactions more directly, via facial micro-expressions, event-related potentials, or neuroimaging (see section 3.5.2 below). 4. *Text analysis* – The Moral Foundations Dictionary has been useful for measuring foundation-related word use in a wide range of applications and disciplines, from computer science analyses of blogs (Dehghani, Gratch, Sachdeva, & Sagae, 2011) to digital humanities analyses of 18th-century texts (Pasanek, 2009) to political science analyses of the discourse of political elites (Clifford & Jerit, in press). The Many methods developed have provided initial convergent and discriminant validity for our pluralistic model (see e.g. Graham et al., 2011), and several of them demonstrate the intuitive nature of moral judgment. Materials for most of the methods described in Table 2 can be found at [www.MoralFoundations.org](http://www.MoralFoundations.org).

### 3.2 Moral Foundations and Political Ideology

In his 1992 speech to the Republican National Convention, Pat Buchanan declared that the United States was engaged in a “cultural war” that was “as critical to the kind of nation we shall be as the Cold War itself.” Exemplifying a thesis laid out in less polemic terms a year earlier by Hunter (1991), Buchanan described a battle between two competing moral visions for America. The first championed the virtues of American exceptionalism, traditional families and institutions, and Judeo-Christian sexual propriety (Hunter called this the “orthodox” worldview). The second vision, in Buchanan's dismissive portrayal, was determined to undermine these time-tested institutions and values with support for gay and abortion rights, a squeamish relationship with American power and moral authority, and a penchant for favoring corrosive welfare policies and the habitats of spotted owls over the homes and jobs of hardworking Americans (Hunter called this the “progressive” worldview). In the three decades since Buchanan's opening salvo, there can be little doubt that this cultural divide between conservative and liberal moral sensibilities has only become deeper and more entrenched in American politics.

MFT was created for research in cultural psychology, not political psychology. Haidt and Joseph (2004) focused on variation in virtue concepts across cultures and eras. The list of foundations was not reverse-engineered from known differences between American liberals and conservatives. Yet the theory mapped on closely and easily to the two sides of the culture war described by Buchanan, and by Hunter (1991). These were the first empirical findings produced with MFT (Haidt & Graham, 2007), and these are the findings for which the theory is best known today, so we begin our review with them.

### 3.2.1 Ideology in Five Dimensions

MFT's deepest roots are in the work of Richard Shweder, who showed that the moral domain is broader in India than among educated respondents in the United States (Shweder, Mahapatra & Miller, 1987). Now that we have the terminology of Henrich, Heine, and Norenzayan (2010), we can say that the moral domain in WEIRD cultures (Western, Educated, Industrialized, Rich, and Democratic) is fairly narrow, comparatively speaking. It focuses on protecting individuals from harm and exploitation. In most traditional societies, however, the moral domain is broader, including concerns about protecting groups, institutions, traditions, and the moral order more generally. Haidt, Koller, and Dias (1993) confirmed Shweder's basic finding, and showed that it holds across social classes in the United States and Brazil: richer people have a narrower moral domain. Haidt and Hersh (2001) provided the first evidence that Shweder's basic argument applied across the political spectrum too: in a small sample of college students who were interviewed about sexual morality, conservatives had a broader moral domain, making greater use of Shweder's ethics of community and divinity.

Shweder's three ethics translate directly into the five foundations (which were derived in part from those three ethics), leading Haidt and Graham (2007) to make the simple prediction that liberals would show greater reliance than conservatives upon the Care and Fairness foundations (which support the moral discourse of the ethics of autonomy), whereas conservatives would show greater reliance upon the Loyalty and Authority foundations (which support Shweder's ethic of community) and the Sanctity foundation (Shweder's ethic of divinity). To test this prediction, Graham, Haidt, and Nosek (2009) created an early draft of the Relevance scale (see Table 2; see also Graham et al., 2011). They found support for their prediction, and this basic pattern has been found in many subsequent studies, using many different methods (see Figure 1). Haidt and Graham (2007) suggested that MFT could help to explain many aspects of the culture war, including the specific issues that become battlefields, the intractability of the debates, and the inability of the two sides to even understand each other (because their moral visions were based on deep differences – differences in the very foundations upon which moral arguments could rest). Consistent with the intuitionist tradition, arguments about culture-war issues such as gay marriage, abortion, art, and welfare spending should not be expected to influence or convince people on the other side, because attitudes about specific issues are based on deep intuitions, not on the specific reasons put forth during a debate.

A number of studies using a variety of different methods and samples, conducted by several different research groups, have now replicated that first empirical finding. Graham, Haidt, and Nosek (2009), for example, used four different methods and consistently found that liberals valued Care and Fairness more than did conservatives, whereas conservatives valued Loyalty, Authority, and Sanctity more than did liberals (See Figure 1). Using a simple self-report political orientation scale (*very liberal* to *very conservative*) and examining large internet samples, Graham et al. (2009) show this pattern in explicit judgments of moral relevance (upper left panel, Figure 1), agreement with foundation-relevant moral statements (upper right panel, Figure 1), and willingness to engage in foundation-related “taboo” acts for money (bottom panel, Figure 1). In each case, care and fairness are valued highly across the political spectrum, with liberals on average endorsing them slightly more than conservatives. Loyalty, Authority, and Sanctity, in contrast, show a clear linear increase in importance moving from extreme liberals to extreme conservatives. In a fourth study, Graham et al. (2009) found the same pattern of liberal-conservative differences comparing the frequency of foundation-related words used in the sermons of liberal and conservative churches (see Table 2).



Additional evidence of the robustness of this basic pattern of foundation differences is reported by Graham, Nosek, and Haidt (in press), who obtained the same results in a representative sample of U.S. citizens. Graham et al (2011) have also replicated this ideological pattern using respondents at YourMorals.org from 11 different world regions (see section 3.4 and Table 3).

Finally, McAdams et al. (2008) conducted life narrative interviews with a group of highly religious and politically-engaged adults, and coded their responses for themes related to the five moral foundations. They found what they characterized as “strong support” for MFT:

When asked to describe in detail the nature and development of their own religious and moral beliefs, conservatives and liberals engaged in dramatically different forms of moral discourse. Whereas conservatives spoke in moving terms about respecting authority and order, showing deep loyalty to family and country, and working hard to keep the self pure and good, liberals invested just as much emotion in describing their commitments to relieve the suffering of others and their concerns for fairness, justice, and equality. (McAdams et al., 2008, p. 987).

### 3.2.2 Personality-Morality-Ideology Linkages

MFT views individual and group differences in reliance on the various moral foundations as emerging from the interactions of differences in biology, cultural socialization and individual experience (see Haidt, 2012, ch. 12). A useful framework for conceptualizing these interactions is McAdams’ three-level model of personality (1995; McAdams & Pals, 2006). At Level 1 are *dispositional traits* such as the Big 5. These are global, decontextualized traits that describe broad patterns of cognitive or emotional responding. At Level 2 are what McAdams calls *characteristic adaptations*, including values, goals, and moral strivings that are often reactions (or adaptations) to the contexts and challenges an individual encounters. Characteristic adaptations are therefore more conditional and domain-specific than dispositional traits, and are thus more variable across life stages and situational contexts. Finally, at Level 3 in McAdams’ framework are *integrative life stories*—the personal narratives that people construct to make sense of their values and beliefs. For many people these life stories include an account of the development of their current moral beliefs and political ideology. Haidt, Graham, and Joseph (2009) elaborated McAdams’ third level for work in political psychology by pointing out that many such stories are not fully self-authored, but rather are often “borrowed” from ideological narratives and stereotypes commonly held in the culture.

We view the moral and personality traits measured by our various methods (as summarized in Table 2) as Level 2 characteristic adaptations, linked closely to particular dispositional traits (Level 1). We can’t measure moral foundations directly – we can’t see the “first draft” of the moral mind. All we can do is read the finished books and quantify the differences among individuals and groups. All we can do is measure the morality of a person and quantify the degree to which that person’s morality is based on each foundation. (We sometimes say that a person scored high on a particular foundation, but that’s a shorthand way of saying that their morality, as we measure it, relies heavily on virtues and concepts related to that foundation.) An individual’s morality is constructed as they grow up in a particular culture, with particular life experiences. But two siblings who bring different dispositional traits to otherwise similar contexts and experiences will develop different moral and political characteristic adaptations. As young adults, they will then find different ideological narratives compelling, and may come to self-identify as members of different political parties.

For example, substantial evidence suggests that political conservatism is associated with personality characteristics that incline individuals toward a general resistance to novelty and change. In a comprehensive meta-analysis of the psychological correlates of conservatism, Jost, Glaser, Kruglanski, and Sulloway (2003) found that, compared to liberals, conservatives have higher needs for order, structure, and closure, and report lower levels of openness to experience. Conservatives have been found to respond less positively to novel stimuli at physiological and attentional levels as well (Amodio et al., 2007; Oxley et al., 2008; Hibbing & Smith, 2007; Shook & Fazio, 2009). Similarly, a growing body of literature has revealed a relation between greater political conservatism and heightened levels of disgust sensitivity (Inbar, Pizarro, & Bloom, 2009; Helzer & Pizarro, 2011; Inbar, Pizarro, Iyer, & Haidt, in press; Dodd, Balzer, Jacobs, Gruszczyński, Smith, & Hibbing, 2012; Smith, Oxley, Hibbing, Alford, & Hibbing, 2011). Together, this constellation of dispositional tendencies may provide the emotional infrastructure underlying conservative reverence for long-established institutions and highly structured systems of social hierarchy and sexual propriety. Conversely, individuals with lower need for structure, greater openness to experience, and dampened disgust sensitivity should be less anxious about challenging traditional authority structures, lifestyle, and sexual practices. These dispositional tendencies may in turn afford greater attraction to liberal policy positions seeking to “reform” traditional values and institutions to reflect greater equality for historically oppressed social groups and a less restrictive view of sexual purity and moral contamination more generally.

Providing empirical support for the causal connections between personality characteristics, moral concerns, and political ideology is a challenging task, and more research in this area is clearly needed. A small set of studies, however, have directly examined these types of associations. Lewis and Bates (2011) measured the Big Five personality traits, moral foundations, and political ideology and found that higher scores on Care-Fairness were related to greater openness, neuroticism, and agreeableness, and that higher Loyalty-Authority-Sanctity scores were associated with greater conscientiousness and extraversion, and lower levels of neuroticism. Importantly, and consistent with McAdams’ three-level personality model, moral foundation endorsements mediated the relationship between Big Five traits and political ideology.

In a similar study, Hirsh, DeYoung, Xu, and Peterson (2010) used a more fine-grained measure of the Big Five personality traits that separates each trait into two separate “aspects” (DeYoung, Quilty, & Peterson, 2007). Like Lewis and Bates (2011), they found an overall measure of agreeableness to be a significant predictor of greater endorsement of the Care and Fairness foundations, but that when examined at the level of aspects, this relation was limited to the aspect of agreeableness they term compassion. The other aspect of agreeableness, politeness, was not related to Care-Fairness scores, but was in fact predictive of higher scores on the Authority foundation. Also, where Lewis and Bates (2011) found openness to be a significant predictor of Care-Fairness, Hirsh et al. (2010) found no significant relation, but they did find a negative relation between openness (particularly the intellect aspect) and the Authority and Sanctity foundations. Hirsh et al. (2010) also found an association between greater overall conscientiousness and endorsement of Loyalty, Authority, and Sanctity foundations, but these relations were driven only by the orderliness (not the industriousness) aspect of that trait. Subtle differences between these and the Lewis and Bates (2011) findings notwithstanding, the Hirsh et al. (2010) findings are consistent with the general thrust of MFT, and their study again provides evidence that moral foundation endorsements mediated the relationships between personality factors and political ideology.

Finally, in an attempt to integrate research on conservative sensitivity to threatening stimuli with MFT, Van Leeuwen and Park (2009) examined whether a conservative pattern of moral foundation endorsement mediated the relationship between perceived social dangers and political conservatism. They found that the tendency to emphasize Loyalty, Authority, and Sanctity over Care and Fairness was related to both explicit and implicit conservatism in the expected directions, and that it also partially mediated the relationship between Belief in a Dangerous World and conservatism. The authors argue that these results suggest that a basic inclination to perceive the environment as dangerous may lead to greater endorsement of the Loyalty, Authority, and Sanctity foundations, perhaps due to the perceived protection these group-oriented values seem to provide.

### **3.2.3 Political Stereotypes and Interpersonal Judgment**

It has often been said that politics is perception. Do the relations between moral foundations and political ideology have implications for how people perceive and make judgments about groups and individuals? Do people recognize the moral differences between liberals and conservatives? Do liberal and conservative moral profiles predict what characteristics they will view favorably in others?

Graham et al. (2011) addressed some of these questions by examining whether people favored or disfavored members of social groups that were conceptually related to the five moral foundations. The researchers began by categorizing 27 social groups according to the moral foundations they exemplify, for example, ACLU members (Fairness), police officers (Authority), and virgins (Sanctity). They found that participants' attitudes toward these groups were predicted most strongly by their endorsement of the corresponding moral foundations, even when controlling for political ideology. In other words, knowing a person's MFQ scores gives you important information, over and above their ideology, about their social group prejudices. These results speak to the tight relationship between social and moral judgment, while also demonstrating the predictive and discriminant validity of the five foundations.

Graham, Nosek, and Haidt (in press) used MFT to examine the moral stereotypes liberals and conservatives hold about each other. Participants filled out the MFQ either normally, or else as a "typical" liberal, or else as a "typical" conservative. Overall, participants correctly simulated the general liberal-conservative pattern predicted by MFT. That is, the typical liberal scores were higher than the typical conservative scores on Care and Fairness, and the typical conservative scores were higher than the typical liberal scores on Loyalty, Authority, and Sanctity. However, participants' estimations of these differences were exaggerated. In fact, the differences in moral foundation scores that participants reported for the typical liberal and the typical conservative were significantly larger than the actual differences observed between even the most extreme partisans. Although participants who identified as liberals, moderates, and conservatives all exaggerated these stereotypes, they did so to varying degrees. Liberals, more than conservatives and moderates, reported the most exaggerated stereotypes of political partisans when estimating all five foundations. Most importantly, conservatives tended to be relatively accurate in their beliefs about how much liberals valued Care and Fairness, but liberals estimated that conservatives valued these foundations far less than they actually did. MFT's pluralistic approach thus allows not only for a better understanding of the moral differences between liberals and conservatives, but also for a more nuanced understanding of the moral stereotypes that contribute to the seemingly intractable nature of partisan conflict.

In terms of judgments of individuals rather than groups, Frimer and colleagues (in press)

asked two groups of respondents (professors solicited from liberal and conservative colleges and visitors to Mechanical Turk) to evaluate the extent to which 40 of the most influential people of the 20th century were “moral exemplars.” A moral exemplar was defined simply as “a highly moral person.” The target individuals had previously been rated by a separate sample of social science professors as to how much each individual embodied each of the five moral foundations. The results were generally quite consistent with the predictions of MFT, although subtle and important differences did emerge. Overall, there was substantial agreement across the ideological spectrum on what led an individual to be perceived as virtuous, with both liberal and conservative respondents relying most heavily in their moral evaluations on the targets’ embodiment of the Care, Fairness, and Sanctity foundations. Ideological agreement regarding the moral importance of Care and Fairness follows directly from MFT, but the importance liberals placed on the Sanctity foundation is more surprising (although it was only a significant predictor of liberal moral evaluations in the academic sample). Also consistent with MFT, liberals were more likely than conservatives to favor those individuals who espoused virtues related to the Care and Fairness foundations, while conservatives were more likely than liberals to favor those who personified virtues related to Authority and Sanctity. The most important divergence from the predictions of MFT was that target individuals’ embodiment of Loyalty and Authority had no significant effect on judgments of virtuousness, even for conservative respondents. In fact, Authority was actually found to be a negative predictor of liberals’ moral evaluations, suggesting that those on the political left may perceive the embodiment of Authority as more vice than a virtue. Frimer et al (in press) conclude from their findings that Care, Fairness, and Sanctity are core foundations of moral evaluation but that Loyalty and Authority may play more complicated, interactive roles that need to be unpacked by future research.

MFT has even been useful in understanding preferences for individual political candidates. Iyer, Graham, Koleva, Ditto, and Haidt (2010) compared two similar groups of Democrats during the 2008 Democratic Primary: supporters of Hillary Clinton and supporters of Barack Obama. Although both candidates were Democrats with only subtle policy differences, their supporters differed on several individual difference measures (psychopathic personality, moral relativism, empathy, and global concern for others). Most importantly, endorsement of the moral foundations predicted which candidate participants were more likely to favor, even when controlling for age, gender, education, and self-placement on the liberal-conservative dimension. Specifically, Clinton supporters showed a more “conservative” profile of moral foundation endorsement, as greater endorsement of both the Loyalty and Authority foundations predicted Clinton favorability. Relative favorability towards Obama, on the other hand, was predicted by greater endorsement of the Fairness foundation. This pattern makes sense given that Clinton polled better with the relatively conservative white working class.

