The Role of Moral Sentiments in Economic Decision Making

Timothy Ketelaar
Department of Psychology
New Mexico State University
Las Cruces, New Mexico
Ketelaar@nmsu.edu


4.14.05

Please do not quote or cite without author’s permission
Comments welcome, please send to Ketelaar@nmsu.edu.

WORD COUNT: 8,920
16 pages of text
The Role of Moral Sentiments in Economic Decision-Making

Explanations of economic behavior often fall into one of two camps: normative and descriptive. Normative accounts typically consist of attempts to fit a wide variety of phenomena into one broad theoretical framework consisting of a single set of axioms governing rational decision-making (e.g., von Neumann & Morgenstern, 1944). Descriptive approaches, on the other hand, consist of attempts to accurately describe how individuals actually behave, regardless of whether these accounts are consistent with any particular normative approach to rational decision-making (Gilovich, Griffin, & Kahneman, 2002; Hogarth & Reder, 1987). What makes this normative-descriptive distinction so intriguing is the observation that many descriptive accounts of economic behavior are often strikingly inconsistent with normative models. It turns out that this contradiction between what people “ought” to do, according to normative models, and what they actually do, as evidenced by descriptive accounts, is a puzzle that has intrigued observers of human nature for centuries.

Along these lines, 18th century moral philosophers such as David Hume (1777) and Adam Smith (1759) attempted to bridge the gap between what people ought to do (normative principles) and what they actually do (descriptive accounts) by exploring the role of “sentiments” in decision-making. Hume and Smith argued that moral sentiments, in the form of emotions and passions, can exert such a powerful influence on judgment that they often compete with more rational deliberations in determining our behavior. In essence, they proposed that emotions can provoke otherwise rational individuals to behave in a manner that is inconsistent with normative principles of “good” judgment. The present chapter expands upon this 18th century conception of moral sentiments to explore the role of emotion in contemporary economic decision research. After discussing the historical role that emotions have played in normative models of “good” judgment and decision-making, I provide a brief review of recent empirical work linking moral emotions to norm-obeying and norm-enforcing behaviors in experimental economics. Finally, it is argued that insights from evolutionary game theory can illuminate the possible adaptive design of moral sentiment systems and, in doing so, may help us understand why decision-makers often do not behave in the strictly “rational” manner that some normative models suggest that they “ought” to behave.

Emotions and Normative Models of “good” Judgment

Just as contemporary decision theorists often employ conflicting standards (e.g., normative vs. descriptive) to explain economic behavior, 18th century moral philosophers once grappled with an equally conflicting set of standards for discerning an individual’s moral character, namely, the standards of absolute moral perfection versus mere adherence to social convention. In his Theory of Moral Sentiments, for example, Adam Smith (1759, p. 29-30) argued that:

when we are determining the degree of blame or applause which seems due to any action, we very frequently make use of two different standards. The first is the idea of complete propriety and perfection....The second is the idea of that degree of proximity or distance from this complete perfection, which the actions of the greater part of man commonly arrive at.

Although the two standards to which Adam Smith referred are strikingly similar to the normative-descriptive distinction employed by contemporary decision researchers, Smith
and other moral philosophers of his era were more interested in the moral (rather than economic) implications of behavior. By contrast, contemporary economists and psychologists are much less interested in whether a given behavior approximates a particular standard of moral propriety and perfection. Instead, economists and psychologists are more concerned with the degree to which descriptive accounts of judgment and decision-making are consistent with particular normative standards of sound economic reasoning such as those standards embodied in the normative rules of inference of the logical calculus (von Neumann & Morgenstern, 1944). Just as 18th century moral philosophers regretted that individuals do not always behave as they ought to in the moral domain, so too, contemporary decision theorists often lament that economic decision-makers routinely enact preferences that are at sharp contrast with normative standards of proper economic behavior. Regardless of whether one considers violations of normative standards to indicate moral turpitude or merely economic irrationality, it is quite obvious that the distinction between what people “ought” to do and what most people “actually do” looms as large today in the minds of economists and psychologists as it did in the writings of 18th century moral philosophers.

Is there any reason to believe that contemporary decision theorists have anything to learn from 18th century moral philosophers? Although one might imagine that distinguishing the moral and logical implications of a particular behavior is a relatively straightforward task, it turns out that moral and logical implications are, in fact, quite often confused. Nowhere is this more evident than in our evaluations of behaviors that violate normative standards of “sound” reasoning are not only viewed as illogical, they are often met with judgments of moral disapproval that include strong emotional reactions. Even the most casual observer of human nature will note just how easy it is to conflate judgments of deviant behavior with judgments of immorality. Social psychologists, for example, have shown that merely observing a person violating a culturally-shared norm is enough to evoke strong moral emotions such as anger and contempt (Haidt, 2003; Rozin, Lowery, Imada, & Haidt, 1999; Ketelaar et al., 2005). Although the types of “non-normative” behaviors studied by social psychologists (e.g., stealing a purse from a blind person) are a bit far removed from the sorts of norm violations that interest economists (e.g., violations of the invariance axiom), empirical evidence now suggests that morally-charged emotions such as anger and guilt are routinely evoked in economic decision tasks that were once conceptualized as purely logical deliberations among rational participants (see Fehr & Gächter, 2002; Ketelaar & Au, 2003; Pillutla & Murnighan, 1996; Van Kleef, De Dreu, & Manstead, 2004; Zeelenberg & Pieters, 2005).

