

**The moral mind: how 5 sets of innate moral intuitions guide the development of many culture-specific virtues, and perhaps even modules**

Craig Joseph  
University of Chicago

Jonathan Haidt  
University of Virginia

May 11, 2006

Invited Submission to P. Carruthers, S. Laurence, and S. Stich (Eds.)  
*The Innate Mind, Vol. 3.*

## **1. Introduction**

Morality is one of the few topics in academe endowed with its own protective spell. A biologist is not blinded by her biological nature to the workings of biology. An economist is not confused by his own economic activity when he tries to understand the workings of markets. But students of morality are often biased by their own moral commitments. Morality is so contested, and the outcomes of those contests are so important to the researchers themselves, that it is often difficult for researchers to set aside their humanity and study morality in a clinically detached way.

One problem is that the psychological study of morality and moral development has been dominated by politically liberal researchers (which includes us). This lack of moral and political diversity among morality researchers has led to an inappropriate narrowing of the moral domain to issues of harm, rights, and justice. Morality in most cultures (and for political conservatives in our own), is in fact much broader, including issues of ingroup/outgroup, hierarchy/authority, and purity/pollution (Haidt & Graham, in press).

This article is about how morality might be partially innate, by which we simply mean organized, to some extent, in advance of experience (Marcus, 2004). We begin by arguing for a broader conception of morality and suggesting that most of the discussion of the innateness to date has not been about morality per se; it has been about whether the psychology of *harm and reciprocity* is innate. Once we have made our case that morality involves five domains, not two, we turn our attention to the ways in which this diverse collection of motives and concepts might be innate. We consider five hypotheses about the origins of moral knowledge and value, and we endorse one of them (a form of flexible and generative modularity) as being the best candidate. Next, we develop this version of modular morality by describing how the innately specified “first draft” gets

modified during development. We discuss virtue theory, an ancient approach that integrates the insights of many modern perspectives. Finally, we discuss the importance of narrative, so often heralded by anthropologists and others who are wary of nativism and reductionism. We show how a narrative approach to morality fits well with the nativist “five foundations” view we developed in the first part of the paper.

## **2. Morality is many things**

Almost as soon as people began to write, they began writing about morality. Many of the earliest moral texts are largely lists of laws and prohibitions (e.g., the Code of Hammurabi; the older parts of the Old Testament). But as the Axial Age progressed (800 BCE – 200 BCE), many cultures, East and West, began to develop a more sophisticated psychology of the virtues. We begin to find more explicit discussions of virtues, often in the context of stories about role models who exemplified them (e.g., Homer and Aesop in Greece; the Mahabharata in India). An important feature of this approach is that moral education is accomplished by shaping emotions and intuitions, rather than by dictating explicit rationales or principles. The wisdom of Confucius and of Buddha, for example, comes down to us as lists of aphorisms and metaphors that produce flashes of intuitive understanding.

A second feature of these virtue-based approaches is that they emphasize practice and habit rather than propositional knowledge and reasoning. Buddha urged his disciples to follow the Eightfold Noble Path – a set of daily practices---to reach moral and psychological perfection. Aristotle and Confucius both compared the development of virtue to the slow practice needed to develop what we now call “virtuosity” on a musical instrument (Aristotle, 1941; Hansen, 1991).

For the ancients there were many virtues, covering most aspects of human activity. Virtues were excellences that people were expected to cultivate in themselves, depending on their social roles and stations in life. Two of the greatest thinkers in ancient Greek philosophy - Plato and Aristotle - conducted much of their inquiries into ethics via examination of the concept of virtue and the individual virtues, although they had very different notions of what virtues were, what grounded them, and how they were acquired.

### *2.1. Quandary ethics and the great narrowing*

The idea that morality is a set of virtues to be cultivated through practice remained the dominant approach throughout the world until at least the time of the Middle Ages. St. Thomas Aquinas followed Aristotle in ethics as in other things, and even Islamic thinkers, such as Miskawayh and al-Ghazali, borrowed from Aristotle in constructing their theories of morality. Even up to the middle of the twentieth century, influential philosophers and psychologists (Dewey, 1922; Hartshorne and May, 1928) continued to assume the essential validity of virtue theory and to base empirical research programs on the assumption that virtues were psychologically real and served to organize much of moral life.

But Western philosophers' ideas about morality began to change in the 18<sup>th</sup> century. For the most part, virtue- and religiously-based moralities are characterized by specific, substantive beliefs and commitments, "thick" ideas about human nature and society. With the Enlightenment, those assumptions came under increasing scrutiny, and philosophers began to search for groundings for moral judgment that did not depend upon specific metaphysical beliefs or group identities. What MacIntyre (1981) has called "the Enlightenment project" was the attempt to ground morality in highly abstract, even logical truths, and to disengage it (especially) from religious belief. Two types of alternatives emerged that are of continuing relevance today: *formalist* theories and *consequentialist* theories. Formalist theories of ethics, of which Kant is the best-known example, define moral judgments by reference to their logical form, for example as maxims or prescriptive judgments, rather than by their content. The moral status of an action is judged by reference to the kind of norm that underlies it. Consequentialist theories, including especially utilitarianism, ground moral judgments in pre-moral assessments of the consequences of actions; the morally right thing to do is defined, fundamentally, as the thing that will have the best consequences (however that very important phrase is understood).

Despite their differences – and they are great – both formalist and consequentialist approaches to morality seek to detach moral judgment as much as possible from moral content. Formalism replaces substantive moral judgment with a logical rationality, while consequentialism replaces it with a calculative rationality. Both approaches privilege parsimony: moral decisions should be made with respect to a foundational principle, such as the categorical imperative or the maximization of utility. Both insist that moral decisions should be governed by reason and logic, not emotion and intuition. And they both devalue the particular in favor of the abstract.

The commonalities between these two approaches to ethics have led to a modern consensus about the scope of ethical inquiry: morality is about resolving dilemmas involving the competing interests of people. The philosopher Edmund Pincoffs (1986) called this modern approach "quandary ethics," and he laments the loss of the older philosophical interest in virtue. Where the Greeks focused on *character* and asked what kind of person we should each *become*, modern ethics focuses on *actions*, trying to determine which ones are right and wrong.

Nevertheless, quandary ethics has continued to flourish in philosophy and in psychology, where it has guided the operationalization of morality. Lawrence Kohlberg's (1969) pioneering method was the longitudinal study of how children resolve moral dilemmas: should Heinz steal the drug to save his dying wife? Kohlberg's conclusion was that children get progressively better at quandary ethics, until they reach the highest stage, at which all decisions are made by reference to universally applicable, self-constructed, and non-consequentialist principles. Carol Gilligan (1982) challenged Kohlberg's conclusions by using a different dilemma: she interviewed women facing the quandary of an unwanted pregnancy, and she offered a competing highest principle: care. Social psychologists have also operationalized morality as quandary, putting research subjects into situations where they must weigh their self interest against the needs of a stranger

(e.g., the “good Samaritan” study, Darley & Batson, 1973; empathy-altruism research: Batson et al., 1983; moral hypocrisy studies: Batson et al., 1997). Baron (1993) has straightforwardly declared that consequentialism is the correct understanding of morality, and much of the research done in connection with his approach involves presenting subjects with tradeoffs between decision alternatives, each of which has costs and benefits. And when moral philosophers conduct experiments, as they are beginning to do, they experiment primarily on quandaries such as trolley and lifeboat problems that pit utilitarian and deontological concerns against each other (Greene et al., 2001; Petrinovich, O’Neill, & Jorgensen, 1993).

