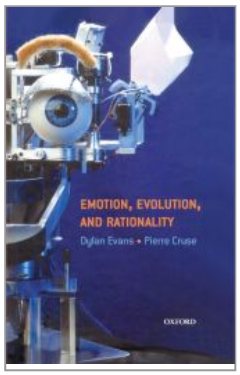

Emotion, Evolution, and Rationality
Dylan Evans and Pierre Cruse

Print publication date: 2004
Print ISBN-13: 9780198528975
Published to Oxford Scholarship Online: March 2012
DOI: 10.1093/acprof:oso/9780198528975.001.0001

Evolution, culture, and the irrationality of the emotions

CHANDRA SEKHAR SRIPADA
STEPHEN STICH

DOI:10.1093/acprof:oso/9780198528975.003.0007

Abstract and Keywords

Most of researchers agree that emotions (or at least the developmental processes that give rise to them) have been designed by natural selection, and that this ensures that emotions will play a positive role in rational decision-making. This chapter, however, casts doubt on whether emotions really are rational in this sense. It argues that emotions are intrinsically linked to mentally represented sets of goals values and standards that this chapter refers to as a ‘value structure’. It claims that there are reasons to think that the contents of value structures will often be maladaptive, and that when they are the emotions and behaviour that they lead to are irrational. Thus there is good reason to think that emotions will sometimes fail to be conducive to rational action.

Keywords: emotions, developmental processes, natural selection, decision-making, rationality, value structure

Introduction

For about 2500 years, from Plato's time until the closing decades of the 20th century, the dominant view was that emotions are quite distinct from the processes of rational thinking and decision-making, and are often a major impediment to those processes. But in recent years this orthodoxy has been challenged in a number of ways. Damasio (1994) has made a forceful case that the traditional view, which he has dubbed *Descartes' Error*, is quite wrong, because emotions play a fundamental role in rational decision-making. When the systems underlying the emotions do not function properly, Damasio maintains,

rational decision-making breaks down. Other theorists, most notably Robert Frank (1988), have argued that if we view the emotions through the longer lens of evolutionary theory, we can see that much of what looked to be irrational in the emotions is actually part of an effective strategy for achieving agents' goals and maximizing their reproductive success. In the wake of this and other recent work, the pendulum of received opinion has swung in the other direction. The emotions are now increasingly regarded as inherently rational, as Frank maintains, and as important components of other rational processes.

One of our goals in this paper is to argue that the pendulum has swung too far, and to push it back a bit in the other direction. Though we will not disagree with either Damasio or Frank, we argue that their work tells only part of the story about the rationality of the emotions. Emotions, we will maintain, are intrinsically linked to a mentally represented set of norms, goals and values which we call a *value structure*. Moreover, there are good reasons—indeed good *evolutionary* reasons—to think that the contents of value structures will **(p.134)** often be *mal*adaptive. When they are, the emotions and the behavior they lead to will typically be *ir*rational.

The paper is divided into four sections. In Section 1, we sketch several quite different accounts of rationality, and focus on the account that we propose to use in assessing the rationality of the emotions. In Section 2 we will introduce the notion of a value structure and explain the central role that value structures play in many recent theories about the mental mechanisms subserving the emotions. In Section 3, we ask how value structures arise. There are, we maintain, three sources from which value structures arise: genes, the environment, and culture. In Section 4 we argue that each of the three sources which influence the formation of value structures can give rise to value structures with importantly irrational components. These are illustrated with some quite varied examples, drawn from the literature on attitudes toward violence, food taboos and psychopathology. Some of the most interesting reasons for thinking that value structures and emotions will often be irrational are suggested by recent work on gene-culture co-evolution by Robert Boyd, Peter Richerson and others. To date, this work has made relatively little impact on research on the emotions. A second major goal of the paper is to argue for the importance of Boyd and Richerson's work in this area.

Some approaches to rationality

As Samuels *et al*. (2003) have argued, accounts of rationality can be usefully divided into two major categories. *Deontological accounts* assess the rationality of instances or patterns of reasoning and decision-making by appeal to normative rules or principles like those of logic and decision theory. In *consequentialist accounts*, by contrast, instances or patterns of reasoning and decision-making are evaluated by attending to their consequences—the states of affairs they lead to—in one or another environment, with different

consequentialist accounts focusing on different sorts of consequences and environments. In reliabilist epistemology, true belief is the consequence that looms large in the assessment of rationality, while in both pragmatist epistemology and decision theory the focus is on desire satisfaction or the maximization of subjective utility.[1] Individual (or occasionally group) well-being is another sort of outcome that a consequentialist account of rationality might seek to maximize. The notion of well-being is, notoriously, much more philosophically controversial and much harder to pin down than the notion of subjective utility. But it is clearly a distinct notion since, sometimes at least, what people **(p.135)** want does not contribute to their well-being (Kahneman *et al*. 1999). Inclusive fitness is yet another property that consequentialist theorists have thought to be important. When evolutionary psychologists like Cosmides and Tooby assess the rationality of belief-forming and decision-making processes, it is clear that what they often have in mind is an evaluation of the contribution that the processes make to the inclusive fitness of those who employ them (Cosmides and Tooby 1994). Similarly, the notion of 'ecological rationality' that is used in the work of Gerd Gigerenzer and his collaborators (Gigerenzer *et al*. 1999, Gigerenzer 2000) is sometimes best understood as a consequentialist notion where what is being maximized is inclusive fitness in some environment. Consequentialist assessments of the rationality of a psychological mechanism or process will, of course, often be quite sensitive to the environment in which the mechanism is assumed to operate. One obvious choice of environment is the one in which the mechanism is actually embedded. For other purposes, however, a theorist might focus on what Sperber (1996) has called the 'proper' environment—the environment in which the mechanism in question evolved. For still other purposes, accounts of rationality might focus on one or another sort of idealized environment in which certain theoretically important assumptions are assumed to obtain (Goldman 1986). The taxonomy of accounts of rationality that we have sketched is summarized in Fig. 7.1.

In exploring the rationality of the emotions in this paper, we will be primarily concerned with the consequentialist notion of rationality that links it to inclusive fitness in the actual environment—the square labelled 'A' in Fig. 7.1.

(p.136) We adopt this notion of rationality for several reasons. First, as argued in Samuels *et al*. (2003), deontological accounts of rationality are problematic in a variety of ways. Second, much discussion of the rationality of the emotions, in recent years, has explicitly or implicitly relied on some version of a consequentialist/inclusive fitness account of rationality (Tooby and Cosmides 1990, Frank 1988, Fessler 2001). Third, assessments of rationality in accordance with the account we will be focusing on often (though not always!) roughly coincide with assessments that would result if we instead used inclusive fitness in the proper environment (square B) or well-being in the actual or proper environment (squares C and D), and when these assessments diverge significantly, there is often something important to be learned. Finally, and perhaps most importantly, sociobiologists, human behavioral ecologists and other advocates of robust versions of adaptationism maintain that the existence of maladaptive mental dispositions or processes is unlikely on evolutionary grounds.[2] This paper can be viewed, inter alia, as an extended argument against this view.