Given the ubiquitous role that emotions appear to play in everyday decisions (Demasio, 1994; Elster, 2000; Nussbaum, 2004), might there be room for a more formal treatment of the role of these moral sentiments in normative accounts of economic decision-making? Along these lines, the present chapter argues that a greater appreciation of the role of emotion might be informative in regards to two questions of interest to economists and psychologists: First, why are emotions so easily provoked in purely economic decisions? Second, why do emotions often compel individual's to pursue strategies that do not appear to be in their immediate material self-interest? It is argued that a large part of the puzzle concerning why “irrational” emotions so often intrude upon our most important decisions may lie in recognizing this link between moral sentiments and judgments of normatively acceptable (and unacceptable) behavior.
A brief history of Moral Sentiments and Decision-making

The role of emotion in judgment and decision-making research has perhaps been most strongly shaped by the views of David Hume for reasons having to do with his provocative treatment of the question regarding whether our moral judgments have their basis in our capacity for reason or instead in our capacity for moral sentiment (Gardner, 1987; Norton, 1993). In his Treatise on Human Nature, Hume (1740) initially espoused the quite radical view that reason was the “slave of the passions,” but later in his Enquires Concerning Human Understanding and Concerning the Principles of Morals (1777) he modified his position to portray moral sentiment and moral reasoning on more equal footing. Thus, although he eventually came to admit that rational deliberation played an important part in moral decision-making, what is most striking about Hume’s view is that he immediately recognized that emotion played a central role in the decision process. By contrast, 20th century psychology began with a relatively balanced treatment of emotion and reason (Gardner, 1987), but quickly became dominated by a strong strain of rationalism soon after the cognitive revolution of the 1960s. This rationalist surge could not have come at a worse time for researchers exploring the role of emotion in decision-making. Just as Von Neumann and Morgenstern’s (1944) Theory of Games and Economic Behavior was beginning to inspire social scientists to model social encounters as strategic interactions, emotions were being banished from normative models of strategic decision-making. Consistent with this rationalist slant, emotions were soon relegated to the status of disruptive influences on judgment and decision-making. For the next several decades (into the late 1980s) the ideas of Kant and Locke were alive and well, whereas the views of Hume and Smith were dead or dying (see Haidt, 2001 for a historical review).

The growing rationalism of post “cognitive-revolution” psychology gave a much different twist to the age-old question of why individuals sometimes strayed from, and at other times upheld, the normative path of “good” judgment. In contrast to the moral philosophers of the 18th century, the cognitively-oriented psychologists of the 20th century emphasized moral reasoning rather than moral sentiments as the basis for adherence to normative standards of “good” behavior (Haidt, 2001). In this light, the ideas of Kohlberg (1971) and Piaget (1932/1965) loomed large, whereas the views of Hume and Smith receded to the shadows. Individuals who acted in a normatively (morally) acceptable manner were believed to be employing a deliberate and conscious process of sound (moral) reasoning; whereas individuals who fell short of rational enlightenment and moral perfection were seen as being influenced by unnecessary and disruptive moral sentiments (Kohlberg, 1971). Although this rationalist view of emotions as disruptive influences was strongest in the moral judgment literature, rationalism managed to extend its reach into the more general literature on social judgment and decision-making (see Clore, 2005; Ketelaar & Clore, 1997). As one emotion researcher observed:

Today the image of man is no longer that of an individual enslaved by his passions, but rather that of a philosopher making decisions on the basis of logical deduction and inference. In this tradition, emotion is seen as a regrettable flaw in an otherwise perfect machine (Scherer, 1984, p. 293).

And the view from economics has not been much rosier:

The standard view of the relation between rationality and emotions is, of course, that emotions interfere with rationality. They are, as it were, sand in the machinery of action. (Elster, 1995, p. 1394).
Although emotionally influenced decisions often appear to be at odds with normative principles of rational inference (von Neumann & Morgenstern, 1944), perhaps consistency with the “cold rules of logic” is the wrong reference point from which to view these sentiments (Clore, 2005; Ketelaar & Clore, 1997; Ketelaar & Todd, 2001). In this regard, recent research on moral judgment suggests that emotions have more in common with relatively automatic processes such as intuition and perception than they do with more calculated processes such as rational deliberation and logical inference (Haidt, 2001). To the degree that our judgments, moral or economic, are often based on moral intuitions and emotional perceptions rather than logic and rational deliberation, one might suspect that the impact of emotion on economic decision-making has been greatly underestimated. Participants in studies of moral judgment, for example, have no difficulty generating “reasons” for their judgments; yet, further analysis reveals that many of these “reasons” are actually post hoc confabulations rather than veridical recollections of the actual thought process that lead to their judgment (Haidt, 2001). Moreover, participants in studies of moral judgment often report that they “don’t really know” why a particular behavior is morally wrong, it just is! (Haidt, 2001). Although these findings don’t necessary establish that moral reasoning is enslaved by the passions, they do suggest, however, that many decision processes that were previously thought to be the product of logical inference and rational deliberation are instead, more accurately described as being the product of relatively automatic emotional intuitions and perceptions. The remainder of this chapter explores how this insight might be integrated into our understanding of certain types of economic decisions.

**Empirical Evidence for Emotional influences on Economic Decision-making**

Emotion researchers have found it useful to conceptualize passions and sentiments in terms of their intrapersonal and interpersonal functions (Bowlby, 1969; Van Kleef et al., 2004). Intrapersonal functions refer to the impact that emotions have on individual decision-making, such as when post-decision regret motivates you to pursue an economic opportunity that you had previously rejected (Zeelenberg, 1999). Interpersonal functions, on the other hand, focus on the impact that emotions have on social decision-making, such as when anger motivates you to punish a selfish contributor in a public goods game (Fehr & Gächter, 2002). In relation to these intrapersonal and interpersonal functions, emotion theorists have identified at least two important clusters of moral sentiments: 1) self-focused sentiments such as shame, embarrassment, regret, and guilt and 2) other-focused sentiments such as contempt, anger, disgust, and schadenfreude (Haidt, 2003; Lewis, 1993; Ortony, Clore, & Collins, 1988). Self-focused emotions emphasize an evaluation of the appropriateness of one’s own behavior, whereas other-focused emotions tend to direct one’s attention to the appropriateness of the behavior of others. Although there exists a great variety of ways in which the relation between emotions and economic decisions can be conceptualized (see Zeelenberg & Pieters, 2005), the remainder of this chapter focuses on two areas in which theoretical and empirical developments have been quite illuminating: 1) self-focused emotions and norm-obeying behaviors and 2) other-focused emotions and norm-enforcing behaviors.