Even when research methods have not used quandaries per se, they have adopted the implicit boundary condition of quandary ethics: moral issues are those that pertain to the rights and welfare of individuals. Morality is about helping and hurting people. Elliot Turiel, a student of Kohlberg, codified this individual-centered view of morality in his influential definition of the moral domain as

prescriptive judgments of justice, rights, and welfare pertaining to how people ought to relate to each other. Moral prescriptions are not relative to the social context, nor are they defined by it. Correspondingly, children’s moral judgments are not derived directly from social institutional systems but from features inherent to social relationships -- including experiences involving harm to persons, violations of rights, and conflicts of competing claims. (Turiel, 1983, p.3)

Turiel’s delimiting of the moral domain seems obviously valid to many people, especially in modern Western (and Westernizing) cultures. However, the fact remains that for many people, in many cultures (including Western ones), the definition does not capture all that is included in the moral domain. In other words, Turiel’s definition (we are asserting) is inadequate as an inductive generalization, and given that, its force becomes more semantic than empirical, an *a priori* definition. And when the moral domain is *defined* in terms of “justice, rights, and welfare,” then the psychology that emerges cannot be a true psychology of morality; it can only be a psychology of perceptions of and judgments about justice, rights, and welfare. And when the domain of morality is narrowed in this way, then overly parsimonious explanations of it become tempting: morality can be explained evolutionarily as the extension of kin-altruism plus reciprocal altruism out to larger groups than those in which we evolved. And morality can be explained developmentally as the progressive extension of the child’s understanding that harming others (which includes treating them unfairly, unreciprocally) is bad.

But what if there is more to morality than harm, rights, and justice? What if these concerns are part of a bigger and more complicated story that can’t be handled so parsimoniously? Might explanations of the origins and development of “morality” have been formulated prematurely?

### 2.3, *The rebirth of breadth*

One of the distinctions that has been most important in the study of morality, but also most problematic, is that between “moral” and “conventional” judgments. As we have just mentioned, Turiel (and cognitive-developmental theorists generally) distinguish the two domains of social judgment on the basis of the presence of issues of harm, justice, rights and welfare. Particularly for modern Westerners, it seems like common sense that morality is about helping and hurting people, whereas the rest of our social rules are just social conventions, that is, revisable by consensus. But this common sense is not as common as we might think. In nearly all cultures, the social order is a moral order, and rules about clothing, gender roles, food, and forms of address are profoundly moral issues (Abu-Lughod, 1986; Meigs, 1984; Parish, 1994; Shweder, Mahapatra & Miller, 1987; Hampshire, 1982). Even a cursory look at foundational religious texts reveals that, while God or the gods do seem to care about whether we help or hurt each other, they care about many other things besides. It would be a gross misunderstanding of ancient Judaism, for example, to describe the Ten Commandments as a mixture of moral rules (about not stealing, killing, or lying) and social conventions (about the Sabbath, and prescribed ways of speaking and worshipping.)

We are cultural psychologists, and we advocate an approach to morality that begins with the cultural facts, not with an *a priori* definition of the moral domain inherited from moral philosophers. We take a bottom-up approach and ask: what is the moral domain as people actually create it? When we examine what people care about, gossip about, legislate, regulate, prohibit, and praise, we find only one culture on earth that limits the moral domain to issues of harm, rights, and justice: well-educated secular Westerners, particularly those who are political liberals (see Haidt, Koller, & Dias, 1993 for evidence on westernization and social class; see Haidt & Graham, in press, and Haidt & Hersh, 2001, for evidence on politics). Research like this has demonstrated that, while a statistically significant distinction can be found in most cultures between people’s judgments about harm/rights/justice issues and other issues, it is a mistake of interpretation to conclude that the other issues are not moral issues. (See also Haidt et al., 1993, and Shweder et al., 1987, for studies that failed to find any significant moral/conventional distinction in at least one population). The moral domain, and its sub-domains, must be discovered inductively, not dictated deductively on the basis of culture-specific common sense or moral theories.

Shweder (1990, Shweder, et al., 1997) offers a useful systematization of the breadth and variation of the moral domain. From a cluster analysis of moral reasons given by a sample of Hindu Indian informants, and from his own reading of the anthropological literature, Shweder proposes that moral discourse around the world generally draws on one or more of three “ethics:” autonomy, community, and divinity. Each ethic is a set of interrelated moral claims that function to protect a different entity. The “ethic of autonomy” functions to protect individuals, using concepts such as harm and suffering, rights and justice, freedom and autonomy. This is the moral domain as Turiel defines it. But in most cultures, people believe that there are things worth protecting besides individuals. The “ethic of community” functions to protect groups, institutions, and other collective entities using concepts such as duty, respect, honor, loyalty, and tradition. The

ethic of divinity functions to protect and glorify God, particularly as God is manifested within each person. This ethic involves moral concepts such as purity, piety, chastity, and other forms of self-restraint that help people live in a more divine, less carnal way. When empirical comparisons of moral discourse are made between more and less Westernized groups (Haidt et al., 1993; Jensen, 1998), or between more or less politically and religiously conservative groups in the United States (Haidt & Graham, in press; Haidt & Hersh, 2001; Jensen, 1997), it is generally found that well-educated secular liberal Westerners largely limit their moral discourse to the ethic of autonomy, whereas other groups make use of a wider set of concepts, drawing heavily on the ethic of community, and often (though not as pervasively) on the ethic of divinity.

This broader conception of morality raises two challenges for innateness theorists. First, they must explain how this full set of moral issues—not just harm, rights, and justice—is innate. Second, they must reconcile their story about innateness with the obvious variation of moral rules and practices, and of the moral domain itself, across cultures.

### **3. Five ways morality could be innate**

Given the theme of this volume, our goal here is to explore whether a broad and heterogeneous set of moral concepts and motives—beyond harm, rights, and justice—might reflect the existence of some kind of innately given mental content. We had better, then, be clear about what we mean by innate. The word has been used in so many different ways by philosophers, biologists and ethologists (Wimsatt, 1999, lists 13 distinct meanings) that some scholars have despaired of finding the concept useful at all (e.g., Griffiths, 2002). But we find a simple and congenial approach in the writings of Gary Marcus (2004), whose research focuses on the developmental pathways by which genes guide the construction of brains. Marcus uses the metaphor that genes create the first draft of the brain, and experience later edits it: “Nature bestows upon the newborn a considerably complex brain, but one that is best seen as prewired – flexible and subject to change – rather than hardwired, fixed, and immutable” (p. 12). Marcus further explains that the editing – the changes in the brain as it learns and grows – is itself governed by genetic processes. Genes are not just templates for making proteins, as was thought decades ago; rather, a part of each gene is devoted to regulatory processes – switching the gene on and off in response to various chemical signals. Marcus (2004, p. 40) explains that “‘built-in’ does not mean unmalleable; it means organized in advance of experience.” (Samuels, 2004, considers many meanings of innateness and reaches a similar conclusion.) We adopt Marcus’s view of innateness, and in the rest of this paper we try to explain the ways that human morality may be “organized in advance of experience,” and we try to explain how cultural and personal experience revises the first (universal) draft during childhood development.

We now describe five ways that morality could be innate. We begin with two theories of moral development – constructivism and connectionism – that are essentially “blank-slate” theories in that they posit only innate processes, not innate moral contents (e.g. ideas, knowledge). We suggest that both theories are partially correct as descriptions of the editing process, and that both can be improved by positing at least some content that

is organized in advance of experience. We then describe three approaches that *do* posit that the mind contains specific innate moral content. We believe that all five of these approaches are useful, particularly when the first draft and editing processes are distinguished. In section 4 we propose our own hypotheses as to what is inscribed in the first draft of moral judgment, and in sections 5 and 6 we describe how we think it is edited.

### *3.1. Piagetian constructivism*

Jean Piaget (1965/1932) got down on his knees and played marbles with children to study the process by which children come to understand rules. In contrast to his contemporaries who stressed the child's passive internalization of the morals of the father (Freud) or society (Durkheim), Piaget showed that children are genuinely active participants in their own development. Development in any domain was, for Piaget, the product of the child's continuous interaction with her environment, as a result of which she continually constructed and re-constructed a progressively more adequate understanding. (Thus Piagetian theory is often referred to as "constructivism.") Lawrence Kohlberg worked out the sequence of progressively more adequate understandings of morality in his famous six stages, and he credited the process of "role taking" with being the driving force of moral development.