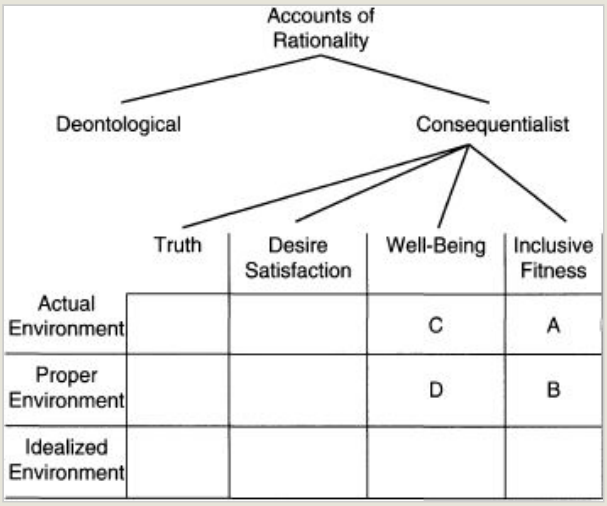

Figure 7.1 A taxonomy of accounts of rationality.

Emotions and value structures

In this section we will explain our notion of a *value structure* and indicate why we think that value structures, or something quite like them, play a role in the account of the emotions offered by a number of leading researchers. We will start with Paul Ekman's highly influential affect program theory. Affect programs, according to Ekman, are universal, largely automated (or involuntary) suites of co-ordinated emotional responses subserved by innate psychological and physiological mechanisms present in all normal members of the species. In humans, the suite of responses often includes an emotion specific facial expression, characteristic autonomic nervous system activity, characteristic subjective experience and emotion-specific action tendencies (Ekman 1992). While the affect programs themselves are taken to be innate and universal, Ekman's work on display rules soon convinced him that emotional responses further ‘downstream’ may be strongly influenced by culturally local beliefs and norms (Ekman 1972).[3] For our purposes, what is most **(p.137)** important is the question of what happens ‘upstream,’ that is, what determines the elicitation of an emotion episode. On Ekman's account, affect programs are typically triggered by an innate ‘appraisal mechanism’ that selectively attends to external and internal stimuli indicating that the emotion is appropriate. It is uncertain whether Ekman ever thought that there are some stimuli which the appraisal mechanism is built to respond to directly, without the mediation of other

cognitive states and processes. By the mid-1990s, however, Ekman had clearly adopted the view that much of the activity of the appraisal mechanism is affected both by culturally local beliefs and by culturally local norms, values and goals (Ekman 1994). The mental representation of these norms, values and goals are what we call a value structure.[4]

Other researchers, influenced by Ekman, have elaborated his notion of an appraisal mechanism in useful ways. Lazarus (1994) suggests that each emotion is linked to its own innate principle, which he calls a ‘core relational theme,’ specifying the conditions under which it is appropriate to have the emotion. The innate specifications of these conditions are quite abstract, however, and thus lots of culturally local information is required to determine when the conditions are satisfied. The core relational theme that Lazarus proposes for anger, for example, is ‘a demeaning offense against me and mine’. Obviously, there is no way of knowing when such an offense has occurred without knowing a fair amount about locally prevailing norms and values. Synthesizing the ideas of a number of theorists, Robert Levenson has proposed a ‘biocultural model’ of the emotions, depicted in Fig. 7.2, which ‘reflects a confluence between innate and learned influences’ (Levenson 1994, 125). The ‘innate hardwired’ parts of the model—corresponding roughly to Ekman's affect program mechanisms—are in the center of the diagram, between the black boxes. Emotion prototypes are the equivalent of Lazarus's core relational themes. The black boxes, indicating the appraisal system and the display and feeling rules are the ‘primary loci of cultural influences’ and can access the agent's beliefs and values, many of which will be culturally local. In Fig. 7.3, we have elaborated on Levenson's model to make the role of beliefs and value structures in this account more explicit.

The idea that emotions are typically elicited by appraisal processes that are sensitive to the contents of a person's value structure is endorsed by theorists who approach the emotions from a variety of theoretical perspectives. **(p.138)**

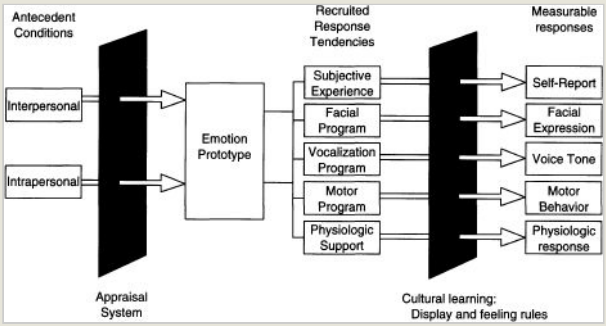

Figure 7.2 Levenson's BioCultural Model of the Emotions. Permission sought.

Nico Frijda, Klaus Scherer, Keith Oatley, and Andrew Ortony and his colleagues and many other leading theorists have all emphasized that the goals, norms and values that constitute a person's value structure are crucial in the antecedents of emotion (for a review see Scherer 1988). Of course, there **(p.139)** are many important differences among these theorists.[5] What is important for our purposes is that these theorists *agree* that value structures—norms, values and goals—do, in fact, play a crucial role in the psychological processes that elicit emotions.

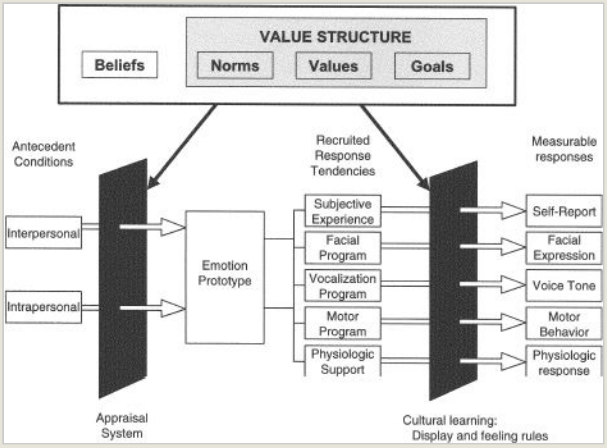

Figure 7.3 An elaboration of Levenson's model in which the role of beliefs and value structures are made explicit.

It might be thought that the widely discussed work of Joseph LeDoux is an important exception to the claim that leading researchers recognize the role of value structures in triggering emotions, since his work shows that there are emotion-triggering pathways which are not mediated by higher cognitive processes at all. LeDoux maintains that some emotions are elicited by a 'low-road' pathway which bypasses appraisal mechanisms and value structures altogether (LeDoux 1996). In claiming that value structures play an important role in the antecedents of emotions, we do not deny that there are alternate routes to the elicitation of emotions such as the route that LeDoux has characterized. Moreover LeDoux himself views appraisal mechanisms as part of the explanation of what establishes some instances of low-road circuits. So even LeDoux would endorse our contention that value structures play a crucial role in the antecedents of many (though not necessarily all) emotions.

What we have argued so far is that there is a broad consensus in support of the claim that the antecedents of many emotion episodes (though perhaps not all) are tightly linked to aspects of people's value structures. In the following two sections, we will argue that some of the sources which causally influence the contents of value structures can (and do) lead to the production of *maladaptive* value structure components. We will also argue that these maladaptive value structure components can generate maladaptive (and thus irrational) emotions.

Three sources of the contents of value structures

There has been little systematic work aimed at explaining how the contents of value structures arise. Following Boyd and Richerson (1985), we propose that it is useful to distinguish three importantly different sources which contribute to the contents of value structures: genes, the environment and culture. **(p.140)** In both this section and the one that follows, we will elaborate on these sources

and offer examples of value structure elements that could well have originated from them. We propose this taxonomy only as a rough first pass at addressing the largely unasked question of why value structures end up having the contents they do. It is important to keep in mind that in distinguishing these three sources, we are not claiming they are independent, and we stress that they can and often do *jointly* contribute to the formation of elements of value structures. With this caveat in place, we now turn to the task of clarifying and explicating the sources from which the contents of value structures arise.