### 3.2.4 Beyond Liberal and Conservative

The research discussed thus far describes ideological differences in reliance on the moral foundations as if all individuals fit neatly along a single liberal-conservative continuum. However, political ideology is a complex and multifaceted construct that can be understood along multiple dimensions. For example, one popular method is to distinguish economic and social political preferences (Conover & Feldman, 1981; Duckitt, 2001; Weber & Federico, 2007; McClosky & Zaller, 1984). MFT offers the opportunity to create a five-dimensional space, and then to examine whether people tend to cluster into some regions and not others.

Haidt, Graham, and Joseph (2009) investigated this possibility by performing a cluster

analysis of over 20,000 American participants who completed the Moral Foundations Questionnaire. They found support for a four-cluster solution, which identified four groups with distinct moral profiles (see Figure 2). The first group, labeled “secular liberals,” scored the highest on Care and Fairness, and they scored very low on Loyalty, Authority, and Sanctity. This group also scored the highest on Openness to Experience and lowest on Right Wing Authoritarianism and Social Dominance Orientation, a pattern that typically exemplifies American liberalism. They also reported high levels of atheism. By contrast, the group labeled “social conservatives” showed a nearly opposite profile of results. Social conservatives were lowest on the Care and Fairness foundations and very high on the other three. They were low on Openness to Experience, high on both Right Wing Authoritarianism and Social Dominance Orientation, and they reported the most frequent religious attendance.

These two clusters offered no surprises. They are just what you’d expect from our common stereotypes about liberals and conservatives, and from the findings of Graham et al. (2009) and Jost et al. (2003). However, the other two groups were different. The third group, dubbed “libertarians,” scored low on all five moral foundations, and they tended to highly value hedonism and self-direction on the Schwartz Values Scale (Schwartz & Bilsky, 1990), and they showed high levels of atheism. The fourth group, labeled “religious left,” scored relatively high on all five foundations, on religious participation, and on the Schwartz values of benevolence, tradition, conformity, security, and spirituality. Importantly, neither the libertarians nor the religious left fits neatly into the categories of “liberal” or “conservative,” but their unique moral and psychological identity was detectable when their moralities were analyzed using the five scores of the MFQ. The left-right dimension is indeed useful as a first pass (Jost, 2006). But the pluralism of MFT gives us greater resolution and detects groups that don’t fit well on that one dimension.

In a similar vein, Weber and Federico (in press) used a mixed model latent class analysis to argue for a more heterogeneous approach to understanding political ideology after identifying six discrete ideological groups (consistent liberals, libertarians, social conservatives, moderates, consistent conservatives, inconsistent liberals). They found each group to have unique sets of economic and social policy preferences that were reflected in distinct patterns of moral foundation endorsement. Further, Care and Fairness concerns were most related to an ideological preference dimension of equality-inequality, while Loyalty, Authority and Sanctity were most aligned with the ideological preference dimension of openness-conformity (Federico, Weber, Ergun, & Hunt, in press).

The most extensive examination of an ideological subgroup that cannot be easily placed along a simple liberal-conservative dimension is the work of Iyer, Koleva, Graham, Ditto, and Haidt (2012) that set out to identify the cognitive, affective, and moral characteristics of self-identified libertarians. Libertarians are an increasingly influential group in American politics, with their ideological positions gaining attention through the popularity of the Tea Party movement and media coverage of the Presidential campaign of Congressman Ron Paul (R-TX). Libertarian values, however, presented a challenge for MFT, as the primary value that libertarians espouse – individual liberty – was not well captured by the existing five foundations. Indeed, the original conception of MFT (Haidt & Joseph, 2004) took Shweder’s ethic of autonomy and created foundations that represented the liberal vision of positive liberty, where individual freedom is defined by opportunity, rather than the libertarian vision of negative liberty, where individual freedom is defined by a lack of obstruction (see Berlin, 1969, for a broader discussion of negative versus positive liberty).

Iyer et al. (2012) compared a large sample of self-identified libertarians (N = 11,994) to self-identified liberals and conservatives across dozens of measures, looking in particular at measures that would shed light on the moral values of Libertarians. They also created a set of MFQ-like items designed specifically to measure endorsement of liberty as a moral value. In the cluster analysis described above (from Haidt, Graham, & Joseph, 2009), the cluster containing the largest number of libertarians reported relatively weak endorsement on all five foundation subscales of the MFQ. Iyer et al similarly found that self-described libertarians showed relatively weak endorsement of all five foundations; both the relatively weaker endorsement of Care and Fairness concerns typical of conservatives, as well as the relatively weaker endorsement of Loyalty, Authority, and Sanctity concerns typical of liberals (see Figure 3).

Does that mean that libertarians have no morality – or, at least, less concern with moral issues than liberals or conservatives? Or might it be that their core moral value was simply not represented among the five foundations measured by the MFQ? Consistent with the latter position, when Iyer et al examined the items tapping the value placed on liberty as a moral concern, they found that libertarians did indeed score higher than both liberals and conservatives. This relative valuation of liberty was found both on items tapping concerns about economic and property-related freedoms (typically valued by political conservatives more than liberals) as well as lifestyle freedoms (typically valued by political liberals more than conservatives). Similar findings emerged from the Good Self measure (Barriga, Morrison, Liao, & Gibbs, 2001), where libertarians reported valuing being independent more than other groups, as well as from the Schwartz Values Scale (Schwartz, 1992), on which libertarians scored highest of all groups on valuing self-direction.

Iyer et al also identified a number of other interesting psychological characteristics of their libertarian sample. Perhaps reflecting the emotional underpinnings of their focus on individual liberty, libertarians scored higher than liberals or conservatives on a scale measuring psychological reactance (Hong, 1996; Brehm & Brehm, 1981). Libertarians also showed a relatively cerebral as opposed to emotional cognitive style (e.g., high in need for cognition, low empathizing and high systematizing [Baron-Cohen, 2009]) and lower interdependence and social relatedness (e.g., low collectivism, low on all three subscales of the Identification with All of Humanity Scale).

Together, these findings paint a consistent portrait of the moral psychology of libertarianism. Libertarians—true to their own descriptions of themselves—value reason over emotion, and show more autonomy and less interdependence. Their central moral value, therefore, is one that grants people the right to be left alone. MFT’s five moral foundations appeared to be inadequate in capturing libertarians’ moral concerns, but the approach that gave birth to these foundations served us well in examining this new group, and stimulated us to consider Liberty/oppression as a candidate for addition to our list of foundations (see section 4.1.5, and further discussion in Haidt, 2012 ch. 8).

### **3.3 Moral Foundations and Other Psychological Constructs**

One of the liveliest areas of current moral psychology is the intersection of moral judgment, attitudes, and emotion. Many of these studies have also focused on ideology variables, either as predictors or as outcomes of situational and individual variation in moral intuition and emotion.

### 3.3.1 Attitudes

Koleva, Graham, Iyer, Ditto, and Haidt (2012) illustrated the utility of MFT's pluralistic framework for understanding the psychological underpinnings of specific policy issues. In two web studies ( $N = 24,739$ ) we used scores on the MFQ to predict moral disapproval and attitude stands on 20 hot-button issues, such as same-sex marriage, abortion, torture, and flag desecration/protection. We found that MFQ scores predicted attitudes on these issues, even after partialling out participants' ideology, gender, religiosity, and other demographic variables. We expected that the foundations would predict variation based on overlapping content – for example, people who scored high on the care foundation would be particularly sensitive to issues involving violence or cruelty, and this was indeed true in general. But unexpectedly, the Sanctity foundation emerged as the strongest predicting foundation for most issues. For example, people who score high on the loyalty foundation tend to be more patriotic, and therefore more strongly in favor of “protecting” the flag from desecration, but scores on the Sanctity foundation were even more predictive. Some people see the flag as merely a piece of cloth; others see it as a sacred object, containing a non-material essence that must be protected. These findings about the importance of Sanctity in ongoing political controversies could not have been obtained using moral theories that limited the moral domain to issues of Care or Fairness.

Another advantage of using a multifaceted approach like MFT is that it helps us understand how a person could hold different attitudes across issues that appear to engender similar moral concerns. For example, even though abortion, euthanasia, and the death penalty all evoke arguments for the sanctity of life, opposition to the first two was best predicted by Sanctity, whereas opposition to the third was best predicted by Care scores. This may help explain why liberals, who score low on Sanctity concerns (Haidt & Graham, 2007; Graham, et al., 2009), do not generally oppose access to abortion and euthanasia, but do tend to oppose the death penalty.

Aside from refining our understanding of ideological opinions, these findings suggest novel approaches to persuasion and attitude change. For example, Feinberg and Willer (in press) showed that framing messages about the environment in terms of Sanctity, rather than just Care, increased conservatives' support for environmental policies, presumably because this framing triggers intuitions which resonate with conservatives.

### 3.3.2 Emotion

In a related line of inquiry, researchers have examined the interplay between morality and emotion, particularly the emotion of disgust, in shaping moral judgments and ideological attitudes and self-identification. Much of this work has either explicitly drawn on MFT or offers indirect evidence that supports its premises. For example, Horberg, Oveis, Keltner, and Cohen (2009) showed that an individual's trait propensity towards feeling disgust, an emotion that is related to Sanctity concerns (Rozin, Haidt & McCauley, 2008), as well as experimental inductions of disgust, intensified the moral importance of maintaining physical and spiritual purity. This effect was specific: other emotions, such as trait or state anger, fear, or sadness did not have an effect on judgments related to purity, and disgust did not affect non-purity moral judgments, such as Care/harm or justice. Finally, Preston and Ritter (in press) showed that the concepts of religion and cleanliness are linked such that priming religion increased the mental accessibility of cleanliness-related concepts and the desirability of cleaning products, whereas priming thoughts of personal cleanliness increased ratings of the value ascribed to religious

beliefs. This work underscores the relevance of experiences with and concerns about the Sanctity foundation to moral judgment.

Building on the finding that conservatives tend to moralize Sanctity concerns more than liberals (Graham, et al., 2009), Helzer and Pizarro (2011) reported two experiments in which subtle reminders of physical purity – standing by a hand sanitizer and using hand wipes – led participants to report being more politically conservative and more disapproving of sexual purity violations, like incest or masturbation. Similarly, Inbar, Pizarro, and Bloom (2011) found that experimental inductions of disgust led participants to report more negative attitudes towards gay men but not towards lesbians or other outgroups. However, unlike Helzer and Pizarro (2011), these researchers did not find a general effect of disgust on political attitudes or on self-identification. Finally, Inbar, Pizarro, Iyer, and Haidt (2011) showed that self-identified conservatives, both in the U.S. and around the world, reported greater propensity towards feeling contamination disgust, and that disgust sensitivity predicted voting patterns in the U.S. Interestingly, Jarudi (2009) found that conservatives were more sensitive to purity concerns about sex (e.g. anal sex), but not about food (e.g. eating fast food), even though disgust sensitivity was related to disapproval in both domains.

Finally, several studies have examined the role of anger and contempt, in addition to disgust, in response to foundation-related violations. For example, Russell and Giner-Sorolla (2011) gave participants scenarios that depicted violations of Care, Fairness, or Sanctity and assessed their moral judgments, anger, and disgust. Next, participants were asked to generate circumstances that could change their opinion and then to re-evaluate the scenarios assuming these new circumstances. Whereas ratings of disgust did not change during re-evaluation, anger for the harm and fairness violations was decreased and this change in emotion predicted a change in moral judgment. These findings suggest that moral anger is a more flexible emotion that is sensitive to context, whereas moral disgust appears to be less so.

Another study on the role of anger, contempt, and disgust in moral judgment showed that even violations outside the Sanctity domain elicit strong moral disgust, suggesting a domain-general function for this emotion (Hutcherson & Gross, 2011). However, the authors also found that moral violations that entail direct harm to the self elicited anger more than disgust, and that contempt was the strongest emotional response to non-moral violations attributed to low competence. Taken together, these studies suggest that anger and disgust are common responses to moral transgressions, but that anger is, in a sense, more open to reason and revision based on new information.

The ability to contrast multiple moral emotions, operating with respect to multiple sets of moral issues, is greatly enhanced by the pluralism and intuitionism of MFT, compared to moral theories that are either monist or rationalist. Among the most important discoveries has been the powerful and until-recently underappreciated role of disgust, and related intuitions about sanctity, in moral judgment, political attitudes, and even voting behavior.

### 3.3.3 Moral Character

In addition to attitudes and emotion, several studies have now explored the moral foundations' association with variables related to moral character. For example, in a large community sample, individuals scoring higher on non-diagnostic psychopathic trait measures indicated less concerns about care and fairness as measured by the MFQ (Glenn, Iyer, Graham, Koleva, & Haidt, 2009); these associations were mediated by their weaker empathic concern. Psychopathic personality was also associated with high endorsement of loyalty, which was



mediated by greater social dominance orientation. However, when morality was measured with the Sacredness Scale (see Table 2) those high in psychopathic personality indicated greater willingness to violate *all five* foundations for money, suggesting that such individuals might be generally similar to those low in psychopathy in terms of their abstract moral evaluations, but are more willing to (hypothetically) violate moral concerns in exchange for a salient reward like money.

Finally, the moral foundations have been examined not just in relation to participants' own character, but also in relation to how they infer others' traits. Specifically, Clifford (2012) found that foundation scores predicted relevant traits' accessibility when describing politicians (e.g. Care scores were positively related to how many Harm-related traits, e.g. kind and compassionate, were brought to mind). Furthermore, a politician's known position on an issue interacted with the individuals' own position in affecting trait-inferences related to the foundation assumed to motivate the position (e.g. if one opposes the death penalty due to strong harm concerns, one will rate a politician who supports the death penalty as low on harm traits).

### 3.3.4 Other Psychological Constructs

In addition to these studies of attitudes, emotion, and character, researchers have used MFT to examine a variety of other constructs. For example, using an evolutionary framework, Kurzban and colleagues found that participants' Care and Sanctity scores related to their opposition to recreational drug use, which the authors argue to be driven by unrestricted views on sexuality (Kurzban, Duker, & Weeden, 2010).

Similarly, building on an evolutionary perspective of sports fandom as a by-product of adaptations that evolved in the context of small-scale warfare, Winegard and Deaner (2010) found that participants' endorsement of moral concerns about group loyalty predicted the extent to which they identified with their favorite sports team. Furthermore, men scored higher on Loyalty than women, and this difference partially explained their higher sports fandom compared to women's.

Within the field of communication, Tamborini and colleagues recently examined the link between moral sensitivity to harm and fairness and perceptions and appeal of violent media (Tamborini, Eden, Bowman, Grizzard, & Lachlan, 2012). Their findings indicated that higher Care scores predicted perceptions of a film narrative that contained gratuitous violence as more graphic and less appealing. Similarly, higher Fairness scores predicted greater appeal of a film narrative that contained strong justification of violence.

Finally, a recent study in environmental and agricultural ethics explored individuals' free associations with the phrases "ethical/morally right food" and "unethical/morally wrong food" and categorized them as relating to the moral foundations (Makiniemi, Pirttila-Backman, & Pieri, in press). Results suggested that the free associations were dominated by concepts related to the Care and Sanctity foundations, followed by Fairness, Loyalty, and Authority concerns.

### 3.4 Cross-cultural differences and intergroup relations

For any theory that claims to be rooted in human nature, the theory must be tested in diverse samples and across different cultures. While MFT is in its toddlerhood, great progress has already been made in examining the moral foundations in several cultures. In this section we describe how MFT has been used so far to investigate cultural differences in morality and phenomena related to intergroup relations more generally.

### 3.4.1 East-West Cultural Differences

Using a large international sample, Graham and colleagues (2011) showed that the moral foundations model was a good fit to the data across world regions. They also found that even after controlling for various demographic variables, world region was a significant predictor of foundation-related concerns. Specifically, participants in Eastern cultures (South Asia, East Asia, and Southeast Asia) expressed slightly greater Loyalty- and Sanctity-related moral concerns than did participants in Western cultures (United States, Canada, United Kingdom, and other Western European countries), which are consistent with established cultural differences in collectivism (Triandis, 1995) and the role of purity concerns in daily life and religious practices (Shweder et al., 1997). Furthermore, compared to the liberal vs. conservative differences in the U.S., these cross-cultural differences were small – consistent with the theory that variation within cultures exceeds variation between cultures (e.g., Vauclair & Fischer, 2011).

However, the findings of Graham et al. (2011) come with two important caveats: they are based on data collected in English, and from participants with access to the internet. Thus these findings likely come from Westernized (or even WEIRD) segments of these populations. Fortunately, researchers in dozens of countries have been translating and back-translating moral foundations measures (translations available at MoralFoundations.org), and collecting data with them among native speakers in various countries (see e.g. Bobbio, Nencini, & Sarrica, 2011; Kim, Kang, & Yun, in press; van Leeuwen & Park, 2009). For example, as of June 2012 the MFQ has been translated into Amharic, Arabic, Bahasa Indonesian, Bengali, Chinese (mainland and Hong Kong), Croatian, Dutch, Farsi (Persian), French, French Canadian, German, Greek, Hebrew, Hindi, Indonesian, Italian, Japanese, Kiswahili, Korean, Malay, Nepali, Polish, Portuguese, Portuguese (Brazilian), Romanian, Russian, Spanish, Spanish-Castilian, Swedish, Tagalog, Thai, Turkish, Ukrainian, Urdu, Vietnamese, and Yoruba (Nigeria). Field work in Nicaragua has enabled foundation measurement among social groups who do not typically speak English or have access to the internet (migrant field workers, residents of the Managua city dump, sex workers), and further field work is planned for India, Iran, Morocco, and Lebanon. This work will be crucial not only for detecting cross-cultural differences in reliance on the foundations, but for exploring within-culture variation as well.