For Piaget and Kohlberg, there was just one word written in the first draft of the moral mind: empathy (or perhaps a compound word: "perspective-taking"). There were also some words written in other (non-moral) chapters of the first draft: like and dislike. As long as the child liked some things (such as pleasure, candy, or friendship) and disliked others (such as pain, frustration, or personal rejection), then as she became increasingly good at taking the perspective of others during the concrete operational stage, the child could feel for herself (empathetically) that actions that hurt others were bad, while actions that made others happy were good. In this way children were said to come to understand the value of different kinds of rules, and to appreciate that rules and social practices had to be justified by reference to something else. For Kohlberg, that something else was tradition, authority, and society for "conventional" moral reasoners, but it was justice for the most advanced moral reasoners. The knowledge of justice was not innate; it was the crowning achievement of the editing process, and the editing process was constructivism driven by the experience of role-taking.

This approach is elegant in explaining how so much can be derived from so little innate knowledge, and it is reasonable if you believe that the moral domain is restricted to matters of harm, rights, and justice. However, if you believe that concepts such as obedience, respect, honor, chastity, temperance, and sacrilege are truly moral concepts that need to be explained (rather than overcome by the child on her way to moral autonomy and perfect justice reasoning), then it is not clear how these can be derived from empathy and role-taking. Why not posit that the first draft of the moral mind has several words written into it, beyond empathy, and that Piaget and Kohlberg are correct that constructivism is a part of the editing process?

### 3.2. *Connectionism*

A second approach that focuses on the editing process is the connectionist paradigm proposed by Paul Churchland (1996, 1998). Churchland starts from the manifest fact that in addition to the physical environment, human beings are born into and live in a sociomoral world that is extremely dense and complex. The problem for such creatures is to learn to navigate this environment successfully by developing adequate representations of it and pairing those representations with appropriate behavioral responses. There is no special faculty for accomplishing this task; it is accomplished in the same way that people learn to represent and live in their physical environments: through the gradual tuning up of expertise by a mind that is produced by a brain that is a neural network.

Churchland's connectionist account of moral functioning is essentially an account of moral learning. For Churchland, moral development does not mean, as it did for Kohlberg, the gradual formulation of abstract, universal moral principles; rather, it is "a matter of slowly generating a hierarchy of moral prototypes, presumably from a substantial number of relevant *examples* of the moral kinds at issue" (Churchland, 1996, p. 102). Churchland's account of moral cognition and competence is closely congruent with one modern view of cognition, and it is also congruent with some of the most ancient understandings of moral development, particularly Aristotle's (Casebeer, 2003). Churchland's approach is essentially a translation of the terms of virtue ethics into connectionist language. It treats moral competence as a set of skills cultivated gradually, by practice, and helped along by adult emphasis on moral stories, fables, and role models.

Churchland's approach is eloquent on how the editing process occurs, but it is silent on the contents of the first draft. In fact, it invites the inference that there is no first draft, other than an innate interest in people and social events. We see this as an easily correctible flaw, for it implies equipotentiality in moral learning: children could just as easily learn to navigate and value any artificial moral world that adults created for them, such as those of communes and kibbutzim that tried, unsuccessfully, to overcome people's preferences for sharing material goods with their close kin, and their aversion to mating with the people they were raised with. Just as children enter the world with some initial settings in the food domain (a liking for sweet, a dislike of bitter) which are then extended by cultural learning, it seems likely that children enter the world with some initial settings in the social domain (a liking for fairness, a dislike of harm) which are then extended by cultural learning. It does little violence to Churchland's theory to propose that the initial state of the system, for example, the initial weights of particular synaptic connections, is "organized in advance of experience," and then edited by experience in the way that Churchland describes. (See Clark's [2000] critique of Churchland for a similar point).

### 3.3. *Relational models*

A third approach to innateness is Alan Fiske's (1991, 1992) theory of "relational models." Fiske's theory is elegant and parsimonious; it was designed to explain both cross-cultural similarity and diversity in a wide range of cultural domains. Fiske (2004, p. 3) gives the following capsule summary:

Relational models theory is simple: People relate to each other in just four ways. Interaction can be structured with respect to (1) what people have in common, (2) ordered differences, (3) additive imbalances, or (4) ratios. When people focus on what they have in common, they are using a model we call Communal Sharing. When people construct some aspect of an interaction in terms of ordered differences, the model is Authority Ranking. When people attend to additive imbalances, they are framing the interaction in terms of the Equality Matching model. When they coordinate their actions according to proportions or rates, the model is Market Pricing. ... In short, four innate, open-ended relational structures, completed by congruent socially transmitted complements, structure most social action, thought, and motivation. That's the theory.

Fiske suggests that the first three models are innate, and are clearly found in other primates (Haslam, 1997). But Market Pricing – social relations based on ratios – appears to be uniquely human, emerging later in both phylogenetic and ontogenetic development than Communal Sharing, Authority Ranking and Equality Matching. The case for the innateness of Market Pricing is therefore much weaker, and Fiske suggests that it might be in the process of becoming an innate psychological mechanism for social relationships. Fiske therefore is very clear on what is contained in the first draft of the social mind: three primitive and inescapable social-perceptual tendencies: to see people as groups in which all are the same, as rank-orderable on one or more dimensions, or as entities whose relative position must be kept equal despite periodic shifts out of balance. These innate models are used in all cultures to structure the social world, although cultures often choose to use different combinations of models to govern any given relationship. For example, the division of household labor may rely on Authority Ranking in some cultures (e.g., the husband is the authority and dictates who does what); or on equality matching (both spouses take turns at each job); or on Communal Sharing (everyone pitches in without keeping track of who does what) or on Market Pricing (jobs are assigned values proportional to their difficulty, and children, or hired help, is paid to perform them).

Most of Fiske's relational models theory is about the editing process: how this first draft, with three (or four) simple and open cognitive frames gets filled in and tuned up during childhood development. We have no criticism of Fiske's theory, except that we think a bit more is given in the first draft, as we explain below.

### 3.4. Massive Modularity

The fourth and most widely discussed approach to innateness is the concept of *modularity*. One can distinguish between *minimalist* and *maximalist* theories of modularity. On the minimalist view (Fodor, 1983), a very small number of mental functions – primarily having to do with sense perception and language – are modular, in the sense that they are innate, fast, informationally encapsulated, functionally specialized computational mechanisms. Fodor offered a very stringent definition of what it takes to

be a module and then claimed that there are very few modules in the mind, and none for handling higher-order tasks like reasoning or moral judgment.

On the maximalist view proposed by evolutionary psychologists (Barkow, Cosmides, & Tooby, 1992; Buss, 2004; Pinker, 1997), the vast majority of the mind is composed of modules—hundreds or thousands of them—designed by natural selection to solve specific problems that were recurrent in the ancestral environment. According to Tooby and Cosmides (2005), many of these modules play an important role in our moral lives. They argue that the study of valuation, even more than other kinds of cognition, reveals just how crucial it is to posit innate mental *content*, not just innate learning processes. Children are born with a preference (value) for sweetness and against bitterness; any parent knows that the preference for candy over broccoli is not learned by “socialization” and cannot be undone by role models, threats, or rewards. Tooby and Cosmides suggest that the same thing is true for valuation in all domains:

The proprietary content introduced by the architecture constitutes a form of knowledge: the architecture must know (in some sense) that living children are better than dead children, social approval is better than disapproval, salt and sweet are better than acrid or putrefying, sex with your mother or father is to be avoided, helping siblings is (within certain tradeoffs) better than helping fungi, your mate copulating with your sexual rival is worse than his or her fidelity, spiders on your cheek are worse than in the garden, understanding is better than confusion, skill mastery is better than inept performance, and so on. (317)

In this passage Tooby and Cosmides gather many kinds of valuation together under the rubric of what they elsewhere call “motivational principles,” and not all of them are relevant for our purpose, which is to think through the ways in which specifically *moral* judgment might have an innate foundation. Salt, sweetness, and spiders, for example, while clearly the objects of tastes and preferences, seem different in kind from preferences connected with understanding, sexual fidelity, and helping, if for no other reason than the latter seem to have more conceptual and less perceptual content. Tooby and Cosmides are interested in developing an inventory and a science of motivation *in general*, rather than a theory of morality, and while the latter is certainly related to the former (and may, perhaps, turn out to be just a special case of it), the moral domain is distinctive enough that their very useful account will need a little modification.