**Self-focused emotions and norm-obeying behavior**

Scottish Philosopher Adam Smith (1759) was among the first to develop a theory of how self-focused emotions can motivate an individual to obey (or disobey) certain normative
standards of “good” judgment. Smith argued that when we experience certain moral sentiments, such as guilt, they intuitively compel us to do what is morally appropriate, despite the fact that rational economic calculation might suggest otherwise. The sentiment of gratitude, for example, can compel an individual to repay an act of kindness, even when the cost of repayment exceeds the benefit initially bestowed upon the actor (Hirschleifer, 1987). So strong are these emotional influences on behavior that certain sentiments, such as greed or lust, can even motivate an individual to disobey these very same moral standards:

The man who acts according to the rules of perfect prudence, of strict justice, and of proper benevolence, may be said to be perfectly virtuous. But the most perfect knowledge of those rules will not alone enable him to act in this manner; his own passions are very apt to mislead him—sometimes to drive him, and sometimes to seduce him, to violate all the rules which he himself, in all his sober and cool hours, approves of (Smith, 1759, p. 349, emphasis added).

In essence, Smith (1759) argued that our capacity to obey normative standards of appropriate behavior was more often determined by our experience of certain moral sentiments than our employment of rational deduction and logical inference. Depending upon the nature of the particular emotion that one experienced (e.g., greed vs. guilt), Smith argued that moral sentiments could compel an individual to pursue, or avoid, a particular course of action even when this choice ran contrary to one’s immediate economic self-interest.

Although contemporary economists and psychologists are more concerned with earthly rewards than heavenly utilities, Smith’s insights regarding moral sentiments can be applied to questions that interest contemporary decision theorists. For example, why do emotions often compel individual’s to pursue economic strategies that do not appear to be in their immediate material self-interest? Along these lines, there is an ever-expanding body of theoretical and empirical research on self-focused emotions and economic decision-making that is consistent with Adam Smith’s insights regarding the influence of moral sentiments on decision-making (Van Kleef et al., 2004; Ketelaar, 2004; Ketelaar & Au, 2003). For example, economist Robert Frank’s (1988, 2004) theory of emotions as commitment devices essentially argues that the experience of self-focused emotions such guilt can explain why some individuals forgo the immediate rewards associated with cheating in favor of more cooperative strategies that pay their benefits, not immediately, but rather in the long run. According to this view, moral sentiments can:

- compete with feelings that spring from rational calculations about material payoffs....Consider, for example, a person capable of strong guilt feelings. This person will not cheat even when it is in her material interests to do so. The reason is not that she fears getting caught but that she simply does not want to cheat. Her aversion to feelings of guilt effectively alters the payoffs she faces (Frank, 1988, p. 53).

Because moral sentiments are experienced as powerful and un-ignoreable moral intuitions, an individual experiencing an emotion like guilt has two concurrent sources of information that are taken into account when deciding how to behave (Ketelaar & Clore, 1997). One source of information (from deliberate economic calculations) informs the individual about the immediate material consequences of a given strategy and the second source (from the
intuitive moral emotions) informs the individual about the future (in this case negative) consequences of that particular strategy choice. In this manner, the mental pain of guilt can compel an individual to not cheat, even when it is in their immediate material interest to do so (Frank, 1988).

For over one decade this view of moral sentiments (Frank, 1988) has been so compelling that it has been accepted at face value without being subject to empirical test. Recently, research at the intersection of social psychology and experimental economics has begun exploring the role of moral sentiments in economic decision-making (Ketelaar & Au, 2003; Van Kleef et al., 2004). Consistent with this emotions-as-commitment devices view (Frank, 1988; Hirschleifer, 1987) several studies have shown that individual differences in the capacity to experience feelings of guilt can translate into rather large differences in the magnitude of monetary offers made in the final round of a repeated ultimatum game (Haley & Fessler, 2005; Ketelaar & Au, 2003). In one study, over 91% of individuals who felt guilty after proposing an unfair monetary offer in the first round of play were later observed to make a generous offer in second round. By contrast, less than 23% of individuals who reported no feelings of guilt (after proposing a similarly unfair offer) were observed to make a generous offer in the second round. Similar findings have obtained in a repeated Prisoner’s Dilemma game where it was observed that individuals who felt guilty after behaving non-cooperatively subsequently displayed 25% more cooperation compared to those who felt no guilt (see Ketelaar & Au, 2003). Finally, in studies of post-decision regret in the ultimatum game, it has been shown that feelings of regret are a good predictor of subsequent offers when the game is repeated with another partner (Zeelenberg & Beattie, 1997). In one study participants reported their level of regret after receiving feedback informing them that their offer was either 1) only slightly higher than the responders minimal acceptable offer or 2) much higher than the responders minimal acceptable offer. Not surprising, participants who received feedback that their offer was much higher than the responders minimal acceptable offer reported significantly more regret than participants who were informed that their offer was only slightly higher than the responders minimal acceptable offer. Interestingly, when analyzed together, only feelings of regret (and not feedback about minimal acceptable offers) were able to predict the magnitude of proposers offers in the subsequent ultimatum game. Taken together, research on guilt and regret suggests that individuals differ in their capacity to experience particular moral sentiments and these individual differences are predictive of subsequent decision behavior (Schwartz, 19xx; Tangney, 1991). If self-focused moral emotions (e.g., regret, guilt) are a universal solution to the so-called commitment problem (Frank, 1988), one might wonder why we observe these individual differences in the capacity to experience guilt and regret? Nowhere is this puzzle regarding individual differences in moral sentiments more evident than in research examining other-focused emotions and norm-enforcing behavior.

Other-focused emotions and norm-enforcing behavior

The existence of individuals willing to enforce norms is an intriguing paradox for any model of decision-making that treats individuals as purely self-interested agents. This is the case because many norm-enforcing behaviors, ranging from specific acts of punishment to credible threats directed at potential norm violators, appear to be altruistic acts (Fehr & Gächter, 2002). The logic behind this claim is straightforward: individuals who display threatening signals and/or perform acts of punishment must directly incur the costs of carrying out these actions, yet the benefits accrued by these norm-enforcing behaviors are often distributed across the entire social group. In this manner one might argue that norm-
enforcing behaviors constitute a form of second-order public good (Fehr & Fischbacker, 2004). As such, one expects that most self-interested agents will refuse to contribute to this public good (fail to threaten or punish violators) and instead will free ride on the actions of few "punisher types" who enforce social norms for the entire group. Interestingly, experimental economists have demonstrated that a significant portion of individuals is often willing to incur a cost to inflict punishment on others who violate social norms such as under-contributing in a public goods game (Fehr & Fischbacker, 2004; Fehr & Gächter, 2002). This section explores the claim that emotions operate as the proximate psychological mechanisms underlying these norm-enforcing behaviors (Fehr & Gächter, 2002).