It is clear that *genes* causally contribute to the formation of some elements of value structures since genes play an important and wide-ranging role in phenogenesis more generally. In some cases, genes play the predominant role in causally contributing to some phenotypic outcome, and we will follow a standard practice in referring to such outcomes as *innate*. There is much dispute about how best to understand innateness (Samuels 2002, Griffiths 2002). However, on any reasonable account, core cases of innateness include phenotypic characteristics such as color vision and blood type, and disorders such as phenylketonuria. We will be discussing some innate psychological traits which contribute to the contents of value structures in the next section.

A second source of causal influences which shape the contents of value structures is the *environment*. Boyd and Richerson use the term *individual learning* as a label for a cluster of processes by which people acquire and modify beliefs, skills and elements of their value structures by interacting with, and getting feedback from, the (non-social) environment. Examples of individual learning processes include classical conditioning, trial and error instrumental learning and various forms of inductive learning. In individual learning, interaction with the environment can lead to changes in a number of different kinds of mental states, including both beliefs and preferences. For example, a person may try many different routes to work and discover which is shortest (change in belief) or may try all the different flavors of Baskin Robbins ice cream and find out which one she likes best (change in preferences).

The last and probably most important source of causal influences which shape the contents of value structures is *culture*. We define culture in terms of its relation to a transmission process often called *social learning*. In social learning, a mental state in one person causally contributes to the formation of a mental state in another person, where the latter mental state *resembles* the former.[6] Two paradigmatic examples of social learning processes are *teaching* and **(p. 141)** *imitation*. Social learning is important because it can lead to the transmission of information such as beliefs, values and skills between individuals and across generations. In this respect, social learning differs profoundly from individual learning, since the products of individual learning will die with the learner unless some social learning process is introduced. Following Boyd and Richerson, we view *culture* as the body of information residing in the minds of

people that was acquired by social learning. We think it is plausible to suppose that social learning contributes significantly to the formation of peoples' value structures, so value structures are intimately connected with culture.

Culture understood in this way can be usefully viewed as an *inheritance system*. The cultural inheritance system can best be explained by analogy with the more familiar genetic inheritance system. In the genetic inheritance system, there is a statistically defined pattern of resemblance between biological parents and children—for example, on average, taller parents have taller children and shorter parents have shorter children—and parents causally contribute to this resemblance by means of well-known genetic mechanisms. The existence of mechanisms which produce patterns of resemblance between individuals, whether based on genes or any other process, is the crucial feature which defines the highly general and useful notion of an inheritance system.

In any inheritance system, a range of forces can potentially act to produce evolutionary change in the distribution of the characteristics of individuals in a population over time. A full specification of these forces depends on the particular structure and properties of the system. For example, one important force which can operate in the genetic inheritance system is *natural selection*. Other forces which might be potentially operative include genetic drift, recombination and migration.

The case of culture is formally analogous to the genetic case. In the social learning process, a mental state of a *cultural parent* causally facilitates the formation of a type-similar mental state in a *cultural child*. As a result, there is a statistically defined pattern of resemblance between cultural parents and children. For example, on average, the religious beliefs and practices of Muslim children resemble those of their Muslim parents, while the beliefs and practices of Mormon children resemble those of their Mormon parents. Thus, culture can be viewed as an inheritance system, with social learning providing the mechanism of inheritance.[7]

As in the case of the genetic inheritance system, a range of forces can potentially act on the cultural inheritance system. The list of forces which can **(p.142)** operate on the cultural system is different than the analogous list for the genetic system because while the cultural system exhibits many properties which are similar to the genetic system, it also exhibits many which are unique and have no parallel in the genetic system. Like the genetic system, change in the cultural system is *cumulative*. Many cultural variants, for example, the technologies of kayak building or watch-making,[8] arose by something like the familiar Darwinian process of descent with modification over extended stretches of time. Unlike the genetic system, however, in the cultural system novel cultural variants can emerge by a *directed process*. A person can create a novel cultural variant whose cultural fitness is better than one would expect were that variant

generated by chance.[9] Furthermore, these novel variants are heritable in that they can be transmitted by social learning. The cumulative and directed nature of the cultural inheritance system makes it an enormously powerful tool for generating innovations that have enabled humans to adapt to environmental change and to survive in a wider variety of environments than any other animal species. Indeed, cumulative social learning is *uniquely human*. There are only a few good examples of social learning in other species, and no other species exhibits the sort of cumulative social learning that humans do (Henrich and Boyd 1998).

Sometimes the social learning process is *unbiased*—an existing cultural variant is picked at random. However, a unique feature of the cultural system is that often social learning is mediated by important and powerful biases. A bias is a process that determines which cultural variant a person will adopt, and there are several possible kinds of biases that a person may utilize. In *direct bias, a* person tries out the different variants that are available and sees which one she prefers. In *indirect bias*, a person selects a cultural variant based on some feature or marker that characterizes the variant. There are two important kinds of indirect biases which we will be focusing on. In *conformist bias*, a person adopts a variant that is common in his or her environment. In *prestige bias*, a person adopts a variant exhibited by a high prestige person.[10]

We have now completed our taxonomy of the sources from which the contents of value structures arise and distinguished three principal sources: genes, **(p.143)** the environment and culture. As we have noted, these three sources need not be independent; rather, they will often interact in jointly contributing to some aspect of phenotype. Indeed, such interaction occurs in many familiar cases such as language learning and moral development. But despite the fact that they interact in many domains, there are a number of reasons why it is important to keep these three sources of causal influence separate, at least conceptually. One reason emerges when constructing models of genetic and cultural evolution in which all three sources play a role. As we have noted, genes, the environment and culture have different properties (for example, genes and culture can exhibit different inheritance structures). Because of these differences, causal influences arising from each source contribute in very different ways to the overall population-level dynamics which such models attempt to capture (Boyd and Richerson 1985). For our purposes, another reason for distinguishing these three sources is that contributions from each source can (and often do) '*fail*' independently of the others. By 'fail,' we mean that causal influences originating from these sources end up contributing to the production of *maladaptive* phenotypic outcomes. By distinguishing these three sources from which the contents of value structures arise and studying how contributions from each might independently fail, we can gain a deeper understanding of how value structures for a single individual, or for a population of individuals in a

community, can come to have maladaptive elements, leading to emotions which are (in our sense) irrational.