### 3.4.2 Other Cultural Differences

Within both Eastern and Western cultures there are a number of consistent patterns of moral concerns. Evolutionary perspectives (e.g., Neuberg, Kenrick, & Schaller, 2010; Van Vugt & Park, 2009) note that pathogens are among the principle existential threats to organisms, so those who could best avoid pathogens would have enhanced evolutionary fitness. Van Vugt and Park contend that human groups develop unique practices for reducing pathogen exposure – particularly in how they prepare their foods and maintain their hygiene. When groups are exposed to the practices of a foreign culture, they may perceive its members as especially likely to carry pathogens that may contaminate one's ingroup. This contamination fear may lead people to place greater emphasis on Sanctity, which Haidt and Joseph (2007) describe as originating in an anti-pathogen system. In a recent analysis, van Leeuwen, Park, Koenig, and Graham (in press) demonstrated that historical pathogen prevalence at the country level is a significant predictor not only of Sanctity, but of Loyalty and Authority as well. Specifically, there was a positive relationship between country-level historical pathogen prevalence and individual-level endorsement of these three foundations, even after statistically controlling for national gross

domestic product (GDP) and individual demographic variables. In addition to providing support for the proposal that Sanctity intuitions are related to mechanisms for combating pathogenic disease, these findings also suggest that pathogen prevalence and contamination fears may enhance group cohesion, collectivism, and adherence to group norms, as a means to minimize the contamination threat.

The threat that other cultures engender may lead to individual differences within Eastern and Western cultures, too. Van Leeuwen and Park (2009) conducted a study showing that perceiving the social world as dangerous predicts increased adherence to the binding moral foundations, which predicts increased conservatism on measures of political orientation. Furthermore, adherence to these binding foundations mediated the relationship between perceptions of a dangerous world and political conservatism.

The link between moral foundations and political orientation appears robust across cultures (see Table 3). Graham and colleagues (2011) found that across the 12 world regions for which data were available, liberals consistently valued Care and Fairness concerns more than conservatives, whereas conservatives consistently valued Loyalty, Authority, and Sanctity more than liberals. Further, they found that these political orientation patterns were robust across national and cultural contexts, both in terms of direction (i.e. a negative relationship between conservatism and valuation of Care and Fairness, and a positive relationship between conservatism and Loyalty, Authority, and Sanctity) and magnitude (i.e. correlations were consistently stronger for Authority and Sanctity, and weakest for Care). Van Leeuwen and Park (2009) found a similar relationship between moral foundations and political orientation among a sample of Dutch participants, providing further evidence of the robustness of this pattern across national and cultural contexts. The relative predictive power of the Authority and Sanctity foundations across studies and across cultures suggests that the most intractable of political conflicts are particularly likely to involve disagreements about respect for tradition, authority, and spiritual purity.

### 3.4.3 Intergroup Relations

Moral differences often lead to poor intergroup relations. Kesebir and Pyszczynski (2011; see also Motyl & Pyszczynski, 2009; Pyszczynski, Motyl, & Abdollahi, 2009) argue that the mere awareness of groups with different moral intuitions and worldviews is existentially threatening and may engender hostility and violence. In one series of studies, McGregor et al. (1998) showed that when participants read a passage that disparaged their political views as amoral and sickening, they were more aggressive and administered more hot sauce to their critic. In related work, Rosenblatt, Greenberg, Solomon, Pyszczynski, and Lyon (1989) found that threats led to increased punitiveness toward moral transgressors. The punishments were greatest when judging prostitutes, whose behavior violates the Sanctity foundation. Together, these findings suggest that disagreement on moral intuitions are especially likely to lead to increased intergroup aggression and conflict.

Intergroup moral conflicts are particularly intractable. As the moral issues at the core of these conflicts are rooted in different intuitions, people on opposing sides of these conflicts simply do not understand how anyone can hold different moral intuitions (Ditto & Koleva, 2011). This empathy gap leads people to view adherents of different moral worldviews as less warm than adherents of similar moral worldviews (Bruneau, Dufour, & Saxe, 2012). Similarly, this empathy gap for moral and political adversaries can make intergroup violence more likely,

as adversaries can more easily view each other as not deserving moral rights (Waytz et al., 2010).

### 3.5 Implicit moral cognition

Although MFT is at base a theory about the intuitive nature of moral concerns, work on the implicit processes involved in foundation-related judgments is still in its infancy. Nevertheless, MFT's pluralist approach can provide a theoretical framework to organize and explain disparate findings connecting morality and automaticity. As discussed in section 3.3.2, incidental disgust (caused by bad smells in the air or dirty surroundings) has been shown to increase the harshness of moral judgments (Schnall et al., 2008), even when the disgust was induced via hypnotic suggestion that participants could not consciously recall afterward (Wheatley & Haidt, 2005). Similarly, individual differences in disgust sensitivity have been shown to predict intuitive negativity toward gays, measured both by an Implicit Association Test and by intentionality assessments (Inbar, Pizarro, Knobe, & Bloom, 2009). Such unconscious effects are not only a blow to our collective self-image as rational moral decision-makers, they are also difficult to explain with monist theories treating all morality as harm (Gray et al., 2012), fairness (Baumard et al., in press), or innate rules of grammar (Mikhail, 2007). In this section we show examples of how MFT is beginning to shed light on automatic or otherwise nonconscious reactions to moral stimuli.

#### 3.5.1 Implicit morality across ideology

Just as the methods of implicit social cognition have begun to transform moral psychology, they have begun to transform the study of political ideology as well (Nosek, Graham, & Hawkins, 2010). In one of the first explorations of how ideology relates to explicit foundation endorsement, Graham et al. (2009; Study 2) used an implicit measure of ideological orientation: a liberal/conservative self-other Implicit Association Test. They found that while implicit ideology added little predictive power beyond explicit ideology in predicting abstract assessments of the moral relevance of foundation-related concerns, it did uniquely predict agreement with moral foundation judgment statements over and above explicit ideology. This finding suggests that while participants' self-reports of political ideology are sufficient to predict what they might *say* they consider morally relevant, their implicit political identities give you additional information about the foundation-related judgments they will actually make.

Given all the work on self-reported differences in foundation endorsement between liberals and conservatives (see Figure 1), a question naturally emerges: how "deep" do those differences go? Do people on opposite ends of the political spectrum have different automatic reactions to various moral cues, or do they experience the same intuitive moral reactions and differentially endorse them? A number of studies suggest that liberals' implicit reactions "look like" the endorsed opinions of conservatives, but that liberals then suppress or correct these first reactions when reporting their explicit opinions. For instance, Skitka, Mullen, Griffin, Hutchinson, and Chamberlain (2002) found that when liberals were tired, distracted, or under cognitive load, they showed levels of personal attributions such as victim-blaming akin to those of conservatives. The authors posited "motivated correction" as the process liberals undergo to bring these automatic reactions in line with their conscious egalitarian goals and values. Similarly, Eidelman, Crandall, Goodman, and Blanchard (2012) found that low-effort thought (induced by cognitive load, time pressure, or alcohol) increased aspects of conservatism such as acceptance of hierarchy and preference for the status quo.

Graham (2010) tested whether MFT could provide an organizing framework for such findings, with the hypothesis that liberals intuitively respond to Loyalty, Authority, and Sanctity cues more strongly than would be suggested by their explicitly-endorsed moral opinions. Using several implicit measures of reactions to foundation-related stimuli – evaluative priming, the AMP, and event-related brain potentials using EEG (see Table 2) – the authors found support for this hypothesis, and found no evidence of such implicit-explicit discrepancy for conservatives. Moreover, when randomly assigned to give their first “gut” reactions on the MFQ, participants across the political spectrum indicated that their answers were the same as their consciously-endorsed opinions, indicating that liberals are unaware of the discrepancy between their implicit and explicit moralities. In contrast to these studies, Wright and Baril (2011) found that cognitive load or ego depletion manipulations decreased MFQ endorsements of Loyalty, Authority, and Sanctity among conservatives. Although two large studies using different samples have failed to replicate this effect, more work needs to be done to test whether conservatives also have implicit-explicit discrepancies in their moralities, particularly for Care and Fairness concerns.

### 3.5.2 Psychophysiological pluralism

We do not expect that anyone will find five distinct and discrete patterns of physiological activity related to the five foundations. Foundations are not spots in the brain (see discussion of modularity in section 2.1), nor are they each identified by one specific physiological signature. Nonetheless, when you broaden the moral domain beyond Care and Fairness and you begin considering a broader range of moral intuitions, it stands to reason that you might find a broader range of central and peripheral psychophysiological responses to moral stimuli. Using facial EMG (see Table 2), Cannon, Schnall, and White (2011) showed that levator activity (disgust micro-expression) was highest for Sanctity violations and second highest for Fairness violations, while corrugator activity (angry micro-expression) was highest for violations of Care. Moreover, muscle activity differentially predicted severity of explicit moral judgments for different types of concerns, with disgust expressions predicting harsher Sanctity and Fairness judgments, anger expressions predicting harsher Care judgments, and smiling predicting less harsh Loyalty judgments.

In a vignette study contrasting judgments about Care (accidental vs. intentional assault) and Sanctity (accidental vs. intentional incest), Young and Saxe (2011) found that intentionality was central to the Care judgments, but was much less crucial for Sanctity judgments. They followed up this finding with an fMRI study and found that the right temporo-parietal junction (TPJ)—an area implicated in theory of mind reasoning, and hence intentionality judgments—was more involved in Care judgments than in Sanctity judgments.

Two other studies have looked for links between moral foundations and brain structures or responses. Lewis, Kanai, Bates, and Rees (2012) gave subjects the MFQ and then collected structural MRI brain scans. They found a variety of significant and interpretable relationships, including: 1. Scores on the Care and Fairness foundations (combined) were associated with larger gray-matter volume in the dorsomedial prefrontal cortex (DMPFC, an area associated with mentalizing and empathy); and 2. Sanctity scores were associated with more gray-matter volume in the left anterior insula (a region active in several moral emotions including disgust). They also found that high scores on the Authority and Purity foundations were associated with more gray matter volume in the subcallosal gyrus, although they did not know how to interpret this finding.

Parkinson, Sinnott-Armstrong, Koralus, Mendelovici, McGeer, and Wheatley (2011) wrote vignettes to trigger a range of moral intuitions, inspired partly by MFT, and then carried

out an fMRI study. They found that stories about people committing intentional physical harm preferentially activated regions associated with understanding and imagining actions; stories about sexual deviance preferentially activated many areas associated with affective processing (including the amygdalae and the anterior insula); and stories about dishonesty preferentially activated brain areas associated with reasoning about mental states (including the DMPFC and the TPJ). Their interpretation of these results was strongly supportive of the pluralist approach we emphasize in this paper:

These results provide empirical support for philosophical arguments against the existence of a functional or anatomical module common and peculiar to all moral judgments... Separate systems were found to characterize different kinds of moral judgment... It is likely that moral judgment is even more multidimensional than what is suggested here, given that there remain other domains of morality that were not examined in the current study (e.g., disrespect, betrayal of an in-group, fairness). These results suggest that, just as disparate systems are now understood to subserve aspects of cognitive faculties once thought to be monolithic (e.g., memory, attention), distinct systems subserve different types of moral judgment. Future research may benefit from working toward a taxonomy of these systems as Haidt and Graham (2007) have suggested (Parkinson et al., 2011, p. 3171).

In a massive measurement validation effort, Knutson, Krueger, Koenigs, Hawley, Escobedo, et al. (2009) compiled standardized ratings for 312 different moral vignettes to be used in behavioral neuroscience research; they explicitly referred to MFT's constructs to categorize vignettes and identify missing areas of moral concern. Although recent work has begun to distinguish the implicit processes involved in Care and Sanctity judgments, much more work is needed to investigate similarities and differences in implicit processing of Fairness, Loyalty, and Authority concerns. To anticipate the next section, we see future investigations of the implicit processes involved in foundation-related concerns, judgments, and reactions as a primary next step not just for MFT, but for moral psychology in general.

## 4 Future Directions

In this section we look toward the future of moral foundations research, with special attention paid to new areas of inquiry and the evolution of the theory itself. We begin by describing notable recent critiques of MFT, which we see as essential for helping to shape its future development. We then offer five criteria for foundationhood, to guide future discussions of what exactly the list of foundations should be, and what it would take to change or expand our current list. Finally, we give additional consideration to what will characterize the next several years of research in MFT and in moral psychology more generally.

### 4.1 Criticisms of the theory

Confirmation bias—the tendency to search only for supportive evidence—is powerful, and nobody has yet found a way to train people out of it (Lilienfeld, Ammirati, & Landfield, 2009). The best cure for the confirmation bias is other people—friends, colleagues, and opponents who don't share your biases, and who may even be motivated to find the disconfirming evidence that is sometimes hiding in plain sight. Scientists who create new theories would be well-advised, therefore, to seek out critics in order to improve their thinking.

Criticism is in fact so valuable that it's worth paying for. That, at least, was our thinking in 2007 when we offered the “moral foundations prize” – one thousand dollars to anyone who

could “demonstrate the existence of an additional foundation, or show that any of the current five foundations should be merged or eliminated.” The challenge was posted at MoralFoundations.org for two years. Nobody won the full prize, which required making a theoretical case and backing it up with empirical evidence, but three people or teams were awarded \$500 each for nominating strong candidates for “foundationhood” (we’ll discuss these candidates below).

In the years since 2007 we’ve been fortunate that many critics have stepped forward and volunteered to criticize MFT for free. These critics have helped us to overcome our confirmation bias, find flawed or under-specified parts of the theory, and make improvements. Most of the criticisms have been directed at one of the four basic claims we made in section 2: nativism, cultural development, intuitionism, and pluralism. We describe them in that order.

#### 4.1.1 Critiques of Nativism

Nobody in psychology today argues that the human mind is truly a “blank slate” at birth, but opinions range widely from minimalist positions, which say that there is hardly any writing on the “first draft” of the mind, to maximalist positions such as massive modularity (Sperber, 2005; Tooby, Cosmides, & Barrett, 2005), which say that the mind is to a great degree organized in advance of experience, including hundreds or thousands of functional modules. We are near the maximalist side of the spectrum, although, like Sperber (2005), we temper our nativism with extensive discussions of cultural development and variation (Haidt, 2012; Haidt & Joseph, 2007). The foundations are part of the first draft of the mind, but experience edits that draft extensively.

Critics of nativism tend to be critics of MFT. Suhler and Churchland (2011) argued that all nativist theories must clear a very high bar to be taken seriously. To be more than mere “hand waving,” they must “be supported by, or at least consistent with,” evidence from genetics, neurobiology, and developmental psychology. (See also Narvaez, 2008, who asked for physiological evidence of modules and asserted that subcortical brain areas include modules but the cerebral cortex does not.) We fully agree that developmental psychology is a crucial testing ground for claims about moral nativism (see section 4.2.4), but we reject their claim that nativists are obligated to point to specific neural circuits, or to genes for those circuits. Given that nobody can find a set of genes that, collectively, explains 5% of the variance in how *tall* people are (Gudbjartsson et al., 2008), what chance is there that anyone will find a set of genes that code for mental modules (such as loyalty or sanctity) whose expression is far more subject to cultural influence than is height? To insist that nativists must point to genes is to ban nativism from psychology.

And yet, psychology has made enormous strides in recent years because of a flood of nativist findings. Personality psychology has been transformed by the discovery that nearly all personality traits are heritable (Bouchard, 1994; Turkheimer, 2000). Developmental psychology has been transformed by the discovery that infants have a great deal of innate knowledge about the physical world (Baillargeon, 1987; Spelke, 2000), and even about the social world (Hamlin, Wynn, and Bloom, 2007). These findings have earth-shaking implications for moral psychology, rendering blank-slate or pure learning approaches non-starters. None of these findings were reduced to “hand waving” by their authors’ failure to point to specific genes or brain circuits. It may have been a defensible strategy in the 1970s to assume that the mind is a blank slate and then require nativists to shoulder the burden of proof, but nowadays, we believe, the discussion

should focus on *how exactly* moral knowledge is innate, not *whether* it is (Tooby, Cosmides, & Barrett, 2005).

Nonetheless, Suhler and Churchland do point out places in which our “how exactly” discussion has been vague or underspecified, giving us an opportunity to improve the theory. In response to their critique, we offered a more detailed discussion of moral modularity (Haidt & Joseph, 2011; see also Haidt & Joseph, 2008). We have also tried to be much more specific in the current article about what exactly a foundation is, and how you know when something is innate (see section 4.2).