Just as certain moral sentiments seem to play a central role in motivating individuals to obey norms, other sentiments appear to play a central role in motivating individuals to enforce these norms. In particular, emotion researchers have identified three other-focused emotions -- contempt, anger, and disgust -- that are activated by detecting violations of normative standards (Rozin, Lowery, Imada, & Haidt, 1999). In addition to being associated with distinct facial signals, these three moral sentiments appear to be triggered by distinct types of norm violations (Haidt, 2003). Contempt, for example, has been shown to be associated with violations of normative rules regarding community standards and customs, as when one observes someone failing to carry out their duties in the community or social hierarchy. Anger, on the other hand, is associated with violations of normative standards governing one’s personal rights, as when one directly experiences another person infringing upon his/her personal liberties. Finally, disgust appears to be triggered by violations of normative standards governing purity and divinity, as when one observes another person disrespecting culturally-shared sacred beliefs or religious traditions. Moreover, these three moral sentiments appear to be excellent candidates for universal and culturally-shared reactions to particular types of norm violations:

"Within any culture, actions that are violations of the ethics of autonomy will be most likely to elicit anger; violations of the ethics of community will be most likely to elicit contempt, and violations of the ethics of divinity will be most likely to elicit disgust." (Rozin et al., 1999, p. 576)

This linkage between specific moral sentiments and specific types of norm violations suggests that the very same event can evoke any one of these three moral sentiments depending, of course, upon how the event is perceived. Consistent with this view, a sample of registered Democrats were observed in one study to spontaneously display a range of moral emotions (contempt, anger, and disgust) when asked how they felt about the very same event: the recent presidential election in which the Republican candidate had won (Ketelaar, 2005). In this context, one supposes that those individuals who reacted with contempt displays (62.5 % of the sample) were focusing on the impact that the election results had on their social group or community. Similarly individuals who reacted with anger (12.5%) may have been focusing on the impact that the election results had on their personal rights whereas individuals who reacted with disgust (25%) were presumably focusing on the impact that the election results had on fundamental beliefs that they held sacred. Such an interpretation would be consistent with the view that contempt, anger, and disgust involve subjective perceptions that can rise to systematic experiences of specific moral emotions in response to violations of specific types of normative standards of appropriate behavior. The clear empirical evidence for several distinct other-focused moral sentiments (contempt, anger, disgust, see Haidt, 2003) leads one to ask why an individual would possess powerful emotions to condemn behaviors in others that they (the observer)
are themselves never motivated to perform? Before discussing the intriguing question of
whether it is in one’s material self-interest to experience these other-condemning emotions, I
briefly review empirical research that suggests that these other-directed moral sentiments
can exert a strong influence on economic decision-making.

Several studies have demonstrated a clear link between anger and punitive behavior
in economic bargaining games. In one study of investment behavior in a public goods
game, punishment took the form of the ability to deduct monetary earnings from other
members of the collective after each round of play (Fehr & Gächter, 2002). Interestingly
these acts of punishment could only be made at a cost to the punisher and resulted in no
monetary gain for the punishing party. Consistent with the view that anger acts as a moral
sentiment that compels an individual to punish individuals who violate normative standards
governing one’s personal rights, these punishing acts were significantly correlated with self-
reports of anger and were typically executed by above average contributors and inflicted
upon below average contributors (Fehr & Gächter, 2002).

Anger has also been shown to motivate individuals to punish norm-violators by
spitefully rejecting monetary offers that appear to violate norms of fairness. In one study of
unfair ultimatum offers, it was observed that self-reported feelings of anger were a better
predictor of rejections of unfair offers than actual perceptions of unfairness (Pillutla &
Murnighan, 1996). One reason that angry reactions may not be synonymous with
perceptions of unfairness may be due to the fact that some unfair offers are not the result of
intentional actions and thus fail to elicit the appropriate attributions of responsibility and
blameworthiness that some theorists would claim are necessarily but not sufficient
conditions for producing the emotion of anger (Lerner, Goldberg, & Tetlock, 1998; Ortony,
Clore, & Collins, 1988; but see Kuppens, Van Mechelen, Smits, & de Boeck, 2003). Along
these lines, other studies have shown that feelings of anger are strongly correlated with
attributions of responsibility and blameworthiness directed towards one’s negotiation partner
(Allred, Mallozzi, Matsui, & Rai, 1997). Consistent with the claim that anger is a norm-
enforcing emotion, other studies have manipulated participant’s perceptions of the emotional
state of their interaction partner (angry versus happy) and find that participants make greater
concessions when they believe that their negotiation partner is angry than when they believe
their partner is happy (van Kleef et al., 2004). In short, there appears to be a clear link
between the experience of anger and the motivation to punish individuals who have violated
norms of appropriate behavior, even when these acts of punishment entail a cost for the
angry enforcer. Because anger is just one of several “other-directed” moral sentiments,
future research might explore the possible norm-enforcing role of additional other-focused
emotions such as contempt, disgust, and schadenfreude.

**Game-Theoretic insights into emotion-based strategy types**

From the perspective of normative models of rational decision-making, it is clear that
emotions are often associated with a number of peculiar economic behaviors. Guilt-prone
individuals often forgo monetary rewards by obeying norms that prescribe cooperation
rather than selfishness (Ketelaar & Au, 2003) and angry individuals often enforce norms by
punishing norm violators even when doing so provides a benefit to the group at a significant
cost to the punisher (Fehr & Gächter, 2002). As paradoxical as these norm-obeying and
norm-enforcing acts may appear to be, it must be emphasized that not all individuals
engage in such behavior. Thus, one might argue that a complete understanding of the
influence of emotion on norm-obeying and norm-enforcing behavior must also explain why
only a certain portion of the population behaves in this manner. In this final section I review several evolutionary game-theoretic insights regarding how individual differences in the capacity to experience certain moral sentiments could be functionally linked to individual differences in strategic economic behavior (Ketelaar, 2004).

