# The Heat of the Moment: Modeling Interactions Between Affect and Deliberation

George Loewenstein
Department of Social and Decision Sciences
Carnegie Mellon University

Ted O'Donoghue
Department of Economics
Cornell University

March 2007

**Abstract**

Drawing on diverse lines of research in psychology, decision making, and neuroscience, we develop a model in which a person's behavior is determined by an interaction between *deliberative processes* that assess options with a broad, goal-based perspective, and *affective processes* that encompass emotions and motivational drives. Our model provides a framework for understanding many departures from rationality discussed in the literature, and captures the familiar feeling of being "of two minds." Most importantly, by focusing on factors that moderate the relative influence of the two processes, our model generates a variety of novel testable predictions. We apply our model to time preferences, risk preferences, and social preferences.

Acknowledgements: For useful comments, we thank David Laibson, Roland Bénabou, Andrew Caplin, Andrew Schotter, Antonio Rangel, John Hamman, Shane Frederick, Joachim Vosgerau, and seminar participants at Princeton University, Duke University, New York University, UC Berkeley, University of Chicago, MIT, and the 2004 ASSA meetings in San Diego. We also thank Christoph Vanberg for valuable research assistance. For financial support, Loewenstein thanks the Integrated Study of the Human Dimensions of Global Change at Carnegie Mellon University (NSF grant # SBR-9521914), and O'Donoghue thanks the National Science Foundation (grant SES-0214043).

> In me there meet a combination of antithetical elements which are at eternal war with one another. Driven hither by objective influences, thither by subjective emotions; wafted one moment into blazing day, by mocking hope; plunged the next into the Cimmerian darkness of tangible despair, I am but a living ganglion of irreconcilable antagonisms.
>
> Gilbert and Sullivan, HMS Pinafore

## I. Introduction

From the writings of the earliest philosophers, there has been an almost unbroken belief among philosophers and psychologists that human behavior is perhaps best understood as the product of two interacting and often competing processes. Referred to at different times as "passion versus reason," "the id and the ego," and more recently "emotion and cognition," the changing labels obscure a surprising degree of consensus about the rough outlines of the two processes --- one more deliberative and focused on broader goals, the other more reflexive and driven by emotions and motivational drives. These ancient intuitions about the human mind have only been reinforced by modern-day neuroscientists, many of whom have argued that human behavior is driven by competing neural processes in the brain.

When it comes to formal modeling, however, one process --- the more cognitive of the two --- has received the lion's share of attention. A considerable amount of intellectual time and energy has gone into formulating what are sometimes referred to as cognitive, or rational-choice, models of decision making, such as the expected utility model and the discounted utility model. Such models are consequentialist in character; they assume that people choose between different courses of action based on the desirability of their consequences. Attempts to increase the realism of such models, many associated with the field of behavioral decision research, have generally adhered to the consequentialist perspective, but modify assumptions about probability weighting, time discounting, or the determinants of the utility or value from consequences.

A major reason for this focus is that the second process --- the more emotional or affective --- has long been viewed as an erratic, unpredictable force that is too complicated to incorporate into formal models of behavior. In recent years, however, there has been a renewed interest in emotion. New research by social psychologists (Epstein, 1994; Sloman 1996; Wilson, Lindsey & Schooler, 2000), neuroscientists (Damasio, 1994; Panksepp, 1998; LeDoux, 1996; Rolls, 1999) and decision researchers (Slovic et al, 2002; Peters & Slovic, 2000; Pham, 1998; Loewenstein, 1996; Loewenstein et al., 2001; Mellers, Schwartz & Ritov, 1997; Lerner &

Keltner, 2000, 2001) has led to a better understanding of the role that emotion plays in decision making, much of it lending new support to the ancient intuitions about human behavior. As of yet, however, there have been few attempts to develop formal, mathematical, models of behavior that incorporate these insights, and in particular to address how emotional or affective processes combine with deliberation to determine human behavior.