We agree with Tooby and Cosmides that valuation – for social behavior as for food – is impossible to explain if one refuses to entertain the notion that there is some innate structure and content built into the mind. As they put it:

there must be an irreducible core set of initial, evolved, architecture-derived content-specific valuation assignment procedures, or the system could not get started. The debate cannot sensibly be over the necessary existence of this core set. The real debate is over how large the core set

must be, and what the proper computational description of these valuation procedures and their associated motivational circuitry is (p. 317).

Massive modularity is a controversial notion. Jerry Fodor, the original author of the concept of mental modules, has said that “the massive modularity thesis pretty clearly *isn't* true” (Fodor 2000, p. 23), and a number of other thinkers have followed his lead for diverse reasons (see, e.g., Buller, 2005, Buller and Hardcastle, 2000). We see two principal problems in applying the massive modularity thesis to morality. The first is one of Fodor’s main concerns, known as the “flexibility problem.” Higher order human cognition—and certainly moral cognition—is quite flexible and to some degree culturally variable, yet a massively modular mental architecture would seem to imply that all people in all cultures have the same basic “reliably developing” moral minds. Without having a large central domain-general integration space, it seems difficult to account for variability either across cultures, or across times within a single adult. The second problem is the encapsulation problem: while many moral judgments meet most of Fodor’s criteria for modularity – including domain specificity and speed – it is implausible to think that moral judgments are as informationally encapsulated as the sorts of phenomena usually used to illustrate modularity at the perceptual level. For example, the Müller-Lyer illusion is unaffected by one’s knowledge of the true lengths of the lines, but moral judgments are easily affected by learning new facts about the situation, or by experimental manipulations of mood or other factors that seem extraneous to the operation of a moral module (e.g., de Steno et al, 2006). We are sympathetic to the possibility of substantial domain-specific knowledge in the first draft of the moral mind, but we would like a version of modularity that can solve these two problems.

### 3.5. “Teeming” modularity

Several theorists (e.g. Carruthers<sup>1</sup>) have sought a middle way between completely non-modular conceptions of the mind and massively modular theories. These thinkers speak of “moderately massive modularity” or “modularity to some interesting degree.” They doubt that there are *no* conceptual modules, but they are also skeptical that the mind is a Swiss army knife crammed with tools that were fully designed long ago. One of the most important of these moderate modularists is the anthropologist Dan Sperber (1994, 2005). As an anthropologist, Sperber’s goal was to explain both the diversity and the stability of culture. Massive modularity with Fodorean modules is incompatible with cultural diversity (the flexibility problem), but the blank slate models assumed by many anthropologists cannot explain either the deep and surprising similarities between cultures (e.g., in gods, ghosts, and witches; Boyer, 2001), or the relative stability of each culture over time (Sperber & Hirschfeld, 2004).

Sperber’s solution is a version of massive modularity, but his modules are decidedly un-Fodorean: they are highly variable (some meet all of Fodor’s criteria, some meet only a few); they are often nested within each other (just as the digestive system is a biological module that contains many sub-modules), and most importantly, most of Sperber’s modules are not innate; they are generated during development by a smaller set of “learning modules” which are innate templates or “learning instincts” (Sperber, 2005,

<sup>1</sup> Available at <http://www.philosophy.umd.edu/Faculty/pcarruthers/Moderate-modularity.htm>.

p.57, citing Marler, 1991). Some of these innate modules have specific perceptual content built in; for example, a fruit-learning module will “know” that fruit is sweet, and will only generate subsequent fruit-recognition modules (e.g., for apples and bananas) for objects in the environment that meet those pre-specified criteria. Other learning modules may be more purely conceptual; for example, if there is an innate learning module for fairness, it generates a host of culture-specific unfairness-detection modules, such as a “cutting-in-line detector” in cultures where people queue up, but not in cultures where they don’t; an “unequal division of food” detector in cultures where children expect to get exactly equal portions as their siblings, but not in cultures where portions are given out by age. Because Sperber envisions a core set of innate modules generating a great diversity of other modules, he uses the evocative term “teeming modularity.”

At this point, any reader who is not already a modularity theorist would be likely to think that we had joined Sperber in a jump off a cliff into a land where everything and everybody is named “module.” Let us explain why we are intrigued by Sperber’s ideas. Our goal is to understand the first draft and the revision processes that create the moral mind. Our empirical research is on moral intuition and moral dumbfounding (for Haidt) and on culture and virtue (for Joseph). For both of us, our research shows that moral judgment is not well described by the domain-general application of rules and principles to specific cases, as though moral judgment was a product of moral reasoning in the Kohlbergian sense. Rather, when people are interviewed about taboo violations (such as consensual sibling incest, or harmless cannibalism), they answer very quickly, and their answers show what appears to be a kind of Müller-Lyer-like encapsulation: people can sometimes be pushed in cross-examination to say “I don’t know why, I can’t explain it, I just know its wrong” (Haidt, 2001; Haidt & Bjorklund, in press). We have argued (Haidt & Joseph, 2004) that the adult mind is full of moral intuitions, which are like little bits of input-output programming connecting the perception of a pattern in the social world (often a virtue or vice) to an evaluation and in many cases a specific moral emotion (anger, contempt, admiration). When people think, gossip, and argue about moral issues, the playing field is not affectively flat and open to any kind of reason; it is more like a minefield or pinball machine where flash after flash of affectively laden intuition bounces around one’s attention and pushes participants toward specific conclusions. These intuitions are not Fodor modules, but they are fast, domain-specific bits of mental structure that strongly influence moral judgment (Haidt, 2001). Where do all these intuitions come from?

Perhaps they are all innate, and people simply learn what conditions, in their culture, count as acts of harm or unfairness (e.g., cutting in line). But Sperber’s approach allows us to explain certain acquired moral tastes, in much the way that other kinds of acquired tastes and fears are explained. People are innately attracted to fruits and to meat, but there are special learning mechanisms that can generate a new and enduring disgust towards specific foods, particularly meats. For example, in 1805, when the Lewis and Clark expedition survived months of starvation in the Bitterroot mountains and first made contact with the Nez Perce tribe along the Columbia river, the men gorged themselves on salmon and on a root vegetable, both of which were new to them. Most of them got sick that night from the barely digestible root vegetable, but because of innate one-trial

learning mechanisms that associate nausea preferentially with meat (meats are much more likely to contain bacterial contaminants than are vegetables), they developed a disgust towards salmon. The disgust was so strong that in subsequent days the men purchased dogs to eat from the locals, because that was the only meat available to them other than salmon (Burns, 1997). Was “salmon-disgust” a new module, generated from an innate learning module? Sperber would say yes. We are not in a position ourselves to say what counts as a module, but we see this as a new intuition (gut feeling) generated by an innate learning process that can radically alter the value of things on the basis of experience, but only within limits related to evolutionary adaptation. The new intuition was partially encapsulated: factual knowledge that the root vegetable was the culprit would not have eliminated the disgust.