One of the most compelling explanations for the apparently “irrational” influence of emotion on economic behavior centers on the claim that certain emotions function as commitment devices that compel individuals to maximize long-term payoffs at the expense of sometimes forgoing immediate rewards (Frank, 1988, 2004; Hirscheifer, 1987). However, this commitment device view of emotion does not explain why certain individuals (e.g., guilt-free individualists and emotion-less sociopaths) lack these important emotional commitments. In other words, if one wishes to explain the disruptive influence of emotion on economic decision-making by saying that certain emotions, such as guilt or gratitude, serve important functions, one can ask why don’t all individuals experience these putatively functional sentiments? In one study of the ultimatum game, for example, only 57% of individuals reported experiencing feelings of guilt after proposing an unfair split of the money (Ketelaar & Au, 2003). However, these individual differences in emotion resulted in quite striking differences in ultimatum offers in the second round of play. When the ultimatum game was repeated one week later, over 91% of individuals who felt guilty (the previous week) now made generous offers, whereas only 22% of individuals who reported no feelings of guilt made similarly generous offers. Such strategically relevant individual differences are not limited to studies of emotional influences on norm-obeying behavior, they also appear to play a prominent role in studies of norm-enforcing behavior. In one study of “altruistic punishment,” for example, approximately 16% of the sample never punished other group members even when they contributed significantly less than the group average in a public goods game. Moreover, only one third of the sample consistently (across several rounds of play) punished group members who made deviant contributions and these costly acts of punishment were shown to be strongly linked to individual differences in the experience of anger directed toward low contributors (Fehr & Gächter, 2002). In sum, not all individuals are compelled by feelings of guilt to resist the temptation to defect (Ketelaar & Au, 2003) and not everyone becomes angry enough to incur the cost of punishing norm violators (Fehr & Gächter, 2002). Although the emotions-as-commitment-device approach is able to explain how emotions can compel individuals to obey norms or punish norm violators, such models leave open the question of why not everyone behaves as if they are emotionally committed to these norm-obeying and norm-enforcing actions.

**Strategies as Decision Rules**

Although Adam Smith (1759) explained individual differences in the capacity to experience moral sentiments in terms of a failure to exercise “self-command” over these emotions, one can also utilize evolutionary game theory to explore another, more intriguing explanation for these individual differences. Specifically, one can examine the possibility that individual differences in the capacity to experience moral sentiments may represent a relatively stable (frequency-dependent) distribution of emotion-based strategy types (Ketelaar, 2004; Mealey, 1995). One way to think about individual differences in emotion-based strategy types is to think of strategies as analogous to computer programmed decision rules (Binmore, 1998). According to this way of thinking there are at least two ways that multiple strategies (i.e., individual differences in strategic behavior) can manifest themselves in a population of agents. First, a population may contain multiple behavioral strategies because the population consists of several distinct strategy types, each
Emotions and Economic Decision-Making

corresponding to a distinct decision rule governing each agent’s behavior (e.g., compare “IF you cooperate, THEN I will cooperate” to “IF you cooperate, THEN I will defect”). A second reason that one might observe different behavioral strategies in the same population concerns the presence of certain types of reactive strategies (see Tooby & Cosmides, 1990). In other words, a population may consist of just one strategy type, but this strategy type may correspond to a complex decision rule that generates different behavioral phenotypes contingent upon the phenotypic characteristics of the other agents that it encounters. For example, imagine a decision rule of the form “IF I am weaker than you AND IF you cooperate, THEN I will cooperate; ELSE IF I am stronger than you AND IF you cooperate, THEN I will defect.” A reactive strategy of this form will appear cooperative when interacting with strong opponents and non-cooperative when encountering weak opponents. The presence of such a reactive strategy might lead one to the erroneous conclusion that their exists two strategy types in the population (cooperators and non-cooperators) when, in reality, the population is comprised of just one strategy type that reacts differentially to particular phenotypic characteristics (e.g., relative strength) that vary across agents.

Finally there exists a third means of obtaining a diversity of strategy types in a single population of agents and I argue that this explanation may be key to understanding the influence of emotions on economic decision-making. This third explanation for multiple strategies is a hybrid of the two types of explanation discussed above. According to this hybrid view, all agents possess essentially the same reactive decision rule (e.g., IF X THEN Y, but only IF B > C) for determining strategic behavior, yet each agent differs in terms of the actual values that they assign to certain parameters in this decision rule. To appreciate this idea, imagine an evolved decision rule comprised of a number of parameters (X, Y, B, C, etc.), some of which are invariant across individuals and others that vary systematically across individuals. Consider, for example, a decision rule that essentially translates into the following form “IF you fail to contribute to the public good by some factor X, THEN I will punish you by some factor Y, but only IF the benefit (B) of inflicting this punishment is greater than the cost (C) associated with punishing you. Further imagine that the amount of punishment (Y) that is inflicted on a norm violator is a function of the degree of norm violation entailed by X, and moreover, that this parameter is some fixed function \(Y = f(X)\) shared by all agents in the population. Thus, all agents who encounter a given instance of norm violation will necessarily punish the same amount as determined by the function \(Y = f(X)\). However, now imagine that the two cost-benefit parameters \(C\) and \(B\) in this hypothetical decision rule] are determined by natural selection according to their evolutionary success. Because agents are constantly interacting with each other across time and space, one might imagine that natural selection could eventually converge on a single algorithm that determines how all agents will calculate these cost and benefit parameters (C and B). Because the algorithm that calculates these cost-benefit parameters would essentially correspond to a sort of threshold for determining whether or not an agent will punish a particular instance of norm violation, if all agents possess the same algorithm for calculating these parameter values, all agents will essentially possess the same behavioral phenotype in regard to punishing norm violators. By contrast, if different agents possessed somewhat different algorithms for calculating these parameters (costs and benefits of punishing) this would necessarily generate several distinct behavioral phenotypes (i.e., punishers and non-punishers). Therein lies the rub, or at least a possible evolutionary source for individual differences in punishment behavior or any other strategic behavior for that matter. The only difference between this explanation for multiple strategies and the reactive strategy explanation (discussed above) is the subtle point that
several distinct strategy types can emerge merely as a result of subtle differences in how
certain parameter values are calculated in the agent’s decision rule. How might this work?