Maladaptive value structures and irrational emotions

We suggest that there are two basic ways in which the contributions from genes, the environment or culture might end up producing maladaptive elements of value structures. For want of better names, we will refer to these two kinds of failure *as failure due to inertia* and *failure due to inevitable misfiring*. We will clarify what these two kinds of failure amount to as we proceed. Since we propose that there are three basic sources from which elements of value structures arise, and two ways in which the contribution from these sources might fail, the result is a two by three matrix as shown in Fig. 4. Our main task in this section is to fill in this matrix with examples that plausibly illustrate cases in which the contributions from these three sources end up producing maladaptive elements of value structures and irrational emotions. Because space is limited, we will not try to fill in every box; we will leave 2 and 4 empty. We are inclined to think that the remaining four boxes in our matrix are in many ways the most interesting. The examples to be used are indicated in Fig. 7.4. **(p.144)**

4.1

We begin with box 1—inertial failure in the genetic system. The idea behind genetic inertia is quite simple: Natural selection operating on genetic variation can be quite slow and phenotypic change often requires extended periods of time. Thus, even though natural selection may have produced phenotypic variants adaptive in an ancestral environment, as environments change there is substantial scope for mismatches between an organism's phenotype and the phenotype that would be most adaptive in the current environment. When such a mismatch occurs, we say it is due to *genetic inertia*. This theme is often emphasized in the evolutionary psychology literature. Indeed, the importance of genetic inertia is a point on which evolutionary psychologists and sociobiologists (or human behavioral ecologists) differ sharply (Tooby and Cosmides 1990, Laland and Brown 2002). We have suggested that genes causally contribute to the formation of some elements of value structures, and it is plausible that one way in which these elements might fail to be adaptive is due to the operation of genetic inertia. We believe that certain kinds of specific phobias illustrate

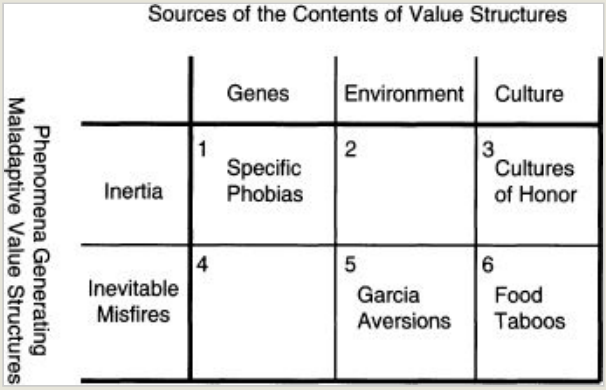

Figure 7.4 Contribution from genes, the environment or culture might end up producing maladaptive elements of value structures via inertia or via inevitable misfires.

genetic inertia leading to maladaptive elements of people's value structures and irrational emotions.

Specific phobias are psychiatric disorders characterized by pathological experience of the emotion of fear. We propose that at least some phobias arise from *innate* fears. Plausible candidates for innate fears are those directed at recurrent threats faced by human ancestors. The underlying adaptive logic is that an innate and rapid fear response to a recurrent threat would have conferred a selective advantage on human ancestors who possessed such a trait. **(p.145)** Some specific situations that are plausible candidates for being recurrent threats to human ancestors, and thus triggers for innate fears, include the following: being constrained in a small space without clear means of leaving, being near snakes, being at a great height from the ground, being alone in open spaces far from home, being near blood or injuries and being near large numbers of unfamiliar individuals. Indeed, there is significant evidence that fears triggered by these situations are, in fact, substantially innate (see Marks 1987).

Innate fears can have a wide-ranging impact on behavior by contributing to the acquisition of elements of value structures. In some cases, these value structure elements can be quite maladaptive. For example, a person with a particularly strong innate fear of closed spaces will typically, over time, come to acquire a host of *avoidances*, that is goals and preferences for avoiding situations or activities which involve closed spaces, as elements of their value structures. In ancestral environments, these value structure elements might have been directed at such things as deep caves or dense underbrush, and were presumably adaptive. However, in modern environments, strong fear of closed spaces can lead to avoidances towards elevators, subways, phone booths and a host of other places and activities. Such avoidances, when they are sufficiently debilitating, are the hallmark claustrophobia. Similar kinds of maladaptive avoidances can also occur with the other innate fears listed above, each of which is implicated in a sub-type of phobic anxiety disorder commonly seen in modern clinical practice, namely: ophidophobia, acrophobia, agoraphobia, hemophobia and social phobia (DSM-IV 1994).

We believe that certain strong innate fears, which presumably led to adaptive avoidances in ancestral environments, now contribute to producing the burdensome and maladaptive constellation of avoidances characteristic of phobias.[11] Some kinds of phobic disorders appear to be heritable, and like many complex psychological disorders which are presumed to involve polygenetic inheritance, one could reasonably expect that the strengths of peoples' innate fear dispositions are distributed over some range. We suspect that the **(p.146)** optimal distribution of strengths of innate fear dispositions for ancestral environments is significantly *mismatched* with respect to what would be most adaptive in modern environments (see also Marks and Nesse 1994, Murphy and

Stich 2000). A high level of sensitivity to an innate fear trigger which would be unproblematic, or even adaptive, in ancestral environments leads to morbidity in modern environments.[12]

4.2

We turn now to box 3—cases of inertia in the cultural domain. Earlier we noted that culture is an inheritance system with social learning providing the mechanism of inheritance. Like the genetic system, the cultural system exhibits cumulative change. But culture is unique in that it is also *directed*, that is, cultural change can be more rapid because novel variants can be introduced by processes that reliably produce variants whose cultural fitness is higher than one would expect by chance. Because culture is both cumulative and directed, culture has a *unique* ability to produce highly complex adaptations to local conditions in relatively short periods of time. For this reason, culture-wielding creatures are typically highly *specialized* to their local surroundings. Consider for example Copper Eskimos (another favorite Boyd and Richerson example) who exhibit a spectacular array of socially acquired traits related to matters such clothing, housing, hunting, alliance formation, and on and on. Two features of these cultural traits are noteworthy. First, they were not acquired by the slow process of genetic evolution (there is no reason to suppose these cold-climate adaptations have a genetic basis) or by individual learning. Second, many of these cultural traits represent ecological specializations; they would be inappropriate in any environmental or social context significantly different from the one which the Copper Eskimos do in fact inhabit. Thus while *Homo sapiens* is perhaps the world's most preeminent ecological and geographic generalist, individual cultural groups are largely specialists, achieving high degrees of adaptation to their local environment by means of the cultural inheritance system. By facilitating specialization, the cultural system certainly does promote adaptiveness. But specialization has the trade-off that a specialized individual is more susceptible to mismatch due to *cultural inertia* when environments change. An example which plausibly illustrates **(p.147)** cultural inertia in the domain of emotions is Nisbett and Cohen's well-known study, *Culture of Honor* (Nisbett and Cohen 1996).

Cultures of honor have been documented by anthropologists in many groups around the world. While these groups differ in many respects, they are classified as cultures of honor because they share important commonalities. A central feature of cultures of honor is that males in these cultures are prepared to protect their reputation for strength and probity by resorting to violence. The importance placed on a reputation for strength leads to a corresponding importance placed on insult and the necessity to respond to it. An individual who insults someone must be forced to retract. If the instigator refuses, he must be punished with violence or even death. Cultures of honor have arisen in many parts of the world in situations where resources are liable to theft and where the state's coercive apparatus cannot be relied upon to prevent or punish theft.