Moral development shows some of these same features. People develop a large set of input patterns to which they then react quickly, automatically, and emotionally. For example, Americans in recent decades have become finely attuned to the issue of sexual abuse of children, so much so that they are horrified by social patterns that are quite normal in other parts of the world, such as having children sleep in the same bed as an opposite sex parent through middle childhood (Shweder, Balle-Jensen, & Goldstein, 1995), or kissing the genitals of infant boys as an expression of affection (Shweder, in press). Simply explaining to Americans that these practices are not thought by participants to have anything to do with sexuality is not going to eliminate the disgust and condemnation. Is there an innate sexual abuse detector? Probably not. But as we will explain below, we think there is innate structure that makes protection of children, and sexual activity, evolutionarily prepared domains for moral concern. Other examples would be the speed, ease, and passion with which the American “religious right” sees sin, temptation, and sacrilege, or the American “left” sees racism, oppression, and victimization. Whether or not these moral reactions are seen as manifestations of acquired (teeming) modules or just as sub-cultural expertise, they are examples of moral intuitions: bits of mental structure that connect the perception of specific patterns in the social world to evaluations and emotions that are not fully controllable or revisable by the person who experiences them.

#### **4. The first draft of morality: The five foundations of intuitive ethics**

We have long been searching for the foundations of intuitive ethics—the psychological primitives that are the building blocks from which cultures create moralities that are unique yet constrained in their variations. We began (Haidt & Joseph, 2004) by examining a number of theories (including Shweder’s and Fiske’s) about the breadth of human morality and about its precursors in other primates (e.g., de Waal, 1996). We tried to identify the full range of phenomena across cultures that would need to be explained by any adequate theory of human morality. We identified five sets of concerns, each linked to an adaptive imperative and one or more moral emotions, as the best candidates for the psychological foundations of human morality. The five foundations we identified are harm, reciprocity, ingroup, hierarchy, and purity. Each of these five is a good candidate for a Sperber-style learning module. However, readers who do not like modularity theories can think of each one as an evolutionary preparedness (Seligman,

1971) to link certain patterns of social appraisal to specific emotional and motivational reactions. All we insist upon is that the moral mind is partially structured in advance of experience so that five (or more) classes of social concerns are likely to become moralized during development. Social issues that cannot be related to one of the foundations are much harder to teach, or to get people to care about.

Table 1 gives our theory in a concise form. The first row lists five longstanding adaptive challenges that highly social mammals such as our ancestors faced for millions of years, creating conditions that favored the reproductive success of individuals who could solve the problems more effectively. For each challenge, effective adaptation meant being able to detect certain patterns in the social world and respond to them with an altered motivational profile. Sperber (1994) refers to the set of objects that a module was designed 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 now in fact trigger the module.

The way to read the table is to read down each column. For example, the harm foundation can be understood by beginning with the fact that mammals by definition face the need to care for vulnerable offspring, and nothing could be more central to evolutionary success than keeping these offspring alive. It is therefore implausible that mammals learn entirely from experience how to recognize suffering or distress in their offspring. Rather, primates (at least) seem to have innate harm-detection modules that were shaped by evolution to be responsive to the proper domain of signs of suffering in their own offspring. In actual practice this module (or set of modules) is responsive to many things besides the suffering and distress of one’s own children. Suffering by or harm to almost any child-like entity is part of the actual domain of this module. (A poster showing neotenuous baby seals being clubbed to death by large men is a deliberately contrived superstimulus for this module.) This module has as one of its outputs the emotion of compassion: the individual is motivated to act so as to relieve the suffering or otherwise protect the child. We do not know whether there is a single harm module that has both innate and learned triggers, or whether the teeming modularity account is correct in which the human mind is innately prepared to generate a host of specific harm-related modules. However, as long as all people have an emotional sensitivity to harm, particularly harm of the weak or vulnerable, and as long as people have language, they are likely to develop a vocabulary for talking about their emotional reactions. They are likely to have virtue and vice words with which to praise and condemn people, and to instruct their children. Such virtue talk can then feed back to fine tune the bounds and applications of the module: cultures can become expert in perceiving certain kinds of harm (e.g., sexual abuse).

We tell a similar story for each of the other four columns. For example, Trivers (1971) described a suite of emotions that help people form cooperative dyadic relationships and reap the gains of reciprocal altruism. People (especially males) also easily aggregate into tribes, gangs, and teams that compete with other tribes, gangs, and teams. Conflicts over territory seem to call upon this ingroup foundation particularly keenly, and these mechanisms seem to be a large part of the explanation of the extreme form of self-sacrifice known as suicide terrorism (Pape, 2005). Many primates, probably including our

direct ancestors (Boehm, 1999), live in dominance hierarchies, and the common display patterns of dominance and submission across cultures and species strongly suggests some kind of innate structure linked to emotions of fear and respect. However, as Fiske points out repeatedly, Authority Ranking is a two way street: subordinates must show respect and deference, but superiors must then protect them from external threats and maintain order within the group. And finally, the omnivorous food strategy of human beings, combined with our relatively large group sizes (compared to other primates; Dunbar, 1993) means that we have long been exposed to very high levels of threat from bacteria and parasites, which spread by physical contact. Humans (but no other animals) therefore developed a suite of cognitive and emotional adaptations related to disgust that make us wary but very flexible about the kinds of things we eat, and about the contact histories of the things we eat (Rozin & Fallon, 1987). This food evaluation and rejection system was well adapted for social evaluation and rejection, and every human culture appears to use some of the vocabulary and logic of physical disgust in its moral life (Haidt, et al., 1997; Rozin, Haidt, & McCauley, 2000).

We think that these five sets of appraisals-linked-to-emotions-linked-to-virtues are the best candidates for being the foundations of intuitive ethics for several reasons. First, in the ways that cultures deal with these five adaptive challenges we find a surprising degree of similarity across cultures – for example, in the logic of initiation rites that create a strong ingroup; in the ways that hierarchy and submission are marked; and in the purity and pollution rules that so often regulate biological processes such as menstruation, birth, and defecation. Second, four of our proposed five foundations (all but purity) appear to involve psychological “building blocks” that are present in chimpanzees (de Waal, 1995), giving us further confidence that something about these foundations is “specified in advance of experience.” Third, our five foundations fit perfectly with Shweder’s three ethics (harm and reciprocity give rise to the discourse of the ethic of autonomy; ingroup and hierarchy support the ethic of community; and purity supports the ethic of divinity). Fourth, three of our foundations are coincident with Fiske’s first three relational models (reciprocity = Equality matching; ingroup = Communal Sharing; Authority = Authority Ranking). To the extent that our five foundations don’t match Fiske’s four models, the discrepancy is due to the fact that harm and purity are not primarily modes of interpersonal relationship. We include them because they are important and innate sources of human moral valuation, but we do not include Market Pricing because we do not think it is clearly innate. We could easily be wrong about excluding Market Pricing. We do not claim that there are *only* five foundations. There are probably many more, but we believe the five we have identified are the most important ones for explaining the origins of human morality.

## **5. The editing process: Developing virtues**

The five foundations are, in a manner of speaking, the innate “taste buds” of the moral sense. The human tongue has five kinds of receptor, each of which translates a chemical pattern in a substance to an affective experience that is positive (for sweet, salt, and glutamate) or negative (for bitter and, beyond a certain level of intensity, for sour). These taste buds tell us something about how our ancestors lived: they ate fruit and meat, and had a variety of perceptual tools in their tongues (and noses and eyes) that meshed with

conceptual tools in their brains to help guide them to fruit and meat. Similarly, the five foundations suggest some things about how our ancestors lived: they were ultrasocial creatures (Richerson & Boyd, 1998), finely tuned for 1) rearing children and helping kin, 2) selectively cooperating with non-kin while remaining vigilant for cheaters, 3) forming strong ingroups for the purpose of cross-group competition, 4) organizing themselves hierarchically, and 5) attending to each other's physical states and altering interactions and contacts accordingly.

But clearly there is more to moral judgment than the operation of five “taste buds.” Mature moral functioning does not consist only, or even primarily, of simple affective or intuitive reactions to gross social stimuli. It is also characterized by the acquisition and use of a very wide variety of moral concepts. Some of these are categories of actions – lies, betrayals, favors, and so on. Others are categories of persons, or more specifically, categories of characteristics of persons, including positively valenced traits<sup>2</sup> such as kindness, loyalty and trustworthiness, and negatively valenced ones such as cruelty, dishonesty and cowardice.