#### 4.1.2 Critiques of Cultural learning

Nobody doubts that cultural learning is a part of moral development, but cognitive developmentalists have long argued that morality is to a large extent self-constructed by the child. Piaget (1932/1965) strongly rejected ideas prevalent in his day that children internalized their moral values from society (e.g., Durkheim, 1925/1973) or from their parents (Freud, 1923/1962). Kohlberg (1969) believed that children go through two stages of “conventional” moral judgment, and Turiel argued that children are adept at distinguishing social conventions (which vary by culture) from true morality (which is universally applicable). But both men believed that real morality (post conventional, for Kohlberg; the moral domain, for Turiel) was something the child identified for herself during social interactions with peers, aided by the process of role-taking. Cognitive developmentalists carried out a variety of cross-cultural studies, but the goal of these studies – and their consistent conclusion – was that the fundamental stuff of morality did not vary across cultures (Kohlberg, 1969; Hollos, Leis, & Turiel, 1986). Again, as Kohlberg (1971) asserted: “Virtue is ultimately one, not many, and it is always the same ideal form regardless of climate or culture... The name of this ideal form is justice.” Any cross-cultural differences in the ability to reason about justice were explained as *developmental* differences: children in some cultures did not have as many opportunities for role-taking in egalitarian interactions, but if they did have those opportunities, they’d reach the same endpoint.

Of MFT’s four main claims, cultural learning has received the least direct criticism. Following Piaget, Kohlberg, and Turiel, researchers in the cognitive-developmental tradition could argue that MFT has overstated the role of cultural learning and underplayed the role of self-construction by conscious reasoning about care and fairness. This argument was made by Turiel, Killen, and Helwig (1987) against Shweder, Mahapatra, and Miller (1987). But none have advanced such a critique against MFT yet.

#### 4.1.3 Critiques of Intuitionism

Social psychologists and neuroscientists are generally comfortable with the enhanced role that automatic processes (including moral intuition) have played in moral psychology in recent years. Some researchers in those fields, however, favor a slightly different arrangement of reasoning and intuition. In particular, Greene et al. (2008) argue that “cognition” (or emotionless deliberative processing, which involves the dorsolateral prefrontal cortex) is not the servant of the “emotions” (rapid intuitive judgments which more heavily rely upon areas of the brain implicated in emotional responding). In Greene’s dual-process model, cognition and emotion are analogized to John Stuart Mill (cool utilitarian reasoning) versus Immanuel Kant (deontological principles, which are, paradoxically, based in emotion), fighting it out in the brain. However, Greene agrees with the basic intuitionist claim that rapid, automatic, affectively-laden processing often drives moral reasoning and turns it into rationalization.



Critiques of intuitionism (in the form of the Social Intuitionist Model, as well as other intuitionists such as Gigerenzer, 2007) are more common from developmental psychologists, particularly those in the cognitive developmental tradition (Narvaez, 2008; 2010; Saltzstein & Kasachkoff, 2004). Narvaez (2010) grants that intuitionism has been “a useful corrective to overly rationalistic approaches that have long dominated moral psychology” (p. 165). She also notes that “the vast research showing that humans often operate using implicit processes cannot be true everywhere except in the moral domain” (p. 165). Nonetheless, she argues that intuition and reasoning are best seen as partners in a dance, in which either partner can lead and the other will follow. She makes the important point that moral “expertise,” like other forms of expertise, often begins with conscious deliberation that gradually becomes automatic. She charges that “moral intuitionist theories often seem to rely on data from novices using seat-of-the-pants intuition—a quick, prereflective, front-end intuition that novices typically display” (p. 171). As a developmentalist, she is interested in how people arrive at “mature moral functioning,” and she is more interested in moral behavior than in moral judgment.

We think that Narvaez is correct that we have focused too much of our attention on the initial moral judgment, and not enough on the processes by which morality develops and improves with experience (see also Bloom, 2010). Given our interest in cultural development and the “revision process,” we believe MFT can be elaborated to address her concerns, and we are pleased that a few developmental psychologists have begun to do this, using MFT to study the development of character (Frimer et al., in press), the development of moral reasoning (Baril & Wright, in press), and the role of morality in adult development, including personal narratives (McAdams et al., 2008).

#### 4.1.4 Critiques of Pluralism per se

Much of the criticism of MFT has focused on its pluralism. We first address criticism from monists who reject the very notion of pluralism. Then we address critics who accept pluralism but argue for a different set of foundations, values, or virtues than the five we first proposed.

The most direct and detailed monist critique of MFT has come from Gray et al. (2012), who argue that all morality can be reduced to perceptions of dyadic harm (intentionally harmful agent plus suffering patient), and so only Care/harm is truly foundational: “A dyadic template suggests that perceived suffering is not only tied to immorality, but that *all* morality is understood through the lens of harm” (p. 108, emphasis added). While it seems reasonable to assert that Care/harm might be the most prototypical moral concern (see discussions of the Care/harm foundation in Haidt & Joseph, 2004; 2007), the reduction of *all* instances of moral judgment to perceptions of dyadic harm illustrates the deficits of moral monism. The idea that every moral judgment occurs through a single mental process (be it perceptions of dyadic harm, fairness intuitions, moral grammar, or another monist account) holds intuitive appeal, and it would certainly be convenient for moral scientists if morality worked this parsimoniously. But such an account quickly becomes Procrustean, cutting off phenomena it cannot explain (e.g., the role of disgust in moral judgments) and stretching its unitary construct to fit everything else (e.g., stretching “harm” to cover anything perceived as morally bad). (For more on this theory, see Ditto, Liu, & Wojcik, 2012; Graham & Iyer, 2012; and Koleva & Haidt, 2012.)

A similar harm-based moral monism has also been suggested by Harris (2010):

Haidt’s data on the differences between liberals and conservatives is interesting, but is his interpretation correct? It seems possible, for instance, that his five foundations of

morality are simply facets of a more general concern about harm. What, after all, is the problem with desecrating a copy of the Qu’ran? There would be no problem but for the fact that people believe that the Qu’ran is a divinely authored text. Such people almost surely believe that some harm could come to them or to their tribe as a result of such sacrileges—if not in this world, then in the next (p. 89 [see also pages 180-181]).

Harris makes his monist critique in the context of the larger normative argument that science should determine human values and pronounce which moral views are correct based on which ones lead to the greatest happiness (which can be measured in the brain by neuroscientific techniques). For the person morally offended by the desecration of a holy book, Harris suggests simply discarding the incorrect view that any deity exists who would cause harm because of it. Once that illusion is gone, one can correctly see, according to Harris, that desecrating a holy book is morally acceptable because it causes no harm. Moral monism is thus necessary for such a project, which requires a single standard by which to measure moral rightness or wrongness. For Harris, that standard is human welfare, defined in a rather narrow way: the absence of suffering.

But even if one agrees with Harris’s normative views, would the reduction of all morality to harm help us understand how morality actually works? Or would it be (to paraphrase William James) another attempt to clean up the litter the world actually contains? A monist model in which all moral judgments (even those based on objectively harmless transgressions) are produced by a single mental process (perceptions of intentional dyadic harm) cleans up much of the “litter” of empirically observed moral life, and in this cleaning suffers as a scientific description of morality. To name just three examples, such an account cannot explain: why incidental disgust harshens moral judgments (Schnall et al., 2008), why cognitive processes differ for Care- and Sanctity-based moral judgments (Young & Saxe, 2011), or why moral judgments of character can be produced by less harmful (Tannenbaum, Uhlmann, & Diermeier, 2011) or even harmless (Inbar, Pizarro, & Cushman, 2012) actions. Although not as explicitly committed to monism, accounts boiling morality down to fairness (Baumard et al., in press) or universal grammar (Mikhail, 2007) can suffer from the same deficits in their ability to adequately describe and explain human morality in all its messiness and complexity (see also Graham, in press).

#### **4.1.5 Alternative Pluralisms**

MFT has never claimed to offer an exhaustive list of moral foundations. We have tried from the beginning to identify the candidates for which the evidence was strongest, and we have actively sought out arguments and evidence for additional foundations. The first winner of the “moral foundations challenge” was John Jost, who suggested that we were missing concerns about liberty and oppression. As described in section 4.1.4, we have already begun empirical work testing Liberty/oppression as a possible sixth foundation. The second winner was the team of Elizabeth Shulman and Andrew Mastrorarde, who proposed that concerns about waste and inefficiency, particularly when a group is trying to achieve a common goal, produce an emotional reaction that is not related to any of the other foundations. The third winner was Polly Wiessner, an anthropologist who noted that issues of ownership and property arise everywhere, even among the !Kung Bushmen whom she studies, and that concerns about ownership have apparent precursors in animals’ ability to recognize and guard their own territories.

We think that Liberty/oppression, Efficiency/waste, and Ownership/theft are all good candidates for foundationhood, and we are conducting further research on those issues, along with Honesty/deception, to determine whether we should add any of them to the current list of

five foundations. We think the issue of identifying foundations is rather like the issue of counting planets. There are millions of objects orbiting the sun, but astronomers don't call them all planets. There are six (including the Earth) that are so visible that they were recorded in multiple ancient civilizations, and then there are a bunch of objects further out that were discovered with telescopes. Astronomers disagreed for a while as to whether Pluto and some more distant icy bodies should be considered planets. Similarly, we are content to say that there are many aspects of human nature that contribute to and constrain moral judgment, and our task is to identify the most important ones – the sets of social sensitivities that are most helpful for understanding intercultural and intracultural moral disagreements, and for understanding moral thought and behavior in general.

Although articulated well before the development of MFT, Turiel's (1979, 1983) moral-conventional distinction pre-figures one of the most common responses to MFT that we have heard from other researchers: two foundations – Care and Fairness – are legitimately moral, holding for all times and places, while the other three are merely conventional—valued in some times and places, but not in the same way as Care and Fairness. This critique was echoed by Jost (2009), who raised the normative objection that calling Loyalty, Authority, and Sanctity “moral” could legitimize anything from jingoism to blind obedience to prejudice and racism. Jost's objection raises a valid critique of some of our writings (Haidt & Graham, 2007; Haidt, 2007b) that blurred the line between the descriptive and the normative, and highlights the importance of carefully distinguishing the two. MFT is designed to provide a purely descriptive understanding of human morality, *not* to provide any normative justification (or condemnation) of any particular moral judgments or concerns. Although the word “moral” can introduce ambiguities because it has both descriptive and normative uses, MFT is about the foundations of morality as it is observed around the world, not about the moral systems that ought to prevail.

In contrast to the critique that MFT has included too much in its mapping of the moral domain, some have criticized it for not including enough. Janoff-Bulman and Sheikh (2012) presented a 2x3 matrix of moral motives, based on their work distinguishing approach-based moral prescriptions and avoidance-based moral proscriptions (Janoff-Bulman, Sheikh, & Baldacci, 2008; Janoff-Bulman, Sheikh, & Hepp, 2009) crossed with three contexts: intrapersonal, interpersonal, and intragroup. They argue that the moral foundations cover some of the six cells in this matrix, but fail to cover others – namely, intrapersonal prescriptions and proscriptions, and the intragroup prescriptions that characterize social justice solidarity concerns. Although MFT's treatment of all foundations involves both prescriptions (virtues to be admired) and proscriptions (vices to be condemned), Janoff-Bulman raises the important point that it's not necessarily the case that Care and Fairness only operate at the individual level, while Loyalty, Authority, and Sanctity (which we have sometimes referred to as the “binding” foundations) always operate at the group level. Graham and Haidt (2010) describe several intrapersonal concerns related to Sanctity (e.g. prescriptions for treating one's body as a temple, proscriptions against masturbation and impure thoughts). Interestingly, our own self-critiques (see below) have also brought up the group-focused fairness concerns we've been missing (e.g., equity and vigilance against free-riders, reciprocal retaliations for outgroup attacks), but unlike social justice concerns (which we see as primarily focused on group members, not the group itself), we predict that these concerns would be endorsed more by conservatives than by liberals. Nevertheless, considering the different fundamental psychological motives (approach/avoid) involved in different moral concerns will be a promising area for future development.

Further, Janoff-Bulman's inclusion of different contexts (intrapersonal, interpersonal, intragroup, and even intergroup) echoes the critique by Rai and Fiske (2010) that MFT doesn't pay enough attention to *relational context*. Specifically, they propose four moral motives – unity, hierarchy, equality, and proportionality – based on Fiske's (1992) relational models described above in section 1, and say that these motives can add to MFT “by grounding the foundations in a theory of social relationships and thereby predicting when and how people will rely on one foundation over another” (p. 67). Jarudi (2009) suggests an expansion of the Sanctity domain, distinguishing between sexual purity and food purity. Finally, in addition to their meta-theoretic critiques of MFT's approach (see section 4.1.1), Suhler and Churchland (2011) suggest other candidate foundations, such as industry and modesty.

Despite the collective coherence suggested by our use of “we” throughout this chapter, we are constantly arguing among ourselves over changes to existing foundations and considerations of new candidate foundations. Iyer (2009) first pointed out that our measures of Fairness concerns centered on equality rather than equity, and that concerns about equality are often motivated by care for others, whereas equity concerns may be motivationally distinct (Iyer, Read, & Correia, 2010). Iyer (2009) also questioned the pragmatic utility of separating Loyalty and Authority, suggesting that both concerns could conceptually be considered part of a single foundation concerning subsuming one's interests for one's group. Analyses of libertarians (described in section 3.2.4) raised the question of whether Liberty/oppression is its own basic moral concern, not reducible to self-interest or existing foundations. And in responses to open-ended questions about what people felt guilty about (or ways in which they were not living up to their values), honesty violations come up more frequently than any other kind of concern (see Iyer, 2010, on treating honesty as a separate foundation). We are currently investigating all of these as part of the method-theory co-evolution of MFT.

#### **4.2 Getting Specific: What does it take to be a foundation?**

One common critique of MFT has been that our list of foundations is arbitrary, chosen originally by Haidt and Joseph (2004) based on their reading of five books and articles. Many scientists would like to see a set of explicit criteria which researchers could use to decide what counts as a foundation. We agree that such a list would be helpful for progress in moral psychology.

We therefore offer a list of five criteria for foundationhood. We looked for guidance to the long-running debate over how many “basic emotions” there are. Ekman, Sorenson, and Friesen (1969) originally offered a list of six emotions, based on their research on facial expressions: joy, sadness, anger, fear, surprise, and disgust. Gradually a few more emotions were added, and eventually Ekman (1992) offered a set of nine criteria for what it takes to be a basic emotion. He made no commitment to parsimony, suggesting that perhaps 17 emotions might eventually qualify as basic emotions (Ekman, 1994). Some emotions, such as fear and anger, meet all of Ekman's criteria very cleanly; they are prototypical emotions, about which there is less debate. Other emotions, such as awe, relief, and contentment, meet most of the criteria to some degree, making them less prototypical exemplars of emotionhood and leaving more room for debate.

We think the same is true of foundationhood. We think that our original list of five foundations did a good job of capturing the most obvious and least debatable foundations, but we acknowledge that there's still room for debate, and, like Ekman, we're confident that our initial list is not the final list.

Here, then, is our list of five criteria (see below and Table 4). The first two criteria establish the kinds of phenomena we’re studying – intuitive moral judgments. The last three indicate that a content area of morality may be related to an innate but variably expressed foundation. We’ll illustrate each criterion by discussing the Fairness foundation, which we believe meets all criteria extremely well.

#### **4.2.1 Criterion 1: A common concern in third-party normative judgments**

One of the most significant steps in the evolution of morality may have occurred when human beings developed “shared intentionality” – the ability of multiple people to hold a shared mental representation of what they are trying to do together (Tomasello, Carpenter, Call, Behne, & Moll, 2005). Chimpanzees seem to have some sense of norms for behavior within the group, and they sometimes get upset when they are not treated according to their expectations. The evidence that they react to third parties who violate norms, however, is mixed or anecdotal at best (de Waal, 1996). But when humans developed the capacity for shared intentionality, our capacity to recognize norms began to grow into a passion for enforcing them on each other (Boehm, 2012). Humans began to live in “moral matrices” – the “consensual hallucinations” that provide a common normative framework against which people can and do judge the actions of others, even when those actions have no direct implications for the self (Haidt, 2012).

The sorts of third-party violations that people in a community react to is a good guide to where moral foundations should be sought. If a putatively moral issue never shows up in gossip, even in communities that are said to endorse values related to that foundation, then that’s a reason to doubt the existence of such a foundation. Gossip about fairness, for example, is ubiquitous. From hunter-gatherers (Wiessner, 2005) to Chaldean-Iraqi merchants in Michigan (Henrich and Henrich, 2007) to college roommates sharing a kitchen, people gossip frequently about members of their group who cheat, fail to repay favors, or take more than their share. In fact, Dunbar (1996) reports that one of the principle functions of gossip is to catch cheaters and free-riders within groups.