If agents are analogous to pre-programmed decision rules and evolutionary
selection pressures do not converge on just method for calculating parameter values for all
agents, the population, by definition, will consist of several different strategy types. For
example, if evolutionary dynamics allowed for different agents to possess different values for
the cost-benefit parameters contained in the hypothetical punishment decision rule “IF X
THEN Y, but only IF B > C,” this could result in a diversity of strategy types in regards to
their propensity to punish norm violators. Depending on how these cost-benefit parameters
varied, some versions of this rule would dictate that certain agents never punished
(punishment would always be perceived as too costly), whereas other versions of this rule,
with somewhat different cost-benefit parameters, would dictate that certain agents would
invariably punish every instance of norm violation that they encountered (it would always be
perceived as beneficial to punish). Yet, one could imagine that evolutionary forces might
invariably converge on just one method of calculating these cost-benefit parameters for all
agents, whereby the population of agents (decision rules) that emerges is at some sort of
game theoretic equilibrium. Yet, it turns out that the ecological conditions under which such
decision rules often operate (i.e., indefinitely repeated interactions) lend themselves to a
vast number of possible equilibria and thus, an equally vast number of possible decision
rules. The implication of this insight---known as the folk theorem—is that evolutionary
selection pressures are capable of generating a vast number of successful distributions of
strategy types (each consisting of a distinct decision rule) and thus, one must consider the
possibility that human populations may be comprised of a multitude of decision rules
governing the implementation of norm-obeying and norm-enforcing behaviors.

Evolutionary game theory gives us several hints at how dynamic forces (such as
those encapsulated in the concept of frequency dependent selection and the folk theorem)
could give rise to more than one equilibria in a single population of agents (Binmore, 1998;
Boyd, 1989; Boy & Richardson, 1992; Maynard Smith, 1982). For example, if one considers
strategy types as analogous to finite decision rules (finite automata) that encounter one
another in indefinitely repeated games, an interesting phenomenon emerges. Specifically,
one observes that:

“any two finite automata playing each other for long enough in a repeated game
will eventually end up cycling through the same sequence of plays. A player’s
per-game payoff can therefore be calculated simply by taking the average payoff
in a cycle” (Binmore, 1998, p. 301).

For example, if the finite strategy of tit-for-tat interacts with itself (or any other finite strategy
that is never the first to defect) these “two” strategies will end up in a continuous cycle of
mutual cooperations and obtain whatever payoffs are associated with this particular cycle.
By contrast, tit-for-tat interacting with a finite strategy of “always defect” will end up in a
continuous cycle of mutual defections and obtain a necessarily different set of payoffs
associated with this very different cycle. The folk theorem emerges from the simple fact that
because finite strategies invariably end up infinitely cycling through a particular sequence of
outcomes, (mutual cooperation, mutual defection, etc.), this sets up an incentive for either
agent (i.e., each finite decision rule) to locate and shift to an alternative pairing of strategies
that can generate a different cycle with a greater payoff than that obtained in their current
cycle (Binmore, 1998). It follows that indefinitely repeated games that occur in a diverse
population of strategy types can allow for a great variety of equilibria simply because this
context (indefinitely repeated interactions, different strategy types) allows for a vast number of more profitable alternative cycles upon which various strategy types can coordinate (see Lomborg, 1996, Binmore, 1998 for a fuller treatment of this so-called Folk Theorem). Because individuals who experience different emotions (e.g., guilt versus no guilt, anger versus no anger) often end up making quite different choices in economic bargaining games, it may be useful to explore the possibility that individual differences in moral sentiments constitute one source of these individual differences in strategy types.

**Emotional Gambits: Moral sentiments as pre-programmed decision rules**

The idea that individual differences in emotion can generate different strategy types can be illustrated by considering a simple analogy borrowed from chess-playing computers. In this regard we have already encountered one important source of individual differences in behavior that is commonplace in the world of computer chess, namely, the idea that strategies can be conceptualized as preprogrammed decision rules. After all, different chess playing computers (e.g., Deep Blue vs. Fritz) are essentially just different strategies for playing chess and as such, they will often generate quite different strategy choices when encountering the very same strategic scenario. Although computer programs for playing chess openings and chess endgames are essentially look-up tables that involve little computation to implement successful game-playing behavior (Newborn, 1996), it turns out that the typical chess mid-game presents a vastly more complex puzzle that requires dedicated computational machinery in the form of special purpose hardware and software components (Hsu, 2002; Levy & Newborn, 1991; Newborn, 1996). Because looking ahead just 10 moves in the mid-game can involve constructing a game tree containing at least 40 billion sequences of behavior, Shannon (1950) proposed a much simpler strategy (compared to considering all possible moves) in which a move “evaluation function” plays a central role in selecting a chess playing computer’s next move. To understand how emotions might effectively alter an agent’s perceptions of the payoffs in a repeated economic bargaining game it might be useful to borrow several insights from the world of chess computers regarding how these move evaluation functions determine the payoffs associated with alternative strategy choices.

Despite the fact that computer programs such as Deep Blue are capable of searching much deeper into the game tree than any unaided human strategist could ever manage, the key to a successful computer chess program lies not in its processing speed and power, but rather in the decision rules that it employs to select its next move (see Hsu, 2002; Newborn, 1996). In one approach, known as the “fixed depth” search, the computer’s next move is selected by:

exploring all lines of play to some fixed depth and then assigning a score to the position at the end of each continuation. The score assigned to the position by a “scoring function” is a measure of how good the position is for the side on the move (Newborn, 1996, p. 9).

This simple strategy can be summarized as follows: 1) look ahead a specific, but limited, number of moves, 2) assign a score to the terminal position at the end of each continuation, and then 3) select a move that guarantees progress toward the terminal positions with the higher scores (see Shannon, 1950; Newborn, 1996 for more detail). Experimental economists will recognize this as a process of automated backward induction.