These conditions often occur in relatively remote areas where herding is the main viable form of agriculture; the ‘portability’ of herd animals makes them prone to theft. Since cultures of honor have repeatedly and independently emerged in circumstances where placing high degrees of importance on a reputation for strength would be a highly effective social strategy with clear adaptive benefits, we believe that they are cultural adaptations which have been independently invented in different places in response to similar environmental pressures. Thus cultures of honor are an example of *convergent evolution* of traits within the cultural system.[13]

What makes cultures of honor particularly interesting for our purposes is that they exhibit considerable cultural inertia and they have a profound effect on the emotions and on the physiological mechanisms subserving them. Both points are vividly illustrated in pair of experiments by Nisbett and Cohen. In the first experiment, unsuspecting white male subjects walked down a long hall, and as they did so they were bumped by a (large!) male confederate who also called the subject an ‘asshole’. Two observers who were pretending to work in the hall observed the subject's face when he was bumped, noted his body language and rated his emotional reactions. The results of the experiment are shown in Fig. 7.5; they indicate (dramatically) that culture of honor subjects were much more likely to be rated as having higher levels of anger versus amusement, while for non-culture of honor subjects, the reverse is true. **(p.148)**

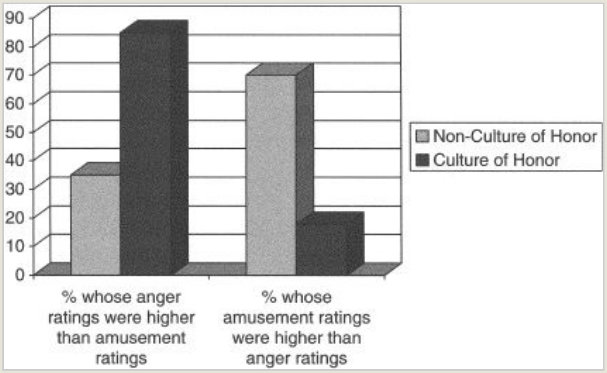

Figure 7.5 The results of an experiment by Nisbett and Cohen in which male subjects were bumped and insulted by a confederate. Most culture of honor subjects were more angry than amused. Permission sought.

The second experiment was similar to the first in that male subjects again walked down a long hallway where they were bumped by a large male confederate who called the subject an ‘asshole’. This experiment differed from the first one in that both before and after walking down the hallway and getting bumped, subjects were asked to provide a saliva sample (a cover story involving measurement of blood sugar during tasks was given). The saliva samples were tested for Cortisol (associated with stress, anxiety, and arousal) and testosterone (associated with aggression and dominance behavior). Again the results, depicted in Fig. 7.6, are dramatic. For culture of honor subjects, both

Cortisol and testosterone were much higher after the bumping and the insult, while this was not the case for non-culture of honor subjects.

Thus far, we have not said anything about who these ‘culture of honor’ and ‘non-culture of honor subjects’ actually are. The crucial point of the experiments, for our purposes, is that all the subjects were students at the University of Michigan. The non-culture of honor subjects were white northerners, while the culture of honor subjects were whites who had spent most of their lives in the American South. Nisbett and Cohen's thesis is that the South is a region whose cultural environment was shaped by members of earlier generations, for example Scotch—Irish immigrants, who had been herders for centuries, and who brought with them and helped entrench a culture of honor within this region. However, the American South has long **(p.149)**

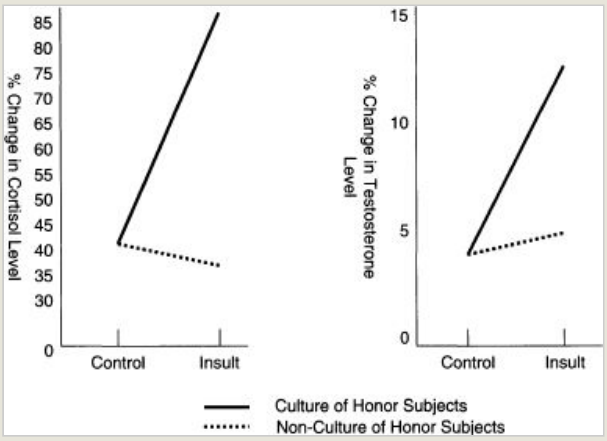

Figure 7.6 The results of a second experiment by Nisbett and Cohen in which levels of Cortisol and testosterone increased much more substantially in culture of honor subjects who were insulted by a confederate. Permission sought.

ceased to be a sparsely populated frontier region, herding has all but disappeared in the South, and reliable policing by the state is available virtually throughout this region. Furthermore, these culture of honor subjects, though raised in the South, were attending the University of Michigan and had relatively high socio-economic status. For several generations, at least, these students' families and their neighbors' families had made their living in non-herding livelihoods like business, the professions and the civil service. Thus the conditions which would have made a culture of honor adaptive in the American South have long since disappeared. Nevertheless, Nisbett and Cohen's experiments show that among subjects raised in the South, a culture of honor persists and these subjects' emotional reactions to insult are profoundly affected by a culturally conveyed value structure that exhibits remarkable *inertia*.[14]

(p.150) 4.3

We turn now to cases of irrational emotions that arise from failure due to inevitable misfires, considering first an example of inevitable misfire in the domain of individual learning from the environment. In a series of important experiments, John Garcia and his colleagues demonstrated that rats acquire a strong aversion to distinctively flavored food if they experience gastrointestinal distress within 12 hours after eating the food (Garcia 1974). Other animals,

including humans, also develop ‘Garcia aversions’—indeed about 30% of Americans report such aversions. Based on these findings, Garcia and his colleagues proposed that that individual learning of food aversions is subserved by a phylogenetically conserved system specifically dedicated to the defense of the gut. The mechanism subserving Garcia aversions differs in many respects from standard classical conditioning. For example, Garcia aversions are elicited by the taste and smell of food (and not by the food's visual, auditory, or tactile properties). Also, Garcia aversions develop despite the fact that the unconditioned stimulus (gastrointestinal distress) is temporally remote (by up to 12 hours) from the conditioned stimulus (the food). For our purposes, the most important difference is the sensitivity of the mechanism. Unlike standard classical conditioning, the mechanism subserving the acquisition of Garcia aversions is extremely sensitive and only a single trial is required for the acquisition of the conditioned response. The high sensitivity of the learning mechanism subserving Garcia aversions was presumably adaptive in an ancestral environment (and may *still* be adaptive in the modern environments)—in a sense the mechanism is following the old adage ‘better safe than sorry’. Nevertheless, because of the high sensitivity of the mechanism, many of the aversions generated by the mechanism are to harmless foods.

The case of Garcia aversions illustrates the important distinction between the adaptiveness of individual learning mechanisms which modify value structure elements in response to environmental contingencies, and the adaptiveness of the value structure elements generated by those mechanisms. Many of the aversions generated by the Garcia mechanism are to perfectly harmless food, and therefore these aversions, and the emotions they trigger, are maladaptive (and, in our sense, irrational). Thus the inevitable misfire of *adaptive* individual learning mechanisms like the Garcia mechanism can lead to *maladaptive* value structure components.

4.4

We believe that some of the most interesting and important examples of maladaptive value structure elements are due to inevitable misfires in the cultural domain. That will be our theme for the remainder of the paper.

(p.151) Earlier we discussed the distinction between directly and indirectly biased social learning. In directly biased social learning, a person tries out different cultural variants in order to see which one works best. In indirectly biased social learning, a person uses some marker, such as the commonness of a cultural variant or the prestige of the person from whom a cultural variant will be copied, to bias the selection of cultural variants. Our focus in this section will be on inevitable misfires that occur as a result of indirect biases. We believe the importance of indirect biases in cultural transmission has not been widely recognized. One of the most original aspects of the work of Boyd and Richerson and their colleagues, and the feature which differentiates them most from other

workers who have adopted Darwinian approaches to culture, is their extensive analysis of the conditions of use and consequences of using indirect biases. For our purposes, indirectly biased social learning is particularly interesting because it produces a number of surprising and counterintuitive effects, one of which is the propagation of maladaptive cultural variants, which Boyd and Richerson sometimes call *rogue memes*. Boyd and Richerson and their colleagues have constructed a family of models demonstrating that natural selection will favor the use of social transmission and indirect biases in a wide variety of circumstances when the environment is variable (but not too variable), and information is costly and thus direct biases are not feasible (Boyd and Richerson 1985). When these circumstances obtain, the optimal cultural system, from the point of view of maximizing genetic fitness, is one that leaves abundant scope for *rogue memes*. Though this is not the place to present these sophisticated mathematical models in detail, it is easy enough to see the intuitive ideas underlying them by examining the conditions under which these models show that indirect biases will tend to be effective, leading to the spread of adaptive cultural variants, and contrasting them with the conditions under which indirect biases tend to lead to the spread of rogue memes.