In the first row of Table 4 we’ve listed studies that show people making third-party moral judgments—condemning others for actions that have no direct consequences for the self. These studies show that people in at least some cultural groups make judgments closely related to the content of all five foundations. Our own studies using multiple measures provide ample documentation of people condemning third parties for violations related to each foundation (e.g., Graham, Haidt & Nosek, 2009).

#### **4.2.2 Criterion 2. Automatic affective evaluations**

MFT is an intuitionist theory – it tries to explain the rapid, automatic reactions people have to violations of what they take to be a shared moral order. There is not just one moral intuition—a general flash of “wrongness”—just as there is not one taste receptor on the tongue whose output tells us “delicious!” Rather, we posit that there are a variety of rapid, automatic reactions to patterns in the social world. When we detect such patterns, moral modules fire, and a fully enculturated person has an affectively-valenced experience. Not just a feeling of “good!” or “bad!”, but an experience with a more specific “flavor” to it, such as “cruel!”, “unfair!”, “betrayal!”, “subversive!”, or “sick!” If a moral reaction can be elicited quickly and easily, with a variety of images, bumper-stickers, or one-sentence stories, that is a point in favor of its foundationhood. Reactions to unequal distributions among children are often visible on the face of the disadvantaged child within one second (LoBue, Chiong, Nishida, DeLoache & Haidt,

2011), and fMRI studies repeatedly show that people have rapid, affectively-laden reactions to being cheated, and those reactions tend to activate brain areas related to emotion, including the anterior insula and the orbitofrontal cortex (Rilling et al., 2002; Sanfey et al., 2003). In an fMRI study of economic games, fair offers (compared to unfair offers of the same value) activated neural reward circuitry, while accepting unfair offers activated self-control circuitry (Tabibnia, Satpute, & Liberman, 2008). It's easy to trigger rapid and affectively-laden judgments of unfairness using still photos, bumper stickers, or a single number on a computer screen that reveals one's partner's choice in a cooperative game. The same is true for images of harm or cruelty activating the Care foundation (e.g., Luo et al., 2006), and stories about sexual violations activating the Sanctity foundation (e.g., Parkinson et al., 2011). There has been less research on automatic reactions to violations of Loyalty and Authority, but here too studies have shown split-second reactions to sentences, words, or pictures showing violations of these foundations (Cannon et al., 2011; Graham, 2010).

#### **4.2.3 Criterion 3. Culturally widespread**

We have proposed that moral foundations are part of the “first draft” of the moral mind. These drafts get edited during childhood development within a particular culture, and some cultures actively suppress some of the foundations. Examples include the ways that Nazi Germany turned compassion into the vice of “softness” (Koonz, 2003), or the way that egalitarian movements such as Occupy Wall Street have tried to create “horizontal” societal structures that do not rely on the Authority foundation. So it is not necessary that a foundation be shown to underlie morality in *all* human cultures. Innate does not mean universally visible in the adult phenotype. It means “organized in advance of experience,” such that we should expect to see it expressed in some form in *most* human cultures.

Additionally, we should not treat all cultures as equally informative. Hunter-gatherer societies should carry added weight, because they may more closely resemble lifestyles of the “environment of evolutionary adaptation” (Cosmides & Tooby, 1994) in which the moral foundations presumably evolved. Traditional societies with small-scale agriculture or herding have also existed for long enough periods to have produced genetic adaptations (e.g., for lactose tolerance and starch metabolism, and quite possibly for behavior too; see Cochran & Harpending, 2009). If moral foundations were shaped by gene-culture co-evolution (Richerson & Boyd, 2005) in response to long-standing adaptive challenges, then a candidate foundation should be easily visible in anthropological reports from these societies. Modern “WEIRD” societies (Henrich, Heine & Norenzayan, 2010) are arguably the worst places to look for moral foundations because such societies have narrowed the moral domain in order to grant individuals the maximum freedom to pursue their projects. Nonetheless, when similar moral concerns are found across WEIRD societies, agricultural societies, and hunter-gatherer societies, the case for foundationhood gets stronger. Fairness certainly passes this test – nobody has yet identified a society in which reciprocity is not an important moral concern (Brown, 1991; Fiske, 1992). The other foundations also show up widely in anthropological accounts (as shown in the third row of Table 4), and in Brown's (1991) list of human universals.

Authority is a particularly interesting case in that hunter-gatherer societies are generally egalitarian. Yet as Boehm (1999) explains, it's not that they lack the innate cognitive and emotional structures for implementing hierarchical relationships, because such relationships emerge very rapidly when groups take up agriculture. Rather, hunter-gatherers generally find cultural mechanisms of suppressing the ever-present threat of alpha-male behavior, thereby

maintaining egalitarian relationships among adult males in spite of the hierarchical tendencies found among most primates, including humans.

#### **4.2.4 Criterion 4. Evidence of innate preparedness**

The fact that a behavior or ability is found in most or all human societies does not prove that anything is innate. All human societies face some similar challenges, and it is quite possible that all societies have hit upon similar solutions using their general-purpose, non-domain-specific intelligence. For example, all societies have invented ways to carry water. Perhaps all societies have invented fairness and turn-taking as efficient solutions to the challenge of dividing scarce resources; perhaps all societies have invented food taboos in response to the real dangers of toxins and contaminants. Perhaps there are no innate moral foundations.

The case for innateness grows much stronger when a behavior or ability is found in non-human primates (particularly chimpanzees and bonobos) and when it can be shown to emerge in young children before they have been exposed to relevant teaching or reinforcement. Contrary to Suhler and Churchland (2011), we do not believe that claims about innateness need to point to specific genes or brain areas. Rather, nativists must offer some reason for believing that a behavior or ability is “organized in advance of experience.”

De Waal (1995) has long argued that the “building blocks” of human morality are present in other primates. We believe that such building blocks have been shown for the Care foundation (i.e., empathy and nurturance; Hrdy, 2009; Preston & de Waal, 2002), the Loyalty foundation (coalitional behavior and inter-coalitional conflict; de Waal, 1982); and the Authority foundation (rank and deference; Boehm, 1999; 2012). There is some evidence for precursors of Fairness (Brosnan, 2006), but it is more anecdotal, and the limited lab evidence (e.g., Brosnan & de Waal, 2003) has been disputed (Brauer, Call, & Tomasello, 2009; see also Hammerstein, 2003). We know of no evidence that non-human primates have any building blocks of the Sanctity foundation, such as the emotion of disgust, or even contamination sensitivity (see Rozin & Fallon, 1987). We presume that Sanctity is the most recently evolved foundation, perhaps co-evolving with human religiosity in the last one or two hundred thousand years.

Recent findings in developmental psychology strongly support the nativist claims of MFT. The fourth row of Table 2 lists examples of such research. In the last six years, infants and young children have been shown to have surprisingly sophisticated social-cognitive abilities, often including affective reactions to third-party violators (i.e., puppets who do bad things to other puppets). For example, infants don’t like puppets who harm others, but they do like puppets who help others (Hamlin, Wynn, & Bloom, 2007). Infants are also sensitive to third-party fairness violations (Sloane, Baillargeon, & Premack, 2012); interestingly, this sensitivity predicted infants’ own altruistic sharing behavior (Schmidt & Sommerville, 2011). Children as young as three are adept at sharing rewards equally, but only when they both cooperated to produce the benefit (Hamann, Warneken, Greenberg, & Tomasello, 2011). Infants notice markers of ingroup membership and prefer members of their ingroup (Kinzler, Dupoux, & Spelke, 2007), and even prefer those who help similar others and harm dissimilar others (Hamlin, Mahajan, Liberman, & Wynn, 2012). We know of no research on how infants process markers of authority and respect, or of purity, sanctity, or contagion; we hope that such research will be done in the future. But we do note that children’s games are often based on a single foundation, giving children the opportunity to practice a portion of their moral repertoire. For example, the game of “Simon Says” appoints a leader who commands followers, and the game of cooties is about contagion and how to remove contagion (i.e., with a “cooties shot”). The

concept of “cooties” is not found universally, but it has been identified in several far-flung cultures (Hirschfeld, 2002; Samuelson, 1980), it seems to emerge with no encouragement from adults, and it emerges in Western societies that discourage the use of caste and contagion as moral categories. Importantly, cooties games tend to emerge around the age of 7 or 8 (Opie & Opie, 1969), which is the age at which disgust sensitivity becomes pronounced (Rozin & Fallon, 1987). In other words, these games seem to reflect the *externalization* of children’s developing social-emotional abilities, not the *internalization* of prevailing cultural norms.

#### **4.2.5 Criterion 5. Evolutionary model demonstrates adaptive advantage**

Anti-nativists often criticize evolutionary psychology as a collection of “just-so” stories. And indeed, given the power of the human imagination and the epistemological predilections of the confirmation bias, one could invent an evolutionary story for just about any candidate foundation, especially if one is allowed to appeal to the good of the *group*. But a good evolutionary theory will specify—often with rigorous mathematical models—exactly how a putative feature conferred an adaptive advantage upon *individuals* (or upon other bearers of the relevant genes), in comparison to members of the same group who lacked that feature. A good evolutionary theory will not casually attribute the adaptive advantage to the group (i.e., appeal to group selection) without a great deal of additional work, e.g., showing that the feature confers a very strong advantage upon groups during intergroup competition while conferring only a small disadvantage upon the individual bearer of the trait (see Wilson, 2002; and see Haidt, 2012, ch. 9, on group-selection for groupish virtues). If no clear adaptive advantage can be shown, then that is a mark against foundationhood.

Another important safeguard against “just-so” thinking is to rely upon already-existing evolutionary theories. As we said in section 1, MFT was inspired by the obvious match between the major evolutionary theories and the major moral phenomena reported by anthropologists. We engaged in no post-hoc evolutionary theorizing ourselves. The fifth row of Table 4 shows evolutionary theories that spell out the adaptive advantages of certain innate mechanisms which we posit to be among the modules comprising each foundation. For example, the fairness foundation is largely just an elaboration of the psychology described by Trivers (1971) as the evolved psychological mechanisms that motivate people to play “tit for tat.”

In sum, we have offered five criteria for foundationhood. Any moral ability, sensitivity, or tendency that a researcher wants to propose as an expression of an additional moral foundation should meet these criteria. At that point, the researcher will have established that there is something innate and foundational about an aspect of human morality. The only hurdle left to clear to get added to the list of moral foundations is to show that the candidate foundation is distinct from the existing foundations. For example, we do not believe that there is an “equality” foundation, not because we think there is nothing innate about equality, but because we think that equality is already accounted for by our existing foundations. Equality in the distribution of goods and rewards is (we believe) related to the Fairness foundation. Equality is a special case of equity: when all parties contributed equally, then all parties should share equally (Walster, Walster, & Berscheid, 1978). People who take more than their share are cheating. Moral judgments related to *political* equality – particularly the anger at bullies and dominators who oppress others – may be an expression of the candidate Liberty/oppression foundation. (See Haidt, 2012, ch. 8, for further discussion of equality, equity, and liberty).



### 4.3 Looking ahead: New Directions for Moral Foundations Research

The preceding sections of this chapter give an indication not only of the work that has been done using MFT, but also the work that has yet to be done. For instance, section 3.1 describes many different methods for measuring foundation-related concerns explicitly and implicitly, and yet the majority of the empirical work described in section 3 relies on just one of those methods (the MFQ; see Table 2). In this penultimate section we describe the future we see for moral foundations research, for refining the theory itself and applying it to new research questions.

#### 4.3.1 Method-Theory Co-Evolution of MFT

We began this “future directions” section with a detailed discussion of various critiques of MFT, because we see such critiques as crucial for the progress and future shaping of the theory. In our vision of method-theory co-evolution, critics are especially needed on the theory side, pointing out problems with existing constructs and offering competing conceptualizations of the moral domain. We expect that work bridging MFT with other theories will be productive, for MFT and for moral psychology overall. Janoff-Bulman and Sheikh (2012) and Rai and Fiske (2010) have both offered expansions or alternate configurations of the moral foundations, and while we may disagree on some particulars, none of these theories are incompatible. They are different ways of approaching the same phenomena. And while a strict moral monism *is* incompatible with MFT, monist critiques such as those offered by Gray and colleagues (2012) can also advance the science by prompting more work on how different kinds of moral concerns can be similar as well as distinct.

Working out where the theories converge and diverge can help advance our understanding of morality – particularly once competing predictions can be spelled out to testable hypotheses. Given the confirmation bias inevitable when researchers test their own theories, adversarial collaborations (in which the adversaries first agree on terminology, predictions, and what counts as evidence for and against specific claims) may be necessary to avoid the kind of unresolved theoretical stalemates described by Greenwald (2012).

In this vein, tests of alternate foundations will likely characterize the next few years of moral foundations development. We hope that the criteria spelled out in section 4.2 will be useful for such efforts: what is the existing evidence along these five criteria for the candidate foundations described in section 4.1, such as liberty, honesty, waste, property, social justice, industry, and modesty? Where is the evidence the strongest, and where does more work need to be done to test candidate foundations? This isn’t to say that more work won’t be done on the five initial foundations. Table 4 highlights areas where little evidence is currently available for particular foundations on particular criteria. For instance, might infants show some ability to detect violations of Authority, and to respond to violators negatively? Such work is likely to lead to the creation of new methods as well. Again, the end goal is for competing conceptualizations and theories to be specified and worked out to the point that new methods are developed to marshal evidence for the claims, which will bring new (often unexpected) findings that can in turn lead to new theoretical syntheses and developments.

#### 4.3.2 Applying MFT to New Areas, and New Questions

Theories typically reflect the strengths and weaknesses of their founders. We (the authors) are all social psychologists with interests in political ideology, and so it should be of no surprise that most of the work described in section 4 falls in the realm of political psychology.

Nevertheless, we are hopeful that as more and more researchers make use of MFT's methods and constructs, the benefits of moral pluralism can be realized in more and more content areas and disciplines. Here are a few areas we see as particularly fertile.

*Implicit social cognition* - First, we are beginning to see more work by cognitive scientists on how judgment processes differ for different kinds of concerns. This work has mostly contrasted Sanctity with Care or Fairness (e.g., Young & Saxe, 2011; Inbar, Pizarro, & Cushman, 2012; Feinberg & Willer, in press), but so far much less work has been done on Loyalty and Authority concerns. Recent (see Table 2) and future implicit measures of foundation-related intuitions and reactions could be used by social/cognitive psychologists and neuroscientists to learn more about the automatic processes associated with foundation-related judgments.

*Development* – Second, developmental psychologists are just beginning to test the earliest signs of emergence for moral concerns other than care and fairness. There is much fertile research ground here for both infant/toddler studies and lifespan development studies – do the “binding” concerns of Loyalty, Authority, and Sanctity become more important as people get older, become parents, or take on leadership positions at work? What are the different patterns of emergence and developmental trajectories for different foundational concerns?

*Culture and social ecology* – As Shweder (1991) says, each culture is expert in some aspects of human flourishing, but not all. Although we are working with researchers in other nations to explore the morality of other cultures (see section 3.4), much more work needs to be done to move beyond WEIRD research samples (Henrich et al., 2011). Variations in social ecology (Oishi & Graham, 2010), such as residential mobility, economic structure, or population density, could also be important for predicting foundation endorsements. In one large-scale study, Motyl (2012) found that moral misfits – partisans living in communities which voted heavily against their party's US Presidential candidate – were disproportionately likely to move to a new community. Furthermore, their new communities voted more heavily for the participant's party's candidate. This research suggests that moral values may steer people to live in morally-segregated groups, with implications for attitude polarization and intergroup conflict.

*Beyond moral intuitions* – As noted above, research should examine what happens after the initial moral judgment is made. A central but largely understudied component of the Social Intuitionist Model (Haidt, 2001) is that while one's initial moral judgment is typically intuitive, explicit moral reasoning plays many important roles as people gossip, argue, and otherwise talk about moral issues with other people. We hope that researchers will study moral disagreements as they play out over the course of many days or months, sometimes shifting in terms of the moral foundations used to justify judgments (e.g. see Koleva et al., 2012).

*Interpersonal morality* – One criticism of MFT, and of morality research in general, is that it largely ignores the role of interpersonal and relational factors (Rai & Fiske, 2011). Do moral judgments based on the moral foundations indeed vary for different relationship contexts? Moreover, what are the interpersonal antecedents and consequences of individual variation in foundation-related concerns? We have recently begun to explore these questions – for example, Koleva (2011) examined the role of moral foundation similarity in romantic ideals and relationship satisfaction, and more recent work is examining the relationships between adult romantic attachment and foundation concerns – but many more questions remain.

*From moral judgment to moral behavior* – The virtues have been central to MFT in theory, but not yet in practice. As Narvaez (2010) asked, what is the relation between moral judgment on one hand, and actual moral behavior on the other? Following Graham, Meindl, and

Beall (in press), how can the pluralism of moral judgments and concerns help researchers capture a wider array of morally-relevant behaviors? Relatedly, what are the practical real-world implications of MFT for persuasion (e.g., Feinberg & Willer, in press) or other aspects of moral disagreements?