To appreciate how emotions might be analogous to a computer chess program’s move evaluation function we need to say a bit more about how these computer chess
programs actually work. The move evaluation function of a chess-playing computer is essentially an algorithm that determines how a particular mid-game position on the “game board” can be translated into a new (and presumably improved) position. Typical move evaluation functions accomplish this task by employing two components: a) a scoring function and b) a move selection algorithm (Newborn, 1996; Shannon, 1950). The scoring function of a chess playing computer is essentially an algorithm that assigns points to each of the terminal positions in the game tree based upon a number of situational factors including relative material (piece) advantage, the mobility of one's pieces, et cetera. In the case of Deep Blue, the computer programmers relied heavily upon the advice of Chess Grand Masters to determine the algorithms for computing these scores (Hsu, 2002; Newborn, 1996). The move selection algorithm, on the other hand, is essentially a decision rule that sorts through these scores and employs backward induction to locate not where you want to end up, but rather how best to execute your next move given your knowledge of these future good and bad endings identified by the scoring function. There are a variety of algorithms that can be used for move selection. Today many chess computers employ von Neumann and Morgenstern’s (1944) loss averse minimax algorithm for selecting moves based upon backward induction (see Newborn, 1996). In this final section, I consider how emotions might be analogous to each of these processes: scoring functions, move selection algorithms, and canned programs.

Utilizing a decision rule known as a scoring function chess playing computers often appear to “act” as if certain behaviors (chess moves) are “good” and others are “bad,” not because the immediate payoffs associated with these moves are good or bad, but rather because the scoring function has computed their likely future payoffs. It is in precisely this sense that emotions may be analogous to the scoring functions of chess playing computers. Recall that one interpretation of Frank’s (1988, 2004) view—that emotions act as commitment devices—suggests that moral sentiments (such as guilt) function to move the future costs of particular actions (such as defection) into the current situation, in the form of a powerful and overwhelming feeling state (e.g., negative affect) evoked by merely contemplating the pursuit of a particular strategic option (Ketelaar & Todd, 2001). One mechanism that could generate these sorts of individually-different, emotionally-biased payoff estimates is evolution by natural selection. In particular, it is not inconceivable that evolutionary selection forces (including frequency dependent selection) could shape the algorithms that determine the specific parameter values in an agent’s decision rule such that these parameter values correspond to estimates of future payoffs rather than rational calculations of immediate payoffs. According to this view, evolved emotional decision weights could determine an agent’s preferences for particular strategy options by serving as psychological “stand-ins” for the imperceptible (and often incalculable) future payoffs associated with pursuing each of those options. In this way emotions can be seen as analogous to the “scoring function” of chess-playing computers in the sense that emotions compel an individual to assign disproportionate weight to the final outcome (rather than the immediate payoff) associated with a particular strategy choice (Ketelaar & Todd, 2001; see Kahneman 1999 for a review of the empirical evidence regarding the emotional over-weighting of future outcomes).

By operating as estimates of future (rather than immediate) payoffs, emotional decision weights could provoke individuals to make strategic gambits in repeated social interactions. In chess, a gambit is a strategic sacrifice, often made very early in the game, when one incurs a significant immediate loss in exchange for a long-term strategic advantage. One might argue the emotions-as-commitment device model essentially
portrays emotions as operating in the same fashion, as devices for provoking strategic gambits in the iterated game known as social life (Ketelaar & Todd, 2001). For example, Frank (1988, p. 53) argues that a person who is guilt-prone may be compelled to avoid a strategic move that yields an immediate benefit (i.e., the temptation to defect) because the "aversion to feelings of guilt effectively alters the payoffs she faces.” The idea is, of course, that the guilt prone individual is making a strategic sacrifice by forgoing an immediate reward in favor of pursuing a strategy that will reap still greater rewards in the future. By contrast, the guilt-free individualist might perceive a very different set of payoffs and thus pursue a very different strategy path (they might actually take the immediate reward associated with defection). It is as if guilt-prone and guilt-free individuals possess somewhat different parameter values in their “scoring functions”. Researchers in the world of computer chess are keenly aware of how these relatively small differences in the parameter values of an agent's scoring function can result in rather striking differences in the behavioral strategies that the agent employs (Hsu, 2002; Newborn, 1996). In sum, if emotions operate in a manner analogous to the scoring function employed by chess-playing computer programs one expects that emotions might play an important role in determining the payoffs that individuals assign to strategy options. Moreover, individual differences in emotion-based strategy types may reflect a frequency dependent distribution of possible payoff structures rather than irrational and mysterious deviations from a single equilibrium strategy.

Finally, the analogy to chess-playing computers suggests at least two more mechanisms through which emotions could influence strategic decision-making: 1) emotions as move selection algorithms or 2) emotions as canned programs. In this context, move selection algorithms function to determine the strategies of backward induction that an agent will employ once scores have been assigned (via the scoring function) to the terminal positions in the game tree. As previously mentioned, many chess playing computers employ a loss averse move selection algorithm (e.g., Von Neumann & Morgenstern’s minimax, see Newborn, 1996). However, there are numerous other alternative algorithms (some loss averse, others gain-seeking) that could conceivably be employed in this process of backward induction. Along these lines, it is not inconceivable that different emotional states (fear vs. happiness) could be associated with different variations (some loss averse, some gain-seeking) of these move selection algorithms. Lastly, emotions could also influence economic decision-making by essentially operating as canned programs that merely locate the appropriate preprogrammed decision rule for mindlessly (without calculating payoffs) playing a particular sub-game that it has just stumbled upon (e.g., a particularly successful end-game strategy). This emotions-as-canned programs explanation suggests that emotions do not correspond to on-line devices for calculating payoffs (scoring functions) and then selecting moves (move selection algorithms), but rather that emotions merely operate as automatic decision rules that simply: 1) match the current situation to a pre-existing template corresponding to a particular sub-game stored in long-term memory and then 2) call up the appropriate preprogrammed algorithm for playing that sub-game. Again, it is not inconceivable that evolutionary selection pressures could evolve specific emotional subroutines (decision rules) that vary across individuals and which are automatically triggered by certain scenarios, much like the canned programs that different chess-playing computers use to play certain endgames. Thus, when we observe Agent A getting angry and Agent B maintaining her cool in the very same situation, we might surmise that Agent A is pre-equipped with a canned program for reacting to this situation, whereas Agent B is not. In this manner, the presence or absence of various canned programs would constitute different strategy types. It would be intriguing to articulate how these sorts of
different strategy types could be maintained in the same population via a process of frequency dependent selection of decision rule parameters (the hybrid explanation discussed above seems like a good starting place for this discussion). In sum, if emotions operate as canned programs this would suggest that certain moral sentiment systems correspond to rather mindless (e.g., they do not rationally calculate payoffs) decision rules that blindly execute a particular behavioral strategy when encountering a particular evolutionary recurrent scenario. Such an approach seems consist with research that shows that moral sentiments, such as anger, behave like rather automatic systems of moral intuition (canned programs) that blindly execute a particular strategy when encountering a particular instance of norm violation (Haidt, 2001). Future research might explore whether any of these three computer-chess inspired ideas—scoring functions, move selection algorithms, canned programs—can help social psychologists and experimental economists appreciate why emotions are so routinely associated with individual differences in norm-obeying and norm-enforcing behavior.