These traits – virtues and vices – are beginning to re-emerge in empirical moral psychology after a long period of exile, occasioned in part by critiques by Lawrence Kohlberg and other theorists. We have previously discussed the role that virtues can play in a comprehensive theory of moral functioning and moral development (Haidt and Joseph 2004). We repeat our main points here, as a prelude to a discussion of one aspect of virtue theory and of our theory in particular, namely, narrative.

First, what is a virtue? There are many views, but most virtue theorists would agree at least that virtues are characteristics of a person that are morally praiseworthy. Virtues are therefore traits, as long as one doesn't think of traits as global tendencies to act in a particular way (e.g., honest, brave) across widely varying circumstances. Rather, we think of traits as John Dewey did: as dynamic patternings of perception, emotion, judgment, and action (Dewey 1922; see also Churchland 1998). Virtues are social skills. To possess a virtue is to have extended and refined one's abilities to perceive morally-relevant information so that one is fully responsive to the local sociomoral context. To be kind, for example, is to have a perceptual sensitivity to certain features of situations, including those having to do with the well-being of others, and for one's motivations to be appropriately shaped and affected. To be courageous is to have a different kind of

---

<sup>2</sup> The word “trait” is fraught with significance in psychology; in particular, it is the focus of a heated debate between personality and social psychologists. Some psychologists have placed traits at the center of the study of personality, while others, for various reasons, are skeptical or dismissive of the very concept (Mischel 1968; Nisbett and Ross 1991). This debate has penetrated philosophical and psychological discussions of morality, with “situationists” like John Doris (1998) and Gilbert Harman (1999) construing virtues as traits and then dismissing their existence, and virtue theorists (Sreenivasan 2002, Merritt 2000) defending versions of virtue theory against the situationist critique. We believe the virtue theorists are right; virtues, as we construe them, are highly situation-specific *skills* or *capacities* rather than broad behavioral dispositions. This way of seeing virtues obviates the basic charge of the situationists, and is consistent with Mischel's original and ongoing critique of trait theories (Cervone and Shoda 1999), and with recent work in cognitive neuroscience (e.g. Casebeer 2003, Churchland 1998).

sensitivity; to be patient, still another.<sup>3</sup>

One of the crucial tenets of virtue theory is that the virtues are acquired inductively – that is, through exposure to -- sometimes with efforts to copy -- many examples of the virtue in practice. Each of these examples contains information about a number of aspects of the situation, including the motivations of the protagonists, their state of being (suffering, disabled, hostile, rich, etc.), the categorization of the situation, and the evaluation of the outcome offered by more experienced others. Only over time will the moral learner recognize what information is important to notice and retain, and what can be safely disregarded.

As philosophers and cognitive scientists have recently been arguing, with respect both to morality and to cognition more generally, this kind of learning cannot be replaced with top-down learning, such as the acceptance of a rule or principle and the deduction of specific responses from it. Interestingly, this aspect of virtue theory shows Aristotle to have been a forerunner of the current application of the connectionist approach to morality that we described above (see May, Friedman, & Clark, 1997). In this model, the mind, like the brain itself, is a network that gets tuned up gradually by experience. With training, the mind does a progressively better job of recognizing important patterns of input and of responding with the appropriate pattern of output.

For those who emphasize the importance of virtues in moral functioning, then, moral maturity is a matter of achieving a comprehensive attunement to the world, a set of highly sophisticated sensitivities embodied in the individual virtues. Of course, reasoning and deliberation play important roles in this conception as well; part of being a virtuous person is being able to reason in the right way about difficult or problematic situations. But virtue theory is nevertheless a departure from theories of morality that see deliberation as the basic moral psychological activity.

Virtue theory posits a particular kind of organization of moral competence, one in which perception, motivation, action and reasoning correspond to demands placed on the person by features of situations. Naturally, the objectivity of these demands, and the moral relevance of features of situations, are to some degree dictated by culture, by the moral concepts, social structures, and narratives that are current in the immediate social context. But this does not mean that the content or structure of a virtue is completely culturally relative. As Aristotle, one of the foundational figures in virtue theory, pointed out, and as current virtue ethicists have elaborated (Nussbaum, 1993), what it means for a personality characteristic to be a *virtue*, and not simply a behavioral regularity, is largely that it consists in functioning well in a specific “sphere of existence.” And what Aristotle and Nussbaum mean by “spheres of existence” is similar, if not identical, to what evolutionary biologists would recognize as persistent adaptive challenges and other types of environmental constraint. Virtues are therefore quite at home in a scientific theory of moral functioning based on evolutionary psychology and cultural psychology.

---

<sup>3</sup> For a classic exposition of the construal of virtues as sensitivities or perceptual capacities, see McDowell (1979).

Virtue theory allows us to draw together all that we find correct and useful in the five approaches that we sketched out in Section 3. The child is indeed an active participant in her own development; moral knowledge and skills are not just “downloaded” into the child’s mind, as blank slate socialization theories would have it. Piaget and Kohlberg are correct that there is a substantial element of self-contruction in moral development. However, what is being learned is best described as the skills of social perception and reaction discussed by connectionists and virtue theorists. Most of these skills are about how to interact with other people – how to fill in the three (or four) innately given models for social relationships described by Fiske. However, some of this knowledge is not about relationships per se; there is also much else that is innate, particularly when we look at the origins of valuation, as described by Tooby and Cosmides (2005). In addition to being “organized in advance of experience” for Fiske’s first three models (which involve reciprocity, hierarchy, and the formation of ingroups), the mind is also innately prepared to perceive and care about harm from a very early age (Zahn-Waxler et al., 1979), and also about disgust, purity, and pollution (from a later age, perhaps not fully until the age of 7 or 8; Rozin, Haidt, & McCauley, 2000). Turiel may have been correct to focus on harm, and the child’s ability to understand and dislike suffering, as the most important intuition of early moral development. We believe he was wrong, however, to suggest that children derive all of their other moral concepts by self-constructing them on this single foundation.

To summarize: the characteristic developmental trajectory in the moral domain is a movement from crude, global judgments articulated using a small number of innate moral intuitions to highly sophisticated and differentiated perceptions, beliefs, emotional responses and judgments. This is consistent with Sperber’s notion of “teeming modularity:” domain-specific, module-like intuitive mini-programs give rise, in the mature moral agent, to an expansive and flexible set of moral modules that are more powerful and subtle than the innate modules that compose the five foundations of intuitive ethics that we have been discussing. Sperber’s approach suggests that virtues are not themselves innate, but rather are acquired through a generative process in which the domain-specific capacities of the modules that compose the five foundations are multiplied, expanded and refined.

## **6. The editing process: Learning narratives**

Constructivists such as Piaget and Kohlberg called our attention to the ways that children actively create ever more nuanced understandings of moral issues. Kohlberg studied the most explicit, discursive, deliberative aspects of moral functioning. Stage growth was measured through the analysis of verbal reasoning used to justify responses to hypothetical moral dilemmas. In the cognitive-developmental tradition, moral thinking was seen as akin to logical thinking; Piaget (1965/1932, p. 398) said explicitly that “Logic is the morality of thought just as morality is the logic of action.”

We agree that children reflect on moral questions, particularly when in discussion with others, and we see the attraction of positing a domain-general workspace where moral thinking (as well as other kinds of thinking) is carried out. However we agree with critics of Kohlberg who advocate a *narrative* approach to moral development (see, for example,

Vitz, 1990). These critics argue, based on the pioneering work of Bruner (1990), Sarbin (1986), and others, that moral functioning, like mental functioning in general, is at least as much a matter of a *narrative* mode of cognition as of a *paradigmatic* or *logico-scientific* mode.