*Beyond psychology* – Researchers in many departments beyond psychology have begun to apply MFT’s methods and constructs to such fields as public policy (Oxley, 2010), media studies (Tamborini, 2011), marketing (Winterich, Zhang, & Mittal, in press), legal studies (Prince, 2012), climate science (Markowitz & Shariff, 2012), business ethics (Sadler-Smith, in press), political science (Jones, 2012), genetics (Smith et al., 2011) neuropsychology (Young & Saxe, 2011), neuroanatomy (Lewis et al., 2012), and even agricultural ethics (Makiniemi et al., in press). Given the importance of values in real-world domains such as philanthropy, politics, and business, we hope that MFT proves useful beyond academia as well.

## 5 Conclusion

A cherished maxim in psychology comes from Kurt Lewin (1951): "There is nothing so practical as a good theory." Putting this maxim together with Einstein’s maxim at the opening of this chapter, we think MFT is a good theory. It is a practical theory—complete with a set of well-validated measurement tools—which has quickly yielded a great variety of new findings, in many fields. It is a non-Procrustean theory which does not force researchers to “surrender the adequate representation” of experience. And it is an open and revisable theory, offering an initial list of foundations along with a list of criteria for how to revise the list. MFT is a theory in motion, a theory to be expanded, constricted, refined, and built upon. Above all, we think it is the right theory for our age—a golden age of cross-disciplinary research in which most scientists studying morality have at least some familiarity with findings in neighboring fields. Conferences on moral psychology nowadays often include researchers who study chimpanzees, psychopaths, infants, hunter-gatherers, or people with brain damage. MFT gives this varied set of researchers a common language for talking about the moral domain. It calms the sometimes-divisive nature-nurture debate by distinguishing the first draft of the moral mind and the experiential editing process.

We think MFT is practical in another way too: it helps researchers as well as the general public look beyond the moral values that are dearest to them, and understand those who live in a different moral matrix. We close with a final quote from Isaiah Berlin (2001), who explains one reason why pluralism is so practical:

If I am a man or a woman with sufficient imagination (and this I do need), I can enter into a value system which is not my own, but which is nevertheless something I can conceive of men pursuing while remaining human, while remaining creatures with whom I can communicate, with whom I have some common values—for all human beings must have some common values or they cease to be human, and also some different values else they cease to differ, as in fact they do. That is why pluralism is not relativism—the multiple values are objective, part of the essence of humanity rather than arbitrary creations of men's subjective fancies.

## References

- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256-274.
- Amodio, D. M., Jost, J. T., Master, S. L., & Yee, C. M. (2007). Neurocognitive correlates of liberalism and conservatism. *Nature Neuroscience*, *10*, 1246-1247.
- Asch, S. E. (1959). A perspective on social psychology. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3, pp. 363-383). New York: McGraw-Hill.
- Baillargeon, R. (1987). Object permanence in 3 1/2 - and 4 1/2 -month-old infants. *Developmental Psychology*, *23*(5), 655-664.
- Baril, G., & Wright, J.C. (in press). Different types of moral cognition: Moral stages versus moral foundations. *Personality and Individual Differences*.
- Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*, 462-479.
- Barkow, J., Cosmides, L. & Tooby, J., eds. (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.
- Baron-Cohen, S. (2009). Autism: the empathizing-systemizing (E-S) theory. *The Year in Cognitive Neuroscience. Annals of the New York Academy of Science*, *1156*, 68-80.
- Barrett, H. C., and Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, *113*, 628-647.
- Bastick, T. (1982). *Intuition: How we think and act*. Chichester, England: Wiley.
- Baumard, N., André, J.B. and Sperber, D. (in press). A mutualistic approach to morality, *Behavioral and Brain Sciences*.
- Baumeister, R. F., & Newman, L. S. (1994). How stories make sense of personal experiences: Motives that shape autobiographical narratives. *Personality and Social Psychology Bulletin*, *20*, 676-690.
- Berlin, I. (1969). *Four essays on liberty*. Oxford University Press, USA.
- Berlin, I. (2001). My intellectual path. In H. Hardy (ed.), *The Power of Ideas 1-23*, Princeton, NJ: Princeton University Press.
- Bloom, P. (2010). How do morals change? *Nature*, *464*, 490.
- Bobbio, A., Nencini, A., & Sarrica, M. (2011). Il Moral Foundation Questionnaire: Analisi della struttura fattoriale della versione italiana. *Giornale di Psicologia*, *5*, 7-18.
- Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Cambridge, MA: Harvard University Press.
- Boehm, C. (2012). *Moral origins: The evolution of virtue, altruism, and shame*. New York: Basic.
- Bouchard, T. J. J. (1994). Genes, environment, and personality. *Science*, *264*, 1700-1701.
- Bowlby, J. (1969). *Attachment and loss: Vol.1: Attachment*. New York: Basic.
- Brauer, J., Call, J., & Tomasello, M. (2006). Are apes really inequity averse? *Proceedings of the Royal Society B*, *273*, 3123-3128.
- Brehm, S. S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control*. London: Academic Press, Inc.
- Brosnan, S. F. (2006). Nonhuman species' reactions to inequity and their implications for fairness. *Social Justice Research*, *19*, 153-185.
- Brown, R. W., & Ford, M. (1964). Address in American English. In D. Hymes (Ed.), *Language in Culture and Society*. New York. Harper & Row.

- Brown, D. E. (1991). *Human universals*. Philadelphia: Temple University Press.
- Bruneau, E.G., Dufour, N., & Saxe, R. (2012). Social cognition in members of conflict groups: Behavioural and neural responses in Arabs, Israelis and South Americans to each other's misfortunes. *Philosophical Transactions of the Royal Society Biological Sciences*, 367, 717-730.
- Bruner, J. S. (1960). *The process of education*. Cambridge, MA: Harvard University Press.
- Cannon, P. R., Schnall, S., & White, M. (2011). Transgressions and expressions: Affective facial muscle activity predicts moral judgments. *Social Psychological and Personality Science*, 2, 325-331.
- Clifford, S., & Jerit, J. (in press). How word do the work of politics: Moral Foundations Theory and the debate over stem-cell research. *Journal of Politics*.
- Cochran, G., & Harpending, H. (2009). *The 10,000 year explosion: How civilization accelerated human evolution*. New York: Basic.
- Conover, P. J. & Feldman, S. (1981). The origins and meaning of liberal-conservative self-identifications. *American Journal of Political Science*, 25(4), 617-645.
- Cosmides, Leda, and John Tooby. (1994). Origins of Domain Specificity: The Evolution of Functional Organization. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*, (pp. 85-116). Cambridge, UK: Cambridge University Press.
- Damasio, A. R. (2000). *The feeling of what happens: Body and emotion in the making of consciousness*. Mariner Books.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113-126.
- Dawkins, Richard (1976). *The selfish gene*. New York City: Oxford University Press.
- Dehghani, M., Gratch, J., Sachdeva, S. & Sagae K. (2011). Analyzing the debate over the construction of the Ground Zero Mosque. Presented at the 2011 meeting of the Society of Judgment and Decision Making (SJDM), Seattle, WA.
- DeLoache, J. S., & LoBue, V. (2009). The narrow fellow in the grass: Human infants associate snakes and fear. *Developmental Science*. 12(1), 201-207.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, 93, 880-896.
- De Waal, F. (1982). *Chimpanzee politics: Power and sex among apes*. London: Jonathan Cape.
- De Waal, F. B. M. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.
- Ditto, P. & Koleva, S. P. (2011). Moral empathy gaps and the American culture war. *Emotion Review* (special issue on "Morality and Emotion" edited by Joshua Greene), 3(3), 331-332.
- Ditto, P. H., Liu, B., & Wojcik, S. P. (2012). Is anything sacred anymore? Commentary on target article, *Mind perception is the essence of morality* (K. Gray, L. Young, & A. Waytz), *Psychological Inquiry*, 23, 155-161.
- Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In B. H. Ross (Series Ed.) & D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Psychology of learning and motivation, Vol. 50: Moral judgment and decision making* (pp. 307-338). San Diego, CA: Academic Press.

- Dodd, M., Balzer, A., Jacobs, C., Gruszczynski, M., Smith, K. & Hibbing, J. (2012). The political left rolls with the good and the political right confronts the bad: connecting physiology and cognition to preferences. *Philosophical Transactions of the Royal Society B*, 367(1589), 640-649
- Douglas, M. (1966). *Purity and danger*. London: Routledge and Kegan Paul.
- Duckitt, J. (2001). A cognitive-motivational theory of ideology and prejudice. In M.P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 33, pp. 41-113). San Diego: Academic Press.
- Dunbar, R. (1996). *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.
- Durkheim, E. (1925/1973). *Moral Education* (E. Wilson & H. Schnurer, Trans.). New York: The Free Press.
- Eidelman, S., Crandall, C.S., Goodman, J.A., & Blanchard, J.C. (2012). Low-effort thought promotes political conservatism. *Personality and Social Psychology Bulletin*, 38, 808-820.
- Einstein, A. (1934). On the method of theoretical physics. *Philosophy of Science*, 1, 163-169.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6, 169-200.
- Ekman, P. (1994). All emotions are basic. In P. Ekman & R. Davidson (Eds.), *The Nature of Emotion* (pp. 15-19). New York: Oxford University Press.
- Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164, 86-88.
- Faulkner, J., Schaller, M., Park, J. H., & Duncan, L. A. (2004). Evolved disease-avoidance mechanisms and contemporary xenophobic attitudes. *Group Processes & Intergroup Relations*, 7, 333-353.
- Federico, C. M., Weber, C. R., Ergun, D., & Hunt, C. (in press). Mapping the connections between politics and morality: The multiple sociopolitical orientations involved in moral intuition. *Political Psychology*.
- Feinberg, M. & Willer, R. (in press). The moral roots of environmental attitudes. *Psychological Science*.
- Ferguson, M. J. (2007). On the automatic evaluation of end-states. *Journal of Personality and Social Psychology*, 92, 596-611.
- Fiske, A. P. (1991). *Structures of social life: The four elementary forms of human relations: Communal sharing, authority ranking, equality matching, market pricing*. New York: Free Press.
- Fiske, S. T. (1992). Thinking is for doing: Portraits of social cognition from daguerreotype to laserphoto. *Journal of Personality and Social Psychology*, 63, 877-889.
- Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. Cambridge, Mass.: MIT Press.
- Frank, R. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.
- Freud, S. (1923/1962). *The ego and the id* (J. Riviere, Trans.). New York: Norton.
- Frimer, J. A., Biesanz, J. C., Walker, L. J., & MacKinlay, C. W. (in press). Liberals and conservatives rely on common moral foundations when making moral judgments about influential people. *Journal of Personality and Social Psychology*.
- Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Viking Press.

- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.
- Glenn, A.L., Iyer, R., Graham, J., Koleva, S., & Haidt, J. (2009). Are all types of morality compromised in psychopathy? *Journal of Personality Disorders*, 23(4), 384-398.
- Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge, MA: Belknap Press.
- Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London*, 205B, 581-598.
- Graham, J. (2010). *Left gut, right gut: Ideology and automatic moral reactions*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (AAT 3437423).
- Graham, J. (in press). Beyond economic games: A mutualistic approach to the rest of moral life. [Commentary on Baumard, André, & Sperber]. *Behavioral and Brain Sciences*.
- Graham, J., Haidt, J., & Nosek, B.A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96, 1029-1046.
- Graham, J., & Haidt, J. (2011). Sacred values and evil adversaries: A Moral Foundations approach. In P. Shaver & M. Mikulincer (Eds.), *The Social Psychology of Morality: Exploring the Causes of Good and Evil*. New York: APA Books.
- Graham, J., & Iyer, R. The unbearable vagueness of "essence": Forty-four clarification questions for Gray, Young, and Waytz. *Psychological Inquiry*, 23, 162-165.
- Graham, J., Meindl, P., & Beall, E. (in press). Integrating the streams of morality research: The case of political ideology. *Current Directions in Psychological Science*.
- Graham, J., Nosek, B.A., & Haidt, J. (in press). The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum. *PLoS ONE*.
- 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.
- Gray, K., Young, L., & Waytz, A. (2012) Mind perception is the essence of morality. *Psychological Inquiry*, 23, 101-124.
- Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E., Cohen, J.D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, 107, 1144-1154.
- Greenwald, A. G. (2012). There is nothing so theoretical as a good method. *Perspectives on Psychological Science*, 7, 99-108.
- Gudbjartsson, D. F., Walters, G. B., Thorleifsson, G., Stefansson, H., Halldorsson, B. V., Zusmanovich, P., et al. (2008). Many sequence variants affecting diversity of adult human height. *Nature Genetics*, 40, 609-615.
- 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., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20, 98-116.
- Haidt, J., Graham, J. (2009). Planet of the Durkheimians, where community, authority, and sacredness are foundations of morality. *Social and Psychological Bases of Ideology and System Justification*, Oxford University Press.
- Haidt, J., Graham, J., & Joseph, C. (2009). Above and below left-right: ideological narratives and moral foundations. *Psychological Inquiry*, 20, 110-119.

- Haidt, J. (2007a). The new synthesis in moral psychology. *Science*, *316*, 998-1002.
- Haidt, J. (2007b). *Moral psychology and the misunderstanding of religion*. Retrieved on July 20, 2012, from [http://www.edge.org/3rd\\_culture/haidt07/haidt07\\_index.html](http://www.edge.org/3rd_culture/haidt07/haidt07_index.html)
- Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. Cambridge, MA: MIT Press. (pp. 181-217).
- Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and reasons of liberals and conservatives. *Journal of Applied Social Psychology*, *31*, 191-221.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*, 55-66.
- Haidt, J., & Joseph, C. (2007). The moral mind: How 5 sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The Innate Mind, Vol. 3* (pp. 367-391). New York: Oxford.
- 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.
- Hamann, K., Warneken, F., Greenberg, J. R., & Tomasello, M. (2011). Collaboration encourages equal sharing in children but not in chimpanzees. *Nature*, *476*, 328-331.
- Hamilton, W. D. (1964). The genetical evolution of social behavior. II. *Journal of Theoretical Biology*, *7*, 17-52.
- Hamlin, K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*, 557-559.
- Hammerstein, P. (2003). Why is reciprocity so rare in social animals? In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation* (pp. 55-82). Cambridge: MIT.
- Harris, S. (2010). *The moral landscape: How science can determine human values*. New York: Free Press.
- Helzer, E. G., & Pizarro, D. A. (2011). Dirty Liberals!: Reminders of physical cleanliness influence moral and political attitudes. *Psychological Science*, *22*, 517-522.
- Henrich, N., & Henrich, J. (2007). *Why humans cooperate: A cultural and evolutionary explanation*. New York: Oxford.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61-83.
- Herdt, G. H. (1981). *Guardians of the flutes*. New York: Columbia University Press.
- Hibbing, J. R., & Smith, K. B. (2007). The biology of political behavior. *The Annals of the American Academy of Political and Social Science*, *617*, 6-14.
- Hirschfeld, L. A. (2002). Why Don't Anthropologists Like Children? *American Anthropologist*, *104*, 611-627.
- Hirsh, J. B., DeYoung, C. G., Xu, X., & Peterson, J. B. (2010). Compassionate liberals and polite conservatives: Associations of agreeableness with political ideology and moral values. *Personality and Social Psychology Bulletin*, *36*, 655-664.
- Hodson, G., & Costello, K. (2007). Interpersonal disgust, ideological orientations, and dehumanization as predictors of intergroup attitudes. *Psychological Science*, *18*, 691-698.
- Hollos, M., Leis, P., & Turiel, E. (1986). Social reasoning in Ijo children and adolescents in Nigerian communities. *Journal of Cross-Cultural Psychology*, *17*, 352-374.
- Hong, S. (1996). Refinement of the Hong psychological reactance scale. *Educational and Psychological Measurement*, *56*, 173-182.



- Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A.B. (2009). Disgust and the moralization of purity. *Journal of Personality and Social Psychology*, 97(6), 963-976.
- Hrdy, S. B. (2009). *Mothers and others: The evolutionary origins of mutual understanding*. Cambridge, MA: Harvard.
- Hunter, J. D. (1991). *Culture wars: The struggle to define America*. New York: Basic Books.
- Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social-functional account of anger, disgust, and contempt. *Journal of Personality and Social Psychology*, 100, 719–737.
- Inbar, Y., Pizarro, D. A., & Bloom, P. (2009). Conservatives are more easily disgusted than liberals. *Cognition and Emotion*, 23, 714-725.
- Inbar, Y., Pizarro, D.A., & Bloom, P. (2011). Disgusting smells cause decreased liking of gay men. *Emotion*, 12, 23-27.
- Inbar, Y., Pizarro, D. A., & Cushman, F. (2012). Benefiting from misfortune: When harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin*, 38, 52–62.
- Inbar, Y., Pizarro, D., Iyer, R., & Haidt, J. (in press). Disgust sensitivity, political conservatism, and voting. *Social Psychological and Personality Science*.
- Inbar, Y., Pizarro, D. A., Knobe, J., & Bloom, P. (2009). Disgust sensitivity predicts intuitive disapproval of gays. *Emotion*, 9(3), 435.
- Iyer, R. (2009). What are the basic foundations of morality? Retrieved June 26, 2012 from <http://www.polipsych.com/2009/11/13/what-are-the-basic-foundations-of-morality/>
- Iyer, R. (2010). The Case for Honesty as a Moral Foundation. Retrieved June 26, 2012 from <http://www.polipsych.com/2010/12/07/the-case-for-honesty-as-a-moral-foundation/>.
- Iyer, R., Read, S. J. and Correia, J. (2010). Functional Justice: Productivity and Well-Being Goals Define Fairness. Available at SSRN: <http://ssrn.com/abstract=1691969> or <http://dx.doi.org/10.2139/ssrn.1691969>.
- Iyer, R., Graham, J., Koleva, S., Ditto, P., & Haidt, J. (2010). Beyond identity politics: Moral psychology and the 2008 Democratic primary. *Analyses of Social Issues and Public Policy*, 10, 293-306.
- Iyer, R., Koleva, S. P., Graham, J., Ditto, P. H., & Haidt, J. (in press). Understanding Libertarian morality: The psychological roots of an individualist ideology. *PLoS One*.
- James, W. (1890/1950). *The Principles of Psychology*. New York: Dover.
- James, W. (1909/1987). *A Pluralistic Universe*. New York: Library of America.
- Janoff-Bulman, R., Sheikh, S., & Baldacci, K (2008). Mapping moral motives: Approach, avoidance, and political orientation. *Journal of Experimental Social Psychology*, 44, 1091-1099.
- Janoff-Bulman, R., Sheikh, S., & Hepp, S. (2009). Proscriptive versus prescriptive morality: Two faces of moral regulation. *Journal of Personality and Social Psychology*, 96, 521-537.
- Janoff-Bulman, R., & Sheikh, S. (2012). The forbidden, the obligatory, and the permitted: Moral regulation and political orientation. Paper presented to the Society for Personality and Social Psychology annual conference, San Diego, CA.
- Jarudi, I.N. (2009). Everyday morality and the status quo: Conservative concerns about moral purity, moral evaluations of everyday objects, and moral objections to performance enhancement. (*Doctoral dissertation, Yale University*).

- Jones, B. (2012). The morality of representation: Constituent moral foundations and position-taking in Congress. Social Science Research Network.
- Joseph, C. (2002). Morality and the virtues in Islam. (*Doctoral Dissertation, University of Chicago*).
- Joseph, C. M., Graham, J., & Haidt, J. (2009). The end of equipotentiality: A moral foundations approach to ideology-attitude links and cognitive complexity. *Psychological Inquiry, 20*, 172-176.
- Jost, J.T. (2006). The end of the end of ideology. *American Psychologist, 61*, 651-670.
- Jost, J. T. (2009). Group morality and ideology: left and right, right and wrong. Paper presented to the Society for Personality and Social Psychology annual conference, Tampa, FL.
- Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin, 129*, 339.
- Joyce, R. (2006). *The evolution of morality*. The MIT Press.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Strauss, Giroux.
- Kesebir, P., & Pyszczynski, T. (2011). A Moral-Existential Account of the Psychological Factors Fostering Intergroup Conflict. *Social and Personality Psychology Compass, 5*, 878–890.
- Killen, M., & Smetana, J. G. (2006). *Handbook of moral development*. Mahwah, New Jersey: Erlbaum.
- Kim, K. R., Kang, J., & Yun, S. (in press). Moral intuitions and political orientation: Similarities and differences between Korea and the United States. *Psychological Reports*.
- Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 104*(30), 12577-12580. doi: 10.1073/pnas.0705345104
- Knutson, K., Krueger, F., Koenigs, M., Hawley, A., Escobedo, J., Vasudeva, V., Adolphs, R. & Grafman, J. (2009). Behavioral norms for condensed moral vignettes. *Social Cognitive and Affective Neuroscience, 5*(4), 378-384.
- 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: Rand McNally.
- Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), *Psychology and Genetic Epistemology* (pp. 151-235). New York: Academic Press.
- Kohlberg, L., Levine, C., & Hower, A. (1983). *Moral stages: A current formulation and a response to critics*. Basel, Switzerland: Karger.
- Koleva, S. (2011). *Birds of a Moral Feather: The Role of Morality in Romantic Attraction and Relationship Satisfaction*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (AAT 3472884).
- Koleva, S., Graham, J., Iyer, Y., 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*(2), 184-194.
- Koleva, S., & Haidt, J. (2012). Let's use Einstein's safety razor, not Occam's Swiss Army knife or Occam's chainsaw. Commentary on target article, Mind perception is the essence of morality (K. Gray, L. Young, & A. Waytz), *Psychological Inquiry, 23*, 175-178.
- Koonz, C. (2003). *The Nazi conscience*. Cambridge, MA: Belknap.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 480-498.

- Kurzban, R., Dukes, A., & Weeden, J. (2010). Sex, drugs and moral goals: reproductive strategies and views about recreational drugs. *Proceedings of Biological Sciences*, 277(1699), 3501-3508.
- Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology*, 68(1), 151.
- Lewin, K. (1951). *Field theory in social science; selected theoretical papers*. D. Cartwright, ed. New York: Harper & Row.
- Lewis, G. J., & Bates, T. C. (2011). From left to right: How the personality system allows basic traits to influence politics via characteristic moral adaptations. *British Journal of Psychology*, 102, 1-13.
- 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.
- LoBue, V., Chiong, C., Nishida, T., DeLoache, J., & Haidt, J. (2011). When getting something good is bad: Even 3-year-olds react to inequality. *Social Development*, 20(1), 154-170.
- Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., & Blair, R. J. R. (2006). The neural basis of implicit moral attitude—An IAT study using event-related fMRI. *Neuroimage*, 30, 1449-1457.
- Makiniemi, J., Pirttila-Backman, A., & Pieri, M. (in press). The endorsement of the moral foundations in food-related moral thinking in three European countries. *Journal of Agricultural and Environmental Ethics*.
- Marcus, G. (2004). *The birth of the mind*. New York: Basic.
- Marcus, G. (2008). *Kluge: The haphazard construction of the human mind*. Boston: Houghton Mifflin.
- Marler, P. (1991). The instinct to learn. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition*. Mahwah, NJ: Erlbaum.
- Marques, J. M., Yzerbyt, V. Y., & Leyens, J. P. (1988). The 'black sheep effect': Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology*, 18, 1-16.
- McAdams, D. P. (1995). What do we know when we know a person? *Journal of Personality*, 63, 365-396.
- McAdams, D., Albaugh, M., Farber, E., Daniels, J., Logan, R., & Olson, B. (2008). Family metaphors and moral intuitions: How conservatives and liberals narrate their lives. *Journal of Personality and Social Psychology*, 95(4), 978-990.
- McAdams, D. P., & Pals, J. L. (2006). A new Big Fig: Fundamental principles for an integrative science of personality. *American Psychologist*, 61, 204-217.
- McClosky, H., & Zaller, J. (1984). *The American ethos: Public attitudes toward capitalism and democracy*. Harvard University Press Cambridge, MA.
- McGregor, H. A., Lieberman, J. D., Greenberg, J., Solomon, S., Arndt, J., Simon, L., & Pyszczynski, T. (1998). Terror management and aggression: evidence that mortality salience motivates aggression against worldview-threatening others. *Journal of Personality and Social Psychology*, 74(3), 590.
- Mercier, H., & Sperber, D. (2010). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34, 57-74.

- Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in cognitive sciences*, 11(4), 143–152.
- Mineka, S., & Cook, M. (1988). Social learning and the acquisition of snake fear in monkeys. In T. R. Zentall & J. B. G. Galef (Eds.), *Social learning: Psychological and biological perspectives* (pp. 51-74). Hillsdale, N. J.: Lawrence Erlbaum.
- Motyl, M. (2012). *How moral migration geographically polarizes the electorate*. Invited talk given at the Society for Experimental Social Psychology's annual conference. Austin, TX.
- Motyl, M., & Pyszczynski, T. (2009). The existential underpinnings of the cycle of violence and terrorist and counterterrorist pathways to peaceful resolutions. *International Review of Social Psychology*, 22, 267-291.
- Narvaez, D. (2008). The social-intuitionist model: Some counter-intuitions. In W. A. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 2, The Cognitive Science of Morality: Intuition and Diversity* (pp. 233-240). Cambridge, MA: MIT Press.
- Narvaez, D. (2010). Moral complexity: The fatal attraction of truthiness and the importance of mature moral functioning. *Perspectives on Psychological Science*, 5, 163-181.
- Navarrete, C. D., & Fessler, D. M. T. (2006). Disease avoidance and ethnocentrism: the effects of disease vulnerability and disgust sensitivity on intergroup attitudes. *Evolution and Human Behavior*, 27, 270-282.
- Neuberg, S. L., Kenrick, D. T., & Schaller, M. (2010). Evolutionary social psychology. In S. T. Fiske, D. T. Gilbert, and G. Lindzey (Eds.), *Handbook of social psychology* (5th Edition, pp. 761-796). New York: John Wiley & Sons.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259.
- Nosek, B. A., Graham, J., & Hawkins, C. B. (2010). Implicit Political Cognition. In B. Gawronski & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition* (pp 548-564). New York, NY: Guilford.
- Nucci, L., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development*, 49, 400-407.
- Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease avoidance mechanism. *Psychological Bulletin*, 135, 303-321.
- Oishi, S., & Graham, J. (2010). Social ecology: Lost and found in psychological science. *Perspectives on Psychological Science*, 5, 356-377.
- Opie, I., & Opie, P. (1969). *Children's Games in Street and Playground*. Oxford, UK: Clarendon Press.
- Oxley, D. (2010). Fairness, justice and an individual basis for public policy. (*Doctoral dissertation, University of Nebraska*).
- Oxley, D. R., Smith, K. B., Alford, J. R., Hibbing, M. V., Miller, J. L., Scalora, M., . . . Hibbing, J. R. (2008). Political Attitudes Vary with Physiological Traits. *Science*, 321, 1667-1670.
- Parkinson, C., Sinnott-Armstrong, W., Koralus, P. E., Mendelovici, A., McGeer, V., & Wheatley, T. (2011). Is morality unified? Evidence that distinct neural systems underlie judgments of harm, dishonesty, and disgust. *Journal of Cognitive Neuroscience*, 23, 3162–3180.
- Pasanek, B. (2009). Ideology and metaphor. Presented to Kellogg School of Management, September 2009.

- Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, 89(3), 277.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2003). *Linguistic inquiry and word count: LIWC2001 manual*. Mahwah, NJ: Erlbaum.
- Piaget, J. (1932/1965). *The Moral Judgment of the Child*. (M. Gabain, Trans.). New York: Free Press.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York: Viking.
- Pinker, S. (2012). *The false allure of group selection*. Retrieved July 20, 2012, from <http://edge.org/conversation/the-false-allure-of-group-selection>.
- Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25, 1-72.
- Preston, J. L. & Ritter, R. S. (in press). Cleanliness and godliness: Mutual association between two forms of purity. *Journal of Experimental Social Psychology*.
- Prince, C. (2010). Moral foundation theory and the law. *Seattle University Law Review*, 33, 1293-1317.
- Pyszczynski, T., Motyl, M., & Abdollahi, A. (2009). Righteous violence: Killing for god, country, freedom, and justice. *Behavioral Sciences of Terrorism and Political Aggression*, 1, 12-39.
- Rai, T. S., & Fiske A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, 118, 57-75.
- Rest, J., Narvaez, D., Thoma, S. J., Bebeau, M. J. (1999). DIT2: Devising and testing a new instrument of moral judgment. *Journal of Educational Psychology*, 91, 644-659.
- Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.
- Ridley, M. (1996). *The origins of virtue*. Harmondsworth, UK: Penguin.
- Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., & Kilts, C. (2002). A neural basis for social cooperation. *Neuron*, 35, 395-405.
- Rosenblatt, A., Greenberg, J., Solomon, S., Pyszczynski, T., & Lyon, D. (1989). Evidence for terror management theory: I. The effects of mortality salience on reactions to those who violate or uphold cultural values. *Journal of Personality and Social Psychology*, 57, 681.
- Rozin, P., Haidt, J., McCauley, C.R. (2000). Disgust. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp 637-653).
- Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review*, 5, 2-14.
- Rozin, P., Haidt, J., & McCauley, C. R. (2008). Disgust. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions*, 3rd ed. (pp. 757-776). New York: Guilford Press.
- Rozin, P., & Royzman, E. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5, 296-320.
- Rozin, P., & Singh, L. (1999). The moralization of cigarette smoking in the United States. *Journal of Consumer Psychology*, 8(3), 321-337.
- Russell, P.S., & Giner-Sorolla, R. (2011). Moral anger is more flexible than moral disgust. *Social Psychological and Personality Science*, 2(4), 360-364.
- Sadler-Smith, E.R. (in press). Before virtue: biology, brain, behavior, and the 'moral sense'.

*Business Ethics Quarterly.*

- Saltzstein, H. D., & Kasachkoff, T. (2004) Haidt's moral intuitionist theory: A psychological and philosophical critique. *Review of General Psychology*, 8, 273-282.
- Samuelson, S. (1980). The cooties complex. *Western Folklore*, 39, 198-210.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300, 1755-1758.
- Schaller, M., & Park, J. H. (2011). The behavioral immune system (and why it matters). *Current Directions in Psychological Science*, 20, 99-103.
- Schmidt, M. F. H., & Sommerville, J. A. (2011). Fairness Expectations and Altruistic Sharing in 15-Month-Old Human Infants. *PLoS ONE*, 6, e23223. doi:10.1371/journal.pone.0023223
- Schnall, S. & Cannon, P.R. (in press). The clean conscience at work: Emotions, intuitions and morality. *Journal of Management, Spirituality & Religion*.
- 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. (1992). Universals in the content and structure of values. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, Vol. 25 (pp. 1-65). New York: Academic Press.
- 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.
- Sherif, M., Harvey, O. J., White, B. J., Hood, W., & Sherif, C. (1961/1954). *Intergroup conflict and cooperation: The Robbers Cave experiment*. Norman, OK: University of Oklahoma Institute of Group Relations.
- Shook, N. J., & Fazio, R. H. (2009). Political ideology, exploration of novel stimuli, and attitude formation. *Journal of Experimental Social Psychology*, 45(4), 995-998.
- Shweder, R. A. (1990). In defense of moral realism: Reply to Gabennesch. *Child Development*, 61, 2060-2067.
- 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: 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.
- Simon, H. (1992). What is an "explanation" of behavior? *Psychological Science*, 3, 150-161.
- Skitka, L. J., Mullen, E., Griffin, T., Hutchinson, S., & Chamberlin, B. (2002). Dispositions, scripts, or motivated correction?: Understanding ideological differences in explanations for social problems. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology*, 83(2), 470.
- Sloane, S., Baillargeon, R., & Premack, D. (2012). Do Infants Have a Sense of Fairness? *Psychological Science*, 23, 196-204.
- Smith, K. B., Oxley, D. R., Hibbing, M. V., Alford, J. R., & Hibbing, J. R. (2011). Linking Genetics and Political Attitudes: Re-Conceptualizing Political Ideology. *Political Psychology*, 32, 369-397.

- Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 39-67). Cambridge, UK: Cambridge University Press.
- Sperber, D. (2005). Modularity and relevance: How can a massively modular mind be flexible and context-sensitive? In P. Carruthers, S. Laurence, & S. P. Stich (Eds.), *The Innate Mind: Structure and Contents, Volume 1* (pp. 53-68). New York: Oxford University Press.
- Spiro, M. (1956). *Kibbutz: Venture in utopia*. Cambridge MA: Harvard.
- Spelke, E. S. (2000). Core knowledge. *American Psychologist*, *55*, 1233-1243.
- Stanovich, W., & West, R. F. (2000). Individual difference in reasoning: implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645-726.
- Stenner, K. (2005). *The authoritarian dynamic*. New York: Cambridge.
- Suhler, C. L., & Churchland, P. (2011). Can innate, modular “foundations” explain morality? Challenges for Haidt’s moral foundations theory. *Journal of Cognitive Neuroscience*, *23*(9), 2103-2116.
- Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, *19*, 339-347.
- Tamborini, R. (2011). Moral intuition and media entertainment. *Journal of Media Psychology: Theories, Methods, and Applications*, *23*, 39-45.
- Tamborini, R., Eden, A., Bowman, N.D., Grizzard, M., & Lachlan, K.A. (2012). The influence of morality subcultures on the acceptance and appeal of violence. *Journal of Communication*, *62*(1), 136-157.
- Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology*, *47*, 1249-1254.
- Tetlock, P. E., Kristel, O. V., Elson, B., Green, M., & Lerner, J. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*, 853-870.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675-735.
- Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19-136). New York: Oxford.
- Tooby, J., Cosmides, L., & Barrett, H. C. (2005). Resolving the debate on innate ideas: Learnability constraints and the evolved interpenetration of motivational and conceptual functions. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The innate mind: Structure and contents* (pp. 305-337). New York: Oxford.
- Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35-57.
- Turiel, E. (1979). Distinct conceptual and developmental domains: Social-convention and morality. *Nebraska Symposium on Motivation*. Lincoln, NE: University of Nebraska Press.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, England: Cambridge University Press.