We begin with prestige bias. Though a person's prestige can be based on a number of factors, there is typically a significant correlation between a person's prestige in a culture and his or her *success* in aspects of life that are of importance in that culture. Mathematical modeling shows that when reliable information about the adaptiveness of cultural variants is hard to come by, prestige bias—adopting the cultural variants that high prestige people have adopted—can be a very successful strategy (Henrich and Gil-White 2001, Boyd and Richerson 1985). But it is often difficult to determine which values, practices and beliefs contribute to prestigious people's success and which do not. People will thus often adopt a sizeable chunk of a prestigious person's repertoire of knowledge, skills, values and practices, some of which might be **(p.152)** useless or even seriously maladaptive. Thus, prestige bias, though an adaptive strategy overall, will nevertheless also lead to the propagation of neutral and maladaptive cultural variants.

Now we turn to conformist bias. In a spatially variable environment, it is reasonable to suppose that different practices, beliefs and values may be adaptive in different places. Mathematical modeling shows that under these conditions, prior episodes of individual learning on the part of others will often make it the case that the most common cultural variant in a given location will also be the most adaptive cultural variant for that location (Henrich and Boyd 1998). Because commonness can serve as a reliable marker for adaptiveness, conformist bias is often effective, which is why its use is so widespread. However, these models also show that there is a range of conditions, including fast environmental change and circumstances in which individual learning is highly inaccurate, under which conformist bias will favor the spread of

maladaptive cultural variants. Conformist bias will also tend to preserve maladaptive cultural variants that become widespread as the result of other processes, e.g. prestige bias.

In addition to the mathematical models of Boyd and Richerson and their colleagues, the importance, and sometimes untoward effects, of indirect biases are also supported by (the admittedly limited) empirical data. The literature on the diffusion of innovations plays a unique role as an empirical database because it offers perhaps the only systematic cross-cultural analysis of the relative role of direct versus indirect biases in adoption of novel cultural variants. Everett Rogers, who is by far the best known investigator in the study of the diffusion of innovations, surveys some 3000 articles in the literature and proposes a model of innovation adoption which is consistent with an important role for prestige and conformist biases in innovation adoption (Rogers 1995, Henrich 2000). If indirect bases are widespread, then we should be able to document cases in which their use leads to acquisition of maladaptive innovations. The diffusion of innovations literature does indeed offer many examples of maladaptive innovations and practices acquired by means of indirect biases. In all these cases, a recurring theme is that people are simply not able to properly assess the goodness or badness of the innovation and must thus make use of indirect biases.[15]

We believe the diffusion of innovations literature suggests two things: the use of indirect biases is widespread and indirect bases do in fact lead to the **(p.153)** propagation of rogue memes. It should be kept in mind that innovations are an ideal case for direct biases since often they are the kinds of things one can try out and ‘see how good they are’. In the case of the social transmission of norms and values, we believe the use of indirect biases will, if anything, be more widespread. In contrast to innovations, it is typically very hard to see how one can use direct biases to guide the acquisition of norms and values since there is little scope to assess how good they are. Thus, even though the empirical evidence is less abundant, we expect that in the social transmission of elements of people's value structures, indirect biases play an even stronger role. And since indirect biases often leave ample scope for the propagation of rogue memes, we suspect that indirectly biased social transmission might account in a significant way for maladaptive elements of people's value structures, and thus for irrational emotions. What we aim to do in the rest of this section is briefly establish at least a circumstantial case for the thesis that irrational emotions do indeed arise as a consequence of rogue meme norms and values propagated by indirect biases. The example we will focus on is the case of irrational reactions of *disgust*, which occur in the context of *food taboos*.

All societies have rules which prohibit consumption of certain types of foods. While the category of food prohibitions is markedly heterogeneous—prohibitions may be applied to an almost boundless variety of food types, be binding on

different classes of people, and may be invoked in a number of different social contexts—we shall refer to all such prohibitions on food as 'food taboos'. The emotion of disgust plays an important role in mediating food taboos. There is abundant anecdotal evidence in the ethnographic literature indicating that the prospect of eating tabooed food items elicits powerful feelings of revulsion, nausea, and the gape expression characteristic of disgust (Whitehead 2000, Simoons 1994). Moreover, consistent with the prominent role of disgust in mediating food taboos, Rozin and his colleagues have proposed that the psychological mechanisms of disgust originated as an elaboration of an ancestral system specialized for the rejection of food (Rozin *et al*. 2001). The question we will be focusing on is: by what mechanisms do particular food items come to be the object of taboos within a society? To the extent that these mechanisms systematically allow for taboos to attach to beneficial foods items rather than harmful ones, the taboos will often be maladaptive, and the result will be irrational emotions, in particular irrational reactions of disgust.

In addressing the question of how taboos become attached to particular food items, we begin by reiterating that across human societies, a startlingly heterogeneous variety of food items are tabooed. This *diversity* alone leads one to suspect that the objects of taboos are quite arbitrary, and that historical accident plays an important role in the explanation of why one food item is **(p. 154)** tabooed and another is not.[16] We believe that in order to gain a deeper understanding of the mechanisms by which taboos attach to particular food items, we must first recognize the unique role that social learning plays in the domain of food. Humans are exceptional in the extent to which they rely on social learning as a source of information. Furthermore, the food domain is particularly apt for utilizing social learning, since copying the food practices of others is relatively easy, while the costs of making errors in food selection can be disastrous (Galef 1998). Thus we believe that a social learning-based approach to understanding food taboos is promising. There are two distinct questions that such an explanation needs to address: how do food taboos *originate*, and how are they *sustained* over generations within cultural groups.

We begin with the first question: How does a particular kind of food originally come to be the object of a taboo within a particular culture? Here we follow evolutionary anthropologists Daniel Fessler and Carlos Navarrete, who suggest an account that invokes a number of interacting psychological and social processes, with indirectly biased cultural transmission playing a crucial role (Fessler and Navarrete unpublished). According to Fessler, food aversions plausibly originate as a consequence of post-ingestion sickness, in the manner of the Garcia phenomenon discussed earlier.[17] Once aversions have been developed, they may spread through direct observation of others—one's sight of a conspecific's aversive response to a food is a potent stimulus for developing **(p.155)** aversive reactions oneself.[18] Crucially, prestige bias and conformist bias are likely to play a pivotal role in propagating aversions from a small subset

of the population to a much wider section of the population; prestige bias when the aversion is still rare, and conformist bias when the aversion has reached a critical mass. Additionally Fessler and Navarrete propose that prestige bias, by itself, might account for the origination of some food taboos—the idiosyncratic dislikes of a prestigious person, even if not related to post-ingestion sickness, might spread by prestige-biased transmission. Finally, they propose that when an aversion is widespread, it eventually comes to be normatively moralized in the form of a taboo. Consumption of a food item which was previously regarded as merely aversive now comes to be regarded as morally wrong and deserving of sanctions (Fessler and Navarrete unpublished manuscript).[19]