**Conclusions**

Moral sentiments—such as feelings of anger—are easily labeled as “irrational” when they are observed to motivate individuals to engage in behaviors (e.g., altruistic punishment) that benefit others but are costly to the individual who performs them. To the degree that the psychological mechanisms underlying our emotions have been shaped by natural selection, one wonders how adaptively-designed moral sentiment systems could routinely generate preferences that compel individuals to pursue outcomes with lower (rather than higher) payoffs? This chapter argues that an intriguing answer to this question may lie in the realization that not all individuals confront identical payoffs, even when they find themselves in the same situation (see Kelley, 1984; Kelley & Thibaut, 1978). For example, an agent who embodies the angry punisher type may experience a quite different set of payoffs associated with inflicting punishment as compared to a relatively anger-less non-punishing type. In this manner, if one is forced to select just one set of payoffs as the “rational” way of viewing the utilities associated with punishing and not punishing, by definition, one of these two agents (strategy types) will be viewed as possessing irrational preferences. Yet, if the payoffs associated with punishment are determined not by one theoretical gold standard, but rather by virtue of the evolutionary (including frequency dependent) success of the specific parameter settings contained in an agent’s decision rule, one might wonder whether the specific parameter values, and the corresponding “payoff estimates” that they generate, must necessarily be identical for all agents? Just as the predicted payoff for approaching elephant dung depends upon whether you are a human or a dung beetle, it may turn out that the ultimate future payoff associated with punishing norm violators depends upon which strategy type you are, as well as the particular distribution of other strategy types in the same environment (recall the hypothetical reactive decision rule for cooperating or defecting as a function of one’s relative strength). Along these lines, it was argued that individual differences in the capacity to experience certain moral sentiments might correspond to frequency-dependent individual differences in the settings (values) of certain decision rule parameters that influence norm-obeying and norm-enforcing behavior. The obvious implication is that by virtue of their ability to generate individual differences in how payoffs are assigned to outcomes, emotion-based strategies might play a significant role in explaining why some decision-makers often do not act in the strictly “rational” manner that some normative models suggest that they “ought” to behave. Along these lines, the present chapter argued that a large part of the puzzle concerning why “irrational” emotions so often
intrude upon our most important decisions lies in recognizing this important link between moral sentiments and judgments of normatively acceptable and unacceptable behavior.
References


Schwartz, B. (19xx). Individual differences in regret citation


---

i Schadenfreude is a German term that refers to the experience of positive emotions while observing another’s misfortune.

ii Frequency dependent selection occurs when the evolutionary success of a strategy type is dependent upon the relative frequency of certain other strategy types in the population (Maynard Smith, 1982).

iii This is a purely hypothetical decision rule that is merely used to illustrate how certain parameters of decision rules might be subject to frequency dependent selection. It is not meant to be taken literally as a model of the determinants of norm-enforcing behaviors (e.g., punishment).

iv The idea here is that even the apparent simple decision rule “IF X THEN Y, but only IF B > C” may contain numerous subroutines (specific algorithms) that determine each of the parameters. For example, we have already considered the possibility that Y is a function of X. The argument that follows, concerning the possible origins for individual differences in the proclivity to punish norm violators, implicitly assumes that different agents could have evolved somewhat different subroutines and functions for determining these parameter values, despite the fact that the overarching decision rule being executed [IF X THEN Y, but only IF B > C] is essentially the same for all agents in the population.

v This same logic could be used to explain individual differences in the proclivity to reward other agents.

vi This argument does not depend on these decision rules being hard-wired or “pre-programmed” so much as it depends upon them being algorithmic. One could imagine an
algorithmic decision rule with ontogenetically (developmentally) calibrated parameters that begin with some pre-programmed starting values that are then subject to change contingent upon the environmental stimuli that the rule encounters.

The claim is not that a cycle will invariably end up repeating just one type of outcome over and over again, such as a string of continuous mutual defections or a string of continuous mutual cooperations. These cycles may correspond to a great variety of rather random looking, but ultimately repeating, sequences of outcomes (see Binmore, 1998 for a nice exposition of this argument.

This argument depends, of course, upon agents being able to maintain credible commitments to sticking with a particular alternative cycle of payoffs (Lomborg, 1996). The claim here is that emotions are a plausible source of these credible commitments (Hirshleifer, 1987; 2002; Ketelaar, 2004).

On the hardware side, successful chess playing computers employ specialized processing chips designed for performing chess mid-game computations (Hsu, 2002).

The size of the game tree that is constructed is determined by two additional components: 1) a move generation program and 2) a search control procedure (Hsu, 2002). In this context, the Move generation program simply locates the current set of legally possible moves as constrained by the search control mechanism whose function it is to limit the scope of the game tree under consideration by pruning it’s breadth and/or depth (see Shannon, 1950; Newborn, 1996 for a discussion of the most common pruning techniques).

In a population comprised largely of defectors, for example, it is more costly to be a punisher than in an environment comprised largely of norm-obeying agents. Along these lines, Maynard Smith (1982) has shown that when the evolutionary success of a strategy type is dependent upon the relative frequency of certain other strategy types in the population, something very interesting can happen. Specifically, under certain circumstances several strategy types can be maintained in a sort of polymorphic equilibrium of strategy types in which each strategy type will experience lower long-term payoffs if it becomes more or less frequent in the population. This evolutionary selection process in which the success of each strategy type depends upon the relative frequency of other strategy types in the population is known as frequency dependent selection (see Lomborg, 1996 for simulation results consistent with these insights, see also Boyd, other references for related insights).