- Turiel, E., Killen, M., & Helwig, C. C. (1987). Morality: Its structure, function, and vagaries. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 155-243). Chicago: University of Chicago Press.
- Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, 9, 160-164.
- Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science*, 20(9), 1092-1099.
- Van Leeuwen, F. & Park, J.H. (2009). Perceptions of social dangers, moral foundations, and political orientation. *Personality and Individual Differences*, 47, 169-173.
- Van Leeuwen, F., Park, J. H., Koenig, B. L., & Graham, J. (in press). Regional variation in pathogen prevalence predicts endorsement of group-focused moral concerns. *Evolution and Human Behavior*.
- Van Vugt, M., & Park, J. H. (2009). Guns, germs, and sex: how evolution shaped our intergroup psychology. *Social and Personality Psychology Compass*, 3(6), 927-938.
- Vauclair, C., & Fischer, R. (2011). Do cultural values predict individuals' moral attitudes? A cross-cultural multilevel approach. *European Journal of Social Psychology*, 41(5), 645-657.
- Walster, E., Walster, G., & Berscheid, E. (1978). *Equity: Theory and research*. Boston: Allyn & Bacon.
- Waytz, A., Epley, N., & Cacioppo, J. T. (2010). Social cognition unbound. *Current Directions in Psychological Science*, 19(1), 58.
- Weber, C., & Federico, C. M. (2007). Interpersonal attachment and patterns of ideological belief. *Political Psychology*, 28(4), 389-416.
- Weber, C., & Federico, C. M. (in press). Moral foundations and heterogeneity in ideological preferences. *Political Psychology*.
- Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed, Vol. 1, pp. 446-496). New York: McGraw-Hill.
- Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, 16, 780-784.
- Wiessner, P. (2005). Norm enforcement among the Ju/'hoansi Bushmen. *Human Nature*, 16, 115-145.
- Wilson, D. S. (2002). *Darwin's cathedral: Evolution, religion, and the nature of society*. Chicago: University of Chicago Press.
- Winegard, B., & Deaner, R. O. (2010). The evolutionary significance of Red Sox nation: Sport fandom as a byproduct of coalitional psychology. *Evolutionary Psychology*, 8(3), 432-446.
- Winterich, K., Zhang, Y., Mittal, V. (in press). How Political Identity and Charity Positioning Increase Donations: Insights from Moral Foundations Theory. *International Journal of Research in Marketing: Special Issue on Consumer Identities*.
- Wright, J. C., & Baril, G. (2011). The role of cognitive resources in determining our moral intuitions: Are we all liberals at heart? *Journal of Experimental Social Psychology*, 47, 1007-1012.
- Wright, R. (1994). *The moral animal*. New York: Pantheon.



- Young, L., & Saxe, R. (2009). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia*, *47*(10), 2065–2072.
- Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*, 202–214.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 151-175.

### **Acknowledgement**

The authors wish to thank Trish Devine, Jeremy Frimer, Ashby Plant, Linda Skitka, and the USC Values, Ideology, and Morality Lab for helpful comments on a draft of this article.

**Table 1.** The original five foundations of intuitive ethics. (Adapted from Haidt, 2012).

<b>Foundation:</b>	<b>Care/ harm</b>	<b>Fairness/ cheating</b>	<b>Loyalty/ betrayal</b>	<b>Authority/ subversion</b>	<b>Sanctity/ degradation</b>
<b>Adaptive challenge</b>	Protect and care for children	Reap benefits of two-way partnerships	Form cohesive coalitions	Forge beneficial relationships within hierarchies	Avoid communicable diseases
<b>Original triggers</b>	Suffering, distress, or neediness expressed by one's child	Cheating, cooperation, deception	Threat or challenge to group	Signs of high and low rank	Waste products, diseased people
<b>Current triggers</b>	Baby seals, cute cartoon characters	Marital fidelity, broken vending machines	Sports teams, nations	Bosses, respected professionals	Immigration, deviant sexuality
<b>Character-istic emotions</b>	Compassion for victim; anger at perpetrator	anger, gratitude, guilt	Group pride, rage at traitors	Respect, fear	Disgust
<b>Relevant virtues</b>	Caring, kindness	Fairness, justice, trustworthiness	Loyalty, patriotism, self-sacrifice	Obedience, deference	Temperance, chastity, piety, cleanliness

**Table 2.** Methods developed to measure MFT's constructs.

<b><u>Method:</u></b>	<b><u>Description:</u></b>	<b><u>Reference:</u></b>
<i>Self-report scales:</i>		
Moral Foundations Questionnaire	Ratings of the moral relevance of foundation-related considerations (part 1); agreement with statements supporting or rejecting foundation-related concerns	Graham, Nosek, Haidt, Iyer, Koleva, & Ditto, 2011
Moral Foundations Sacredness Scale	Reports of how much one would need to be paid to violate the foundations in different ways (including an option to refuse the offer for any amount of money)	Graham & Haidt, 2011
<i>Implicit measures:</i>		
Evaluative priming	Foundation-related vice words (hurt, cruel, cheat, traitor, revolt, sin) used as primes flashed for 150ms before positive or negative adjective targets	Graham, 2010 (adapted from Ferguson, 2007)
Affect Misattribution Procedure	Pictures representing foundation-related virtues and vices flashed for 150ms before Chinese characters, which participants rate as more or less positive than other characters	Graham, 2010 (adapted from Payne et al., 2005)
Foundation Tradeoff Task	Quick dichotomous responses to "which is worse?" task pitting foundation violations against each other	Graham, 2010
<i>Psychophysiological and neuroscience methods:</i>		
Facial electromyography	Measurement of affective micro-expressions while hearing sentences describing actions supporting or violating foundations	Cannon, Schnall, & White, 2011
Time-specified stimuli for psychophysiological studies	Sentences presented one word at a time, with critical word indicating moral opinion supporting or rejecting a foundation	Graham, 2010 (adapted from van Berkum et al., 2009)
Neuroimaging vignettes	Scenarios describing possible violations of Care (assault) or Sanctity (incest), varying intent and outcome, for use in fMRI studies	Young & Saxe, 2011
<i>Text analysis:</i>		
Moral Foundations Dictionary	Dictionary of foundation-related virtue and vice words, for use with Linguistic Inquiry and Word Count program (Pennebaker et al., 2003)	Graham, Haidt, & Nosek, 2009

**Table 3.** Foundation correlations with political ideology by world area. (Adapted from Graham et al., 2011).

	<u>USA</u>	<u>UK</u>	<u>Canada</u>	<u>Australia</u>	<u>Western</u> <u>Europe</u>	<u>Eastern</u> <u>Europe</u>	<u>Latin</u> <u>America</u>	<u>Africa</u>	<u>Middle</u> <u>East</u>	<u>South</u> <u>Asia</u>	<u>East</u> <u>Asia</u>	<u>Southeast</u> <u>Asia</u>	<u>Average</u>
<i>N</i>	80,322	2,579	4,314	1,563	3,766	888	1,345	153	575	884	479	550	
Harm	-0.35	-0.25	-0.31	-0.28	-0.22	-0.17	-0.16	-0.04	-0.19	-0.14	-0.19	-0.12	<b>-0.20</b>
Fairness	-0.44	-0.40	-0.36	-0.38	-0.33	-0.28	-0.33	-0.35	-0.32	-0.21	-0.24	-0.29	<b>-0.32</b>
Ingroup	0.47	0.42	0.34	0.44	0.35	0.35	0.32	0.39	0.42	0.33	0.28	0.33	<b>0.37</b>
Authority	0.56	0.51	0.50	0.53	0.47	0.43	0.46	0.51	0.56	0.42	0.37	0.48	<b>0.48</b>
Purity	0.58	0.46	0.47	0.52	0.46	0.45	0.48	0.51	0.51	0.45	0.42	0.49	<b>0.49</b>

**Table 4.** Criteria for foundationhood, with evidence for the current foundations.

<b>Foundation criteria:</b>	<b>Care/harm</b>	<b>Fairness/cheating</b>	<b>Loyalty/Betrayal</b>	<b>Authority/subversion</b>	<b>Sanctity/degradation</b>
Criterion 1: Common in third-party normative judgments	Playground harm: Nucci & Turiel (1978)	Catching cheaters: Dunbar (1996)	The Black Sheep effect: Marques, Yzerbyt, & Leyens (1988)	Disrespect for authority: Shweder, Miller, & Mahapatra, 1987	Food and sex taboos: Haidt, Koller, & Dias (1993)
Criterion 2: Automatic affective evaluations	To cruelty and violence: Luo et al (2006), Cannon et al. (2011), Graham (2010)	To cheating: Sanfey et al (2003); to unfairness or inequality: Cannon et al. (2011), Graham (2010)	To ingroup betrayals: Cannon et al. (2011), Graham (2010)	To subversion: Cannon et al. (2011), Graham (2010)	To sexual violations: Parkinson et al. 2011; to degradation: Cannon et al. (2011), Graham (2010)
Criterion 3: Culturally widespread	Bowlby (1969)	Fiske (1992)	Herdt (1981)	Fiske (1992)	Douglas (1966)
Criterion 4: Evidence of innate preparedness	NHP: Hrdy (2009); Preston & de Waal (2002); Infants: Hamlin, Wynn, & Bloom (2006)	NHP: Brosnan (2006); Infants: Schmidt & Sommerville (2011); Sloane, Baillargeon, & Premack (2012)	NHP: De Waal (1982); Infants: Kinzler, Dupoux & Spelke (2007); Hamlin et al. (2012)	NHP: Boehm (1999, 2012); Not yet shown in infants	Not yet shown in NHP or infants
Criterion 5: Evolutionary model	Kin selection: Hamilton (1964); Attachment theory: Bowlby, (1969)	Reciprocal Altruism: Trivers (1971)	Multi-level selection: D. S. Wilson (2002); Tribalism: Richerson & Boyd (2005)	Rank and dominance: de Waal (1982); Boehm (1999)	Disgust: Rozin, Haidt & McCauley (2008); Behavioral immune system: Schaller & Park (2011)

*Note.* NHP = non-human primates.

Figure 1. Ideological differences in foundation endorsement across three methods. (Adapted from Graham, Haidt, & Nosek, 2009).

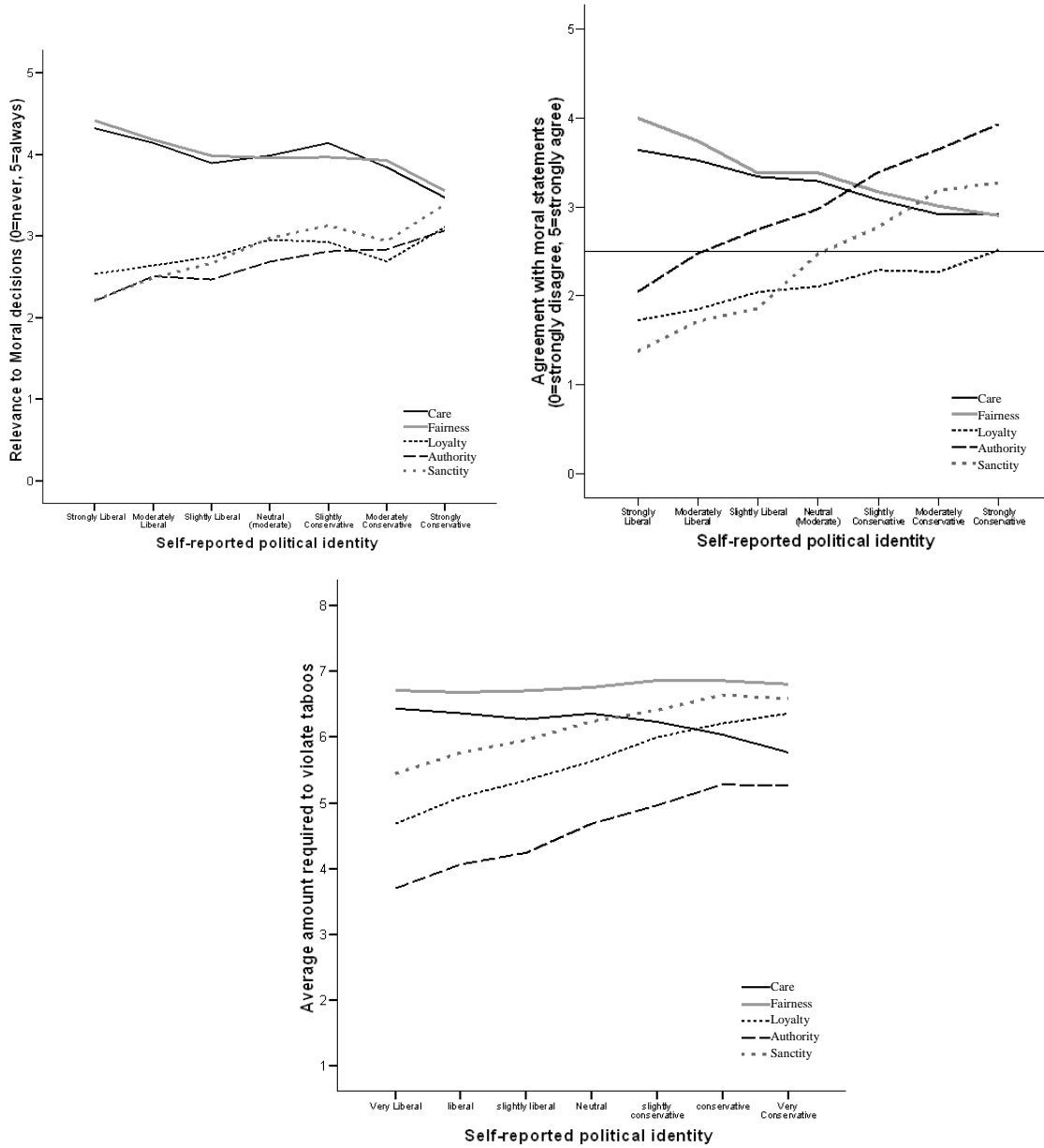
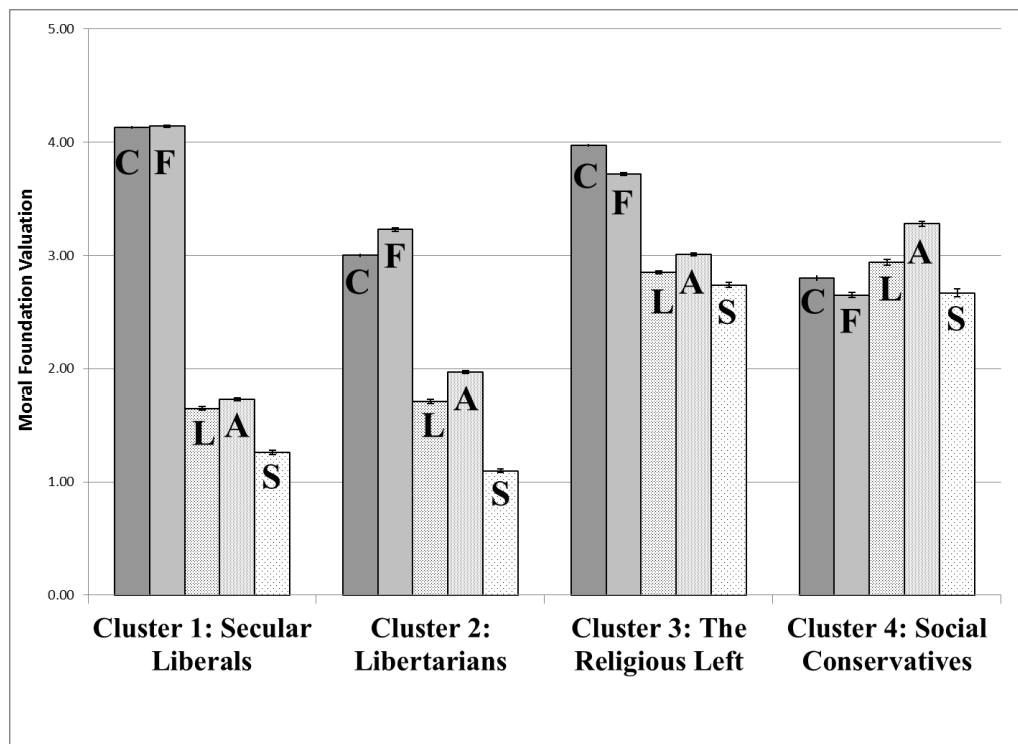


Figure 2. Cluster analysis of Moral Foundations Questionnaire responses. (Adapted from Haidt, Graham, & Joseph, 2009).



*Note.* C=Care, F=Fairness, L=Loyalty, A=Authority, S=Sanctity. Ns for each cluster are as follows: 5946 (cluster 1), 5931 (cluster 2), 6397 (cluster 3), 2688 (cluster 4). Error bars represent +/- 2 S.E.

Figure 3. Moral concerns of libertarians as compared to liberals and conservatives (Adapted from Iyer et. al., in press).