While Fessler and Navarrete's account explains how taboos originate, a second puzzle concerns how taboos are sustained over time within cultural traditions. In order to see that there is a puzzle here, one needs to keep in mind that the objects of taboos are enormously diverse across cultural groups, but quite uniform within cultural groups. This stable pattern of uniformity and diversity requires explanation because it is inconsistent with many kinds of cultural transmission. Boyd and Richerson and their colleagues have constructed an elegant series of mathematical models which demonstrate an interesting and unanticipated result: Given plausible assumptions about base rates for mutation and migration, processes like directly biased social transmission or individual learning cannot maintain stable differences between groups and instead lead to the dissipation of diversity between groups, contrary to what is in many cases in fact observed. However, indirect biases, in particular conformist biases, are capable of maintaining sharp withingroup homogeneity and between-group differences with respect to practices such as taboos, thus providing a plausible mechanism for the pattern of diversity we actually see (Boyd and Richerson 1985, Henrich and Boyd 1998).[20] In the absence of empirical evidence which addresses questions about proximal mechanisms directly, these models serve as part of a reasonable circumstantial case that indirect biases are in fact operative in sustaining cross cultural diversity with **(p.156)** respect to cultural practices such as taboos. Thus, to sum up, we believe the inevitable misfiring of indirectly biased social transmission can and does lead to the genesis, propagation and maintenance of maladaptive food taboos, and the norms and values which mediate these maladaptive practices. When indirectly biased social transmission misfires and leads people to acquire maladaptive norms and values as elements of their value structures, the result is irrational emotions.

Conclusion

In recent years, the question of the rationality of the emotions has often been addressed by theorists interested in showing how emotions can perform rational mental functions or at least can be important components in other rational processes. A central theme of this paper has been that such approaches are not complete because they fail to deal with the antecedents of emotions, that is the cognitive structures which underwrite the link between emotion episodes and

particular contexts of elicitation. In constructing a general framework for analyzing the antecedents of emotions, we introduced the crucial idea of a value structure, which consists of the mental representation of the full range of a person's evaluative attitudes, such as goals, values and norms. Using Boyd and Richerson's dual-inheritance analytic framework, we suggested three sources from which elements of value structures arise and two ways in which these three sources might contribute to maladaptive elements of value structures. We illustrated these proposed kinds of failure with specific examples drawn from the empirical literature. The kinds of cases of irrational emotions we described have largely been neglected by theorists interested in the question of the rationality of the emotions. By suggesting new and more precise ways of thinking about the antecedents of emotions, we hope to have provided a novel perspective on the venerable and contentious question of the rationality of the emotions.

References

Bibliography references:

Boyd, R. and Richerson, P. (1985). *Culture and the evolutionary Process*. The University of Chicago Press, Chicago.

Chewlos, G. and Oatley, K. (1994). Appraisal, computational models, and Scherer's expert system. *Cognition and Emotion*, **8**(3), 245–57.

Chomsky, N. (1981). Principles and parameters in syntactic theory. In *Explanation in linguistics: the logical problem of language acquisition* (ed. N. Hornstein and D. Lightfoot), pp. 32–75, Longman, London.

Cosmides, L. and Tooby, J. (1994). Better than rational: evolutionary psychology and the invisible hand. *American Economic Review*, **84**(2), 327–32.

(p.157) Damasio, A. (1994). *Descartes error*. Avon Books, New York.

Davey, G. (1994). The ‘disgusting’ spider. *Society and Animals*, **2**(1).

Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (1994) American Psychiatric Association, Washington D.C.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotions. In *Nebraska Symposium on Motivation 1971*, Vol. 4 (ed. J. K. Cole), pp. 207–83. University of Nebraska Press, Lincoln.

Ekman, P. (1992). Are there basic emotions? *Psychological Review*, **99**(3), 550–53.

Ekman, P. (1994). All emotions are basic. In *The nature of emotions: fundamental questions* (ed. P. Ekman and R. J. Davidson), pp. 15–19. Oxford University Press, New York.

Fessler, D. M. T. (2001). Emotions and cost/benefit assessment: the role of shame and self-esteem in risk taking. In *Bounded rationality: the adaptive toolbox* (ed. R. Selten and G. Gigerenzer), pp. 191–214. MIT University Press, Cambridge, MA.

Fessler, D. M. T. and Navarrete, C. D. (unpublished manuscript). Meat is good to taboo: dietary proscriptions as a product of the interaction of psychological mechanisms and social processes.

Frank, R. (1988). *Passions within reason*. W. W. Norton and Company, New York.

Frijda, N. (1986). *The emotions*. Cambridge University Press, New York.

Galef, B. G. (1998). Communication of information concerning distant diets in a social, central-place foraging species: *Rattus norvegicus*. In *Social learning: psychological and biological perspectives* (ed. T. R. Zentall and B. G. Galef), pp. 119–40. Lawrence Erlbaum Associates, Hillsdale, NJ.

Garcia, J., Hankins, W. G., and Rusiniak, K. (1974). Behavioral regulation of the milieu interne in man and rat. *Science*, **185**, 824–31.

Gigerenzer, G. (2000). *Adaptive thinking: rationality in the real world*. Oxford University Press, Oxford.

Gigerenzer, G., Todd, P., and the ABC Research Group. (1999). *Simple heuristics that make Us smart*. Oxford University Press, Oxford.

Goldman, A. (1986). *Epistemology and cognition*. Harvard University Press, Cambridge, MA.

Griffiths, P. E. (2002). What is Innateness? *The Monist*, **85**(1), 70–85.

Harris, M. (1979). *Cultural materialism: the struggle for a science of culture*. Random House, New York.

Henrich, J. (2000). Cultural Transmission and the Diffusion of Innovations. Unpublished Manuscript.

Henrich, J. and Boyd, R. (1998). The evolution of conformist transmission and the emergence of between group differences. *Evolution and Human Behavior*, **19**, 215–41.

Henrich, J. and Gil-White, F. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, **22**, 165–96.

Laland, K. and Brown, G. (2002). *Sense and nonsense: evolutionary perspectives on human behavior*. Oxford University Press, Oxford.

Lazarus, R. (1994). Universal antecedents of the emotions. In *The nature of emotions: fundamental questions* (ed. P. Ekman and R. J. Davidson), pp. 163–71. Oxford University Press, New York.

LeDoux, J. (1996). *The emotional brain*. Touchstone, New York.

(p.158) Levenson, R. (1994). Human emotion: a functional view. In *The nature of emotion: fundamental questions* (ed. P. Ekman and R. J. Davidson), pp. 123–30. Oxford University Press, New York.

Marks, I. M. (1987). *Fear, phobias, and rituals*. Oxford University Press, New York.

Marks I. M. and Nesse, R. (1994). Fear and fitness: and evolutionary analysis of anxiety disorders. *Ethology and Sociobiology*, **15**, 247–61.

Murphy, D. and Stich, S. (2000). Darwin in the madhouse: evolutionary psychology and the classification of mental disorders. In *Evolution and the human mind: modularity, language and meta-cognition* (ed. P. Carruthers and A. Chamberlain), pp. 62–92. Cambridge University Press, Cambridge, MA.

Nisbett, R. and Cohen, D. (1996). *Culture of honor*. Westview Press, Boulder, CO.

Ortony, A., Clore, G., and Collins, A. (1988). *The cognitive structure of emotions*. Cambridge University Press, New York.

Rogers, E. (1995). *The diffusion of innovations*, 4th edition. The Free Press, New York.

Rozin, P., Haidt, J., and McCauley, C. (2000). Disgust. In *Handbook of emotions* (ed. M. Lewis and J. Haviland), pp. 637–53. The Guilford Press, New York.

