CHAPTER TWO

Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism

Jesse Graham[*], Jonathan Haidt[†], Sena Koleva[*], Matt Motyl[‡], Ravi Iyer[*], Sean P. Wojcik[§], Peter H. Ditto[§]

[*]Department of Psychology, University of Southern California, Los Angeles, California, USA
[†]Stern School of Business, New York University, New York, USA
[‡]Department of Psychology, University of Virginia, Charlottesville, Virginia, USA
[§]School of Social Ecology, University of California, Irvine, California, USA

Contents

Advances in Experimental Social Psychology, Volume 47
ISSN 0065-2601
http://dx.doi.org/10.1016/B978-0-12-407236-7.00002-4

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 moral psychology.

> *"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 overvaluation of parsimony.

In this chapter, 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, 2013). 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 will 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." 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 are 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 will 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 will provide an overview of empirical results that others and we 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 will discuss criticisms of the theory and future research directions that are motivated in part by those criticisms. We will 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 et al., 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 do not know how many moral foundations there really are. There may be 74, or perhaps 122, or 27, or maybe only 5, 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 MFT

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; Joyce, 2006; Ridley, 1996; 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 overvalue conformity and tradition. But if children get more opportunities to role-take, they will progress to the postconventional level (stages 5 and 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 and an ethic of care that could not be derived from the former. Kohlberg eventually acknowledged that she was right (Kohlberg, Levine, & Hewer, 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 are up to three. Or maybe it's four? 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 could not 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.[1] We believe that there are more than five; for example, Haidt

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

(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 will 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

MFT 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 MFT was created precisely to integrate the two (see also Fiske, 1991; Richerson & Boyd, 2005). Our definition of nativism makes this clear: Innate means *organized in advance of experience*. We do 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 and 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 cannot 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 is 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 (Pinker, 1997; Tooby & Cosmides, 1992). 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 past half-million years (see Haidt, 2012, chapter 9, which draws on Richerson & Boyd, 2005; Tomasello, Carpenter, Call, Behne, & Moll, 2005).

MFT 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 2.1). 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 do not 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 would 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., Pinker, 2002; Spiro, 1956). Conversely, if cultural learning played no formative role, then the first draft would be the final draft, and there would be no variation across cultures.[2] 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

[2] Other than those due to individual development, for example, 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.

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. 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 had 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 cannot 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 cannot see the foundations directly, and you cannot 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 will 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 (Kahneman, 2011; Stanovich & West, 2000; 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 esthetic 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 SIM. 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, 2010), and people think and act more like intuitive lawyers than intuitive scientists (Baumeister & Newman, 1994; Ditto et al., 2009; Haidt, 2007a, 2007b, 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 (Fiske, 1992; James, 1890/1950), and so innate mental structures, such as the moral foundations, are likely[3] to be responses to adaptive challenges that faced our ancestors for a very long time.

Table 2.1 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"[4] 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 will use the equivalent terms offered by Haidt (2012): the *original triggers* and the *current triggers*.

We will explain the first column—the Care/harm foundation, in some detail, to show how to read the table. We will 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 will 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,

[3] Spandrels aside (Gould & Lewontin, 1979).

[4] Evolution *is* a design process; it is just not an intelligent design process. See Richerson and Boyd (2005).

Table 2.1 The original five foundations of intuitive ethics

Foundation	Care/harm	Fairness/cheating	Loyalty/betrayal	Authority/subversion	Sanctity/degradation
Adaptive challenge	Protect and care for children	Reap benefits of two-way partnerships	Form cohesive coalitions	Forge beneficial relationships within hierarchies	Avoid communicable diseases
Original triggers	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
Current triggers	Baby seals, cute cartoon characters	Marital fidelity, broken vending machines	Sports teams, nations	Bosses, respected professionals	Immigration, deviant sexuality
Characteristic emotions	Compassion for victim; anger at perpetrator	Anger, gratitude, guilt	Group pride, rage at traitors	Respect, fear	Disgust
Relevant virtues	Caring, kindness	Fairness, justice, trustworthiness	Loyalty, patriotism, self-sacrifice	Obedience, deference	Temperance, chastity, piety, cleanliness

Adapted from Haidt (2012).

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, baby animals (which often share the proportions of children), stuffed animals and cartoon characters that are deliberately crafted to have the proportions of children, and 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.

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 nonzero-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 cannot 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.[5] Sherif, Harvey, White, Hood, & Sherif (1961/1954) classic Robber's Cave study activated (and then deactivated) 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 upward and downward have an advantage over those who fail to perceive or react appropriately in these complex social interactions (de Waal, 1982; 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 United States—but can be seen as neutral or even as vices in others—such as among social liberals (Frimer, Biesanz, Walker, & MacKinlay, in press; Haidt & Graham, 2009; Stenner, 2005).

[5] 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), Chapter 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, for example, to immigrants and sexual deviants (Faulkner, Schaller, Park, & Duncan, 2004; Navarrete & Fessler, 2006; Rozin et al., 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-developmentalist, 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 will 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 will show how our initial five foundations have been measured and used in psychological studies.

3. EMPIRICAL FINDINGS

In this chapter, 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 is 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,