There are two modes of cognitive functioning, two modes of thought, each providing distinctive ways of ordering experience, of constructing reality.... A good story and a well-formed argument are different natural kinds. Both can be used as a means for convincing another. Yet what they convince *of* is fundamentally different: arguments convince one of their truth, stories of their lifelikeness. The one verifies by eventual appeal to procedures for establishing formal and empirical proof. The other establishes not truth but verisimilitude. (Bruner, 1990, p. 11)

Bruner observes that we know a great deal more about the paradigmatic mode of thought, because cognitive psychology has concentrated its attention on it while the narrative mode has been comparatively ignored. He also points out that each mode of thought relies upon its own “prostheses” – aids to thinking provided by a culture. For the paradigmatic mode, prostheses include logic, mathematics, and the sciences; for the narrative mode, the most common prosthetic devices are texts. Texts, among their many other functions, serve to store up cultural meanings, and, through both their content and their structure, they help to guide the thinking of individuals. We think that moral thinking, argument, and reflection (outside of philosophy departments, at least) is much better described as a kind of narrative thinking than as a kind of paradigmatic thinking.

There are many different kinds of narratives, of varying levels of complexity, and as a result narrativity (see Carrithers 1991 for a discussion of this concept) shapes moral functioning and moral development at several levels of organization. Some of the most powerful moral narratives are the simplest. For Westerners, parables such as those found in Jesus’ teaching in the New Testament are familiar examples, and other religions have similar tools, for example the *hadith*, or sayings and doings of the Prophet Muhammad, in Islam. Shweder and Much (1991) found that narratives are commonly invoked in Hindu cultures as a mode of moral argument. In their interviews with Indian informants on moral dilemmas, they often found that questions about the rightness or wrongness of a particular act elicited a response beginning, “Let me tell you a story about ...” (for example, stealing).

More recently, the sociologist Christian Smith (2003, p.64) has observed that we are “animals who make stories but also animals who are *made by* our stories.” Smith describes a variety of high-order often unconscious narratives that organize identity and moral judgment at both the individual and group levels. For example, he notes that Americans and “militant Muslims” interpret the 9/11 attacks in the light of very different metanarratives: Americans see things through what Smith calls “the American Experiment narrative” in which Americans fled the oppression of the old world and ever since have been a shining beacon of liberty and hope, while the “Militant Islamic Resurgence” narrative gives a radically different perception of the causes and

significance of the attacks. There are other narratives, each of which Smith spells out almost like a recipe. Among them are the “Capitalist Prosperity” narrative, the “Progressive Socialism” narrative, the “Expressive Romantic” narrative, and the “Scientific Enlightenment” narrative.

Smith is especially helpful in making explicit the narratives that motivate and guide American sociologists and other academics. For example, the “liberal progress” narrative tells the story of how

Once upon a time, the vast majority of human persons suffered in societies and social institutions that were unjust, unhealthy, repressive, and oppressive. These traditional societies were reprehensible because of their deep-rooted inequality, exploitation, and irrational traditionalism... But the noble human aspiration for autonomy, equality, and prosperity struggled mightily against the forces of misery and oppression, and eventually succeeded in establishing modern, liberal, democratic, capitalist, welfare societies. [However] there is much work to be done to dismantle the powerful vestiges of inequality, exploitation, and repression. This struggle... is the one mission truly worth dedicating one’s life to achieving.

This narrative clearly draws heavily on the harm and reciprocity foundations to tell a story of triumph. It explicitly rejects the hierarchy foundation as a source of value, portraying authority and its attendant valuation of tradition as the very cause of evil in the world.

In contrast, the “Community Lost” narrative is more politically conservative; it relies primarily on the ingroup and hierarchy foundations to tell a story of decline and decay:

Once upon a time, folk lived together in local, face to face communities where we knew and took care of each other... life was securely woven in homespun fabrics of organic, integrated culture, faith, and tradition. We truly knew who we were and felt deeply for our land, our kin, our customs. But then a dreadful thing happened: Folk community was overrun by the barbarisms of modern industry, urbanization, rationality, science, fragmentation, anonymity... Faith began to erode, social trust dissipate, folk customs vanish.... All that remains today are tattered vestiges of a world we have lost. The task of those who see clearly now is to memorialize and celebrate folk community, mourn its ruin, and resist and denounce the depravities of modern, scientific rationalism that would kill the Human Spirit.

Neither narrative is correct in any objective sense. Both are ways that sociologists have tried to make sense of social change that are, like cuisines, methods of artfully combining and recombining a few favored elements. Without the innately-given five foundations

there could be no emotionally compelling moral narratives. But without narrative, our moral concepts would be disjointed and hard to integrate into coherent action plans.

## 7. Conclusion

de Waal (1996) suggests that a building block of human morality visible in chimpanzees is the desire for peace and harmony within the group. Celebrations break out when long-simmering power struggles are resolved. We think this desire is related to the ingroup foundation: group-living creatures prefer (have an innate valuation of) harmony within the cooperative groups upon which they depend both for material sustenance and inter-group defense. We find this desire in ourselves: we are a part of the community of morality researchers that has long been divided on the question of moral innateness. This makes us uncomfortable, for we really like and value the many members of our community, and we have tried, in this essay, to show how all are right about something, all have something to contribute. We propose that some degree of harmony and synergy can be restored if most morality researchers were willing to endorse this statement: the first draft of the moral mind has diverse moral content that was specified in advance of experience, but this innately given content gets revised and greatly extended during the course of development as children actively construct their moral knowledge within a cultural context that uses narrative to shape and guide the development of specific virtues. Is anyone ready to celebrate with us?

## References

- Abu-Lughod, L. (1986). *Veiled sentiments*. Berkeley: University of California Press.
- Aristotle. (1941). *Nicomachean ethics* (W. D. Ross, Trans.). In R. McKeon (Ed.), *The basic works of Aristotle* (pp. 927-1112). New York: Random House.
- Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The Adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford.
- Baron, Jonathan (1993). *Morality and rational choice*. Springer-Verlag.
- Batson, C. D., O'Quinn, K., Fuly, J., Vanderplass, M., & Isen, A. M. (1983). Influence of self-reported distress and empathy on egoistic versus altruistic motivation to help. *Journal of Personality and Social Psychology*, *45*, 706-718.
- Batson, C. D., Kobryniewicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, *72*, 1335-1348.
- Boehm, C. (1999). *Hierarchy in the forest: The evolution of egalitarian behavior*. Cambridge, MA: Harvard University Press.

- Boyer, P. (2001). *Religion explained: The evolutionary origins of religious thought*. New York: Basic.
- Bruner, Jerome (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.
- Buller, D. J. (2005). *Adapting minds: Evolutionary psychology and the persistent quest for human nature*. Cambridge, MA: MIT Press/Bradford Books.
- Buller, D. J. and Hardcastle, V. G. (2000). Evolutionary psychology, meet developmental neurobiology: Against promiscuous modularity. *Brain and Mind*, 1, 307-325.
- Burns, K. (1997). Lewis & Clark - The Journey of the Corps of Discovery.
- Buss, D. M. (2004). *Evolutionary psychology: The new science of the mind. 2nd Edition*. Boston: Allyn & Bacon.
- Carrithers, Michael (1991). Narrativity: Mind-reading and making societies. In Andrew Whiten (Ed.), *Natural theories of mind: Evolution, Development and Simulation of Everyday Mindreading* (pp. 317-331). Oxford: Basil Blackwell.
- Carruthers, P. (Date?). Moderately massive modularity. Available on the Internet at <http://www.philosophy.umd.edu/Faculty/pcarruthers/Moderate-modularity.htm>.
- Casebeer, William D. (2003). *Natural ethical facts: evolution, connectionism, and moral cognition*. Cambridge, MA: MIT Press.
- Cervone, D. and Shoda, Y. (Eds.) (1999). *The coherence of personality: Social-cognitive bases of consistency, variability, and organization*. New York: Guilford Press.
- Churchland, Paul M. (1996). *The engine of reason, the seat of the soul: a philosophical journey into the brain*. Cambridge, MA: MIT Press.
- Churchland, Paul M. (1998). Toward a Cognitive Neurobiology of the Moral Virtues. *Topoi* 17, 83-96.
- Clark, Andy (2000). Word and action: reconciling rules and know-how in moral cognition. In R. Campbell and B. Hunter, eds., *Moral Epistemology Naturalized, Canadian Journal of Philosophy*, supplementary volume XXVI.
- Desteno et al. (2006) xx
- de Waal, F. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Cambridge, MA: Harvard University Press.
- Darley, J. M., & Batson, C. D. (1973). From Jerusalem to Jericho: A study of situational and dispositional variables in helping behavior. *Journal of Personality & Social Psychology*, 27, 100-108.
- Dewey, John (1922). *Human Nature and Conduct: An Introduction to Social Psychology*. New York: Holt.
- Doris, J. (1998). Persons, situations, and virtue ethics. *Nous*, 32(4): 504-530.
- Doris, J. (
- Dunbar, R. I. M. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral & Brain Sciences*, 16, 681-735.
- Fiske, A. P. (1991). *Structures of social life*. New York: Free Press.
- Fiske, A. P. (1992). Four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, 99, 689-723.
- Fiske, A. P. (2004). Relational Models Theory 2.0. In N. Haslam (Ed.), *Relational models theory: A contemporary overview* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge, MA: Harvard University Press.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI study of emotional engagement in moral judgment. *Science*, 293, 2105-2108.
- Griffiths, P. (2002). What is innateness? *The Monist*, 85(1): 70-85.

- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814-834.
- Haidt, J., & Bjorklund, F. (in press). Social intuitionists answer six questions about morality. In W. Sinnott-Armstrong (Ed.), *Moral psychology, vol. 2: The cognitive science of morality*.
- 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, Fall*, 55-66.
- Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, *65*, 613-628.
- Haidt, J., Rozin, P., McCauley, C. R., & Imada, S. (1997). Body, psyche, and culture: The relationship between disgust and morality. *Psychology and Developing Societies*, *9*, 107-131.
- Hansen, C. (1991). Classical Chinese Ethics. In P. Singer (Ed.), *A companion to ethics* (pp. 69-81). Oxford, UK: Oxford.
- Harman, G. (1999). Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. *Proceedings of the Aristotelian Society*, *99*, 315-32.
- Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character. Vol. 1: Studies in deceit*. New York: MacMillan.
- Haslam, N. (1997). Four grammars for primate social relations. In J. A. Simpson & D. T. Kenrick (Eds.), *Evolutionary social psychology* (pp. 297-316). Mahwah, NJ: Lawrence Erlbaum.
- Jensen, L. A. (1997). Culture wars: American moral divisions across the adult lifespan. *Journal of Adult Development*, *4*, 107-121.
- Jensen, L. A. (1998). Moral divisions within countries between orthodoxy and progressivism: India and the united states. *Journal for the Scientific Study of Religion*, *37*, 90-107.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago: Rand McNally.
- MacIntyre, A. (1981). *After virtue*. Notre Dame: University of Notre Dame Press.
- Marler, P. (1991). The instinct to learn. In Susan Carey and Rochel Gelman (eds.), *The epigenesis of mind*, 37-66. Hillsdale: Erlbaum.
- Marcus, G. (2004). *The birth of the mind*. New York: Basic Books.
- May, L., Friedman, M., & Clark, A. (Eds.). (1995). *Mind and morals: Essays on ethics and cognitive science*. Cambridge, MA: MIT Press.
- McDowell, J. (1979). Virtue and reason. *Monist* *62*: 331-350.
- Meigs, A. (1984). *Food, sex, and pollution: A New Guinea religion*. New Brunswick, NJ: Rutgers University Press.
- Merritt, M. (2000). Virtue ethics and situationist personality psychology. *Ethical theory and moral practice*, *3*: 365-83.
- Nussbaum, Martha C. (1993). Non-relative virtues: an Aristotelian approach. In M. C. Nussbaum & A. Sen (Eds.), *The quality of life* (pp. 242-269). New York: Oxford University Press.
- Robert A. Pape (2005) *Dying to win: The strategic logic of suicide terrorism. xx*
- Parish, S. M. (1994). *Moral knowing in a Hindu sacred city*. New York: Columbia University Press.
- Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, *64*, 467-478.
- Piaget, J. (1965/1932). *The moral judgement of the child* (M. Gabain, Trans.). New York: Free Press.
- Pinker, S. (1997). *How the mind works*. New York: Norton.

- Pincoffs, E. L. (1986). *Quandaries and virtues: Against reductivism in ethics*. Lawrence, Kansas: University of Kansas.
- Prinz, 2006 xx
- Richerson, P. J., & Boyd, R. (1998). The evolution of human ultra-sociality. In I. Eibl-Eibesfeldt & F. K. Salter (Eds.), *Indoctrinability, ideology, and warfare: Evolutionary perspectives* (pp. 71-95). New York: Berghahn.
- Rozin, P., & Fallon, A. (1987). A perspective on disgust. *Psychological Review*, 94, 23-41.
- 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). New York: Guilford Press.
- Sarbin, Theodore R. (Ed.) (1986). *Narrative psychology: The storied nature of human conduct*. New York: Praeger.
- Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy*, 2, 307-320.
- Shweder, R. (1990). In defense of moral realism: Reply to Gabennesch. *Child Development*, 61, 2060-2067.
- Shweder, in press xx [recent article on human rights crusades]
- Shweder, R. A., Balle-Jensen, L., & Goldstein, W. M. (1995). Who sleeps by whom revisited: A method for extracting the moral goods implicit in practices. In J. Goodnow, P. Miller & F. Kessel (Eds.), *Cultural practices as contexts for development* (pp. 21-39). San Francisco: Jossey-Bass.
- 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.
- Smith, C. (2003). *Moral, believing animals: Human personhood and culture*. Oxford, UK: Oxford.
- 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. Stich (Eds.), *The innate mind: Structure and contents* (pp. 53-68). New York: Oxford.
- Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences*, 8, 40-46.
- Sreenivasan, G. (2002). Errors about errors: Virtue theory and trait attribution. *Mind*, 111: 47-68.
- Tooby, J., Cosmides, L., & Barrett, 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.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35-57.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, England: Cambridge University Press.
- Vitz, Paul C. (1990). The use of stories in moral development: new psychological reasons for an old education method. *American Psychologist*, 45, 709-720.
- Wimsatt, W. C. (1999). Generativity, entrenchment, evolution, and innateness: Philosophy, evolutionary biology, and the conceptual foundations of science. In V. G. Hardcastle (Ed.), *Where biology meets psychology: Philosophical essays* (pp. 139-180). Cambridge, MA: MIT Press.

Zahn-Waxler, C., Radke-Yarrow, M., & King, R. (1979). Child rearing and children's prosocial initiations toward victims of distress. *Child Development, 50*, 319-330.

	<b>Harm</b>	<b>Reciprocity</b>	<b>Ingroup</b>	<b>Hierarchy</b>	<b>Purity</b>
Adaptive challenge	help young, vulnerable, or injured kin	reap benefits of dyadic cooperation with non-kin	Reap benefits of group cooperation	Negotiate hierarchy	Avoid microbes and parasites
Proper domain (original triggers)	Suffering, distress, or threat to one's kin	Cheating, cooperation, deception	Kin or co-residence boundaries	Signs of dominance and submission	Waste products, diseased people
Actual domain, (modern triggers)	Baby seals, cartoon characters	Marital fidelity, broken vending machines	Sports teams one roots for	Bosses, respected professionals	Taboo ideas (communism, racism)
Characteristic emotions	Compassion	anger, gratitude, guilt	Group pride, belongingness	Respect, fear	Disgust
Relevant virtues [and vices]	kindness, compassion [cruelty]	fairness, justice, trustworthiness [dishonesty]	Loyalty, patriotism, self-sacrifice [treason, cowardice]	Obedience, deference [disobedience, uppitiness]	cleanliness, purity, temperance, chastity [hedonism, lust, intemperance]

**Table 1.** The five foundations of intuitive ethics.