Samuels, R. (2002). Nativism in cognitive science. *Mind and Language*, **17**(3), 233–65.

Samuels, R., Stich, S., and Faucher, L. (2003). Reason and rationality. In *Handbook of epistemology* (ed. M. Sintonen, I. Niiniluoto, M. Sintonen, and J. Wolenski), pp. 1–50. Kluwer, Dordrecht.

Scherer, K. (1988). Criteria for emotion antecedent appraisal: a review. In *Cognitive perspectives on emotion and motivation*, Vol. 10 (ed. V. Hamilton and G. H. Bower), pp. 89–126. Nijhoff, Dodrecht.

Scherer, K. (1993). Studying the emotion-antecedent appraisal process. *Cognition and Emotion*, **7**(3/A), 325–55.

Simoons, F. (1994). *Eat not this flesh*. The University of Wisconsin Press, Madison.

Sperber, D. (1996). *Explaining culture: a naturalistic approach*. Cambridge University Press, Cambridge, MA.

Sripada, C. and Stich, S. (in preparation) The psychology and evolution of moral diversity.

Stich, S. (1990). *The fragmentation of reason*. MIT Press, Cambridge, MA.

Tooby, J. and Cosmides, L. (1990). The past explains the present: emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, **11**, 375–424.

Whitehead, H. (2000). *Food rules*. The University of Michigan Press, Ann Arbor.

Zental T. R. and Galef B. G. (1998). *Social learning: psychological and biological perspectives*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Notes:

(1) For a defense of reliabilist epistemology, see Goldman (1986); for a defense of pragmatist epistemology, see Stich (1990).

(2) See Laland and Brown (2002), chs. 3 and 4 for a useful overview of the relevant literature.

(3) In one experiment Ekman found that when Japanese subjects were shown unpleasant films in the presence of an authority figure they would begin the muscle contractions required to produce the facial expressions of negative emotions, but then immediately mask these expressions with a polite smile. American subjects, by contrast, made no attempt to mask the expression of negative emotions, nor did Japanese subjects when they viewed the distressing films alone. Ekman and his colleagues explained these findings by positing the existence of culturally local ‘display rules’ which can override or radically alter the pattern of emotional expression after an affect program has begun to unfold (Ekman 1972).

(4) There is much more to be said about the distinctions between norms, values and goals. But the details are a long story which we don't have time for here.

(5) For example, they disagree about the exact sequences of cognitive evaluations which occur during the appraisal process (Scherer 1993, Ortony et al 1988), about how appraisals might be neurally implemented (Chewlos and Oatley 1994) and even about the methods by which such questions could be answered (Ekman 1994). There is also considerable disagreement about how emotions should be typed, with Scherer and Ortony et al. claiming that appraisal principles are constitutive of emotion types, Frijda (1986) favoring an approach which also relies on emotion-specific action tendencies and Ekman favoring the individuation of emotions in terms of structural mechanisms. We will not take any sides in these important disputes.

(6) There are a number of ways in which the notion of resemblance can be formalized and made more precise. For the purposes of this paper, an intuitive notion of resemblance will suffice.

(7) It is important to keep in mind that ‘cultural parents’ need not be biologically related to (nor older than) their ‘cultural children.’ Moreover, with respect to a given trait, a cultural child may have one, several or many cultural parents.

(8) Two of Boyd and Richerson's favorite examples!

(9) Analogous to genetic fitness, the cultural fitness of a cultural variant is the variant's chances of surviving and being socially transmitted.

(10) Our terminology here differs slightly from Boyd and Richerson's; as they use the term, prestige bias counts as an ‘indirect bias’ but conformist bias does not.

(11) While we believe that certain strong innate fears dispose one to phobias, we do not mean to imply that all phobias arise from innate fears. The phobic fear of guns, hoplophobia, for example, presumably does not arise from an innate fear, but rather reflects the highly salient danger of these weapons in contemporary environments. Interestingly, at least one worker has argued that pathological fear of spiders—arachnophobia—may not arise from an innate fear, but rather from a culturally transmitted fear. Spiders appear to have become (falsely) associated with disease in Western societies during the Middle Ages during the course of several epidemics, and arachnophobia appears less prominent in non-Western cultures (Davey 1994). We thank Luc Faucher for bringing this case to our attention.

(12) While phobias clearly lead to significant dysfunction, it is less clear that they lead to a reduction in biological fitness, since claims about reproductive success are notoriously hard to test. Nevertheless, we believe that it is at least plausible that dysfunction due to phobias does indeed impact fitness. Additionally, it is worth noting that the dysfunction associated with phobias

clearly does impact well-being in the actual environment, which is another notion of 22 rationality that we distinguished earlier.

(13) An alternative suggestion is that cultures of honor are part of evoked, as opposed to transmitted, culture. That is, one might claim that innate human nature provides both culture of honor and non-culture of honor options, and that one or the other of these options is toggled by local environmental cues. We regard this suggestion as unlikely because agriculture and herding are too recent for it to be plausible that cultures of honor are evoked biological adaptations.

(14) Cohen and his colleagues have also conducted a series of experiments that suggest that honor norms are no longer as deeply internalized among contemporary Southerners as they once were. In particular, Southerners no longer view another person's failing to respond to insult as a reason to view him as less manly, though they continue to expect that others would view the person as less manly. The result is that people continue to heed honor norms because they mistakenly believe that others would think less of them if they did not. Cohen and Vandello (2001) call this state ‘a plurality of ignorance,’ which again underscores the maladaptive nature of honor norms in the contemporary South (see Cohen and Vandello 2001 for a review).

(15) An illustrative example is the case of third-world mothers who rapidly adopted the highly maladaptive practice of formula feeding, at least partly due to advertising deliberately designed to suggest that high-status people formula feed their own children (Rogers 1995).

(16) One well-known approach to explaining taboos draws on the functionalist perspective in anthropology which assumes that many seemingly arbitrary cultural practices are actually functional, and their adaptive benefits, which are often hidden, play an important role in explaining how the practices originated or are sustained. (For a contemporary approach to functionalism, see the defense of cultural materialism in Harris 1979). While there are few decisive arguments in the social sciences, the weight of evidence strongly suggests that, at least in the domain of food taboos, functionalist justification is often absent, and indeed many food taboos appear to be clearly maladaptive (Fessler unpublished manuscript). Particularly difficult for a functionalist hypothesis to explain are cases in which closely related groups living in the same environment, and otherwise sharing similar social structures, exhibit widely different food taboos. Henrich (2000) reports a series of cases of this type from the ethnographic literature. For example, among the Machigueanga of the Peruvian Amazon, snake, a widely available source of protein, is rejected even when it is known to be non-poisonous. Among the Warao of the Peruvian Orinoco river delta, large mammals are not hunted because they have ‘blood like people’.

Nevertheless, closely related groups living in the same region have no such prohibitions (Henrich personal communication).

(17) Fessler and Navarrete add that meat, for various reasons which we will not elaborate on here, may be disproportionately salient as a source of aversions.

(18) Fessler and Navarrete call this process ‘socially mediated ingestive conditioning’—it is a member of a family of social learning processes which humans and other animals use in the acquisition of food preferences (see Zental and Galef 1998 for discussion of a number of other social learning processes used in the domain of food).

(19) The process of normative moralization is poorly understood. For one account of this process, see Sripada and Stich (in preparation).

(20) In addition to conformist bias, there are other processes which maintain stable withingroup homogeneity and between-group differences. See Sripada and Stich (in preparation).