In this paper, we develop a formal, dual-process model of interactions between deliberation and affect. We assume that a person's behavior is influenced both by *deliberative processes* that assess options with a broad, goal-based perspective (along the lines of the standard consequentialist conception) and by *affective processes* that encompass emotions such as anger and fear and motivational drives such as those involving hunger and sex. Our model provides a conceptual framework for understanding many of the documented departures from the standard rational-choice model discussed in behavioral decision research and behavioral economics. At the same time, it captures the familiar feeling of being "of two minds" — of simultaneously wishing one were behaving one way while actually behaving in a different way. Most importantly, by focusing on factors that moderate the relative influence of the two processes, our model generates a number of novel testable predictions.

In Section II, we motivate and describe our general model of interactions between affect and deliberation. Our formal model has three key features. First and foremost, there are two competing systems with distinct objectives, which we formalize by assuming that each system has its own objective function. Second, each system is potentially influenced by environmental stimuli. But since this effect is more pronounced --- and more predictable --- for the affective system, our formal model focuses on how the *intensity* of affective motivations depends on factors such as the temporal and non-temporal proximity of reward and cost stimuli. Third, behavior is the outcome of activity in both systems. The key issue, then, is how the two systems are combined to generate behavior. Based on research by Baumeister and colleagues (e.g., Baumeister & Vohs, 2003), we assume that the affective system has initial control, but that the deliberative system can influence behavior through the exertion of effort, or *willpower*. Our model has two major sources of predictions. First, there are two factors --- willpower depletion and cognitive load --- that tend to reduce a person's ability to exert willpower and thus push behavior further from the deliberative optimum. Second, factors that predictably increase the intensity of affective motivations will also push behavior further from the affective optimum unless stronger affective motivations tend to better align the objectives of the two systems.

Our general model can be applied quite broadly. To illustrate, in Sections III, IV, and V, we apply our model to three specific domains. In each application, we begin by specifying the objectives of the two systems, guided by introspection and, when possible, direct (e.g., neural) evidence. Roughly speaking, our general modeling strategy is to assume that the deliberative system accords with normative models of decision making, whereas the affective system incorporates many descriptive features of decision making identified in psychology. Once we have specified the objective functions, we demonstrate how the general predictions described above become more precise and testable.

In Section III, we apply our model to intertemporal choice. We make the natural assumption that the affective system is primarily driven by short-term outcomes, whereas the deliberative system cares about both short-term and longer-term outcomes. These assumptions imply that a person will exhibit hyperbolic time discounting even if the deliberative system weighs costs and benefits exponentially. More importantly, our model generates novel testable implications about what situational factors should make time preferences more or less hyperbolic.

In Section IV, we apply our model to risky decision making. A natural assumption for the deliberative system is that its objectives correspond roughly to expected-utility theory. It is less obvious what to assume about the affective system; the perspective we suggest is that two prominent features of many descriptive theories of risk preferences --- S-shaped probability weighting and loss aversion --- derive from the affective system. With this perspective, our model generates novel testable implications in terms of factors that should make people more or less S-shaped and factors that should make people more or less loss averse.

In Section V, we apply our model to social preferences, and describe specifically how it can be applied to altruistic preferences. Here, the perspective we suggest is that the deliberative system has a stable concern for others driven by moral and ethical principles for how one ought to behave, whereas the affective system can be driven toward behaviors at any point between the extremes of pure self-interest and extreme altruism depending on the degree of sympathy that is triggered. Once again, with this perspective, our model generates novel testable predictions in terms of factors that should lead people to exhibit more or less altruism.

In Section VI, we focus on the extent to which our model can help address a major question in behavioral decision research and behavioral economics: when we investigate the public-policy implications of our research, how should we evaluate an individual's well-being? Our model suggests two possibilities, using the objectives of the deliberative system, or using the objectives of the deliberative system combined with any disutility from exerting willpower (thus incorporating the role of affect). We argue, however, that neither seems appropriate, and that 'true' well-being will lie somewhere in between these two extremes.

Finally, we conclude in Section VII by discussing a number of ways to expand upon the framework that we propose.

## II. A Two-System Model of Behavior

In this section, we develop our model of interactions between affect and deliberation, along with some motivations for our approach. Figure 1 provides an overview.

**Figure 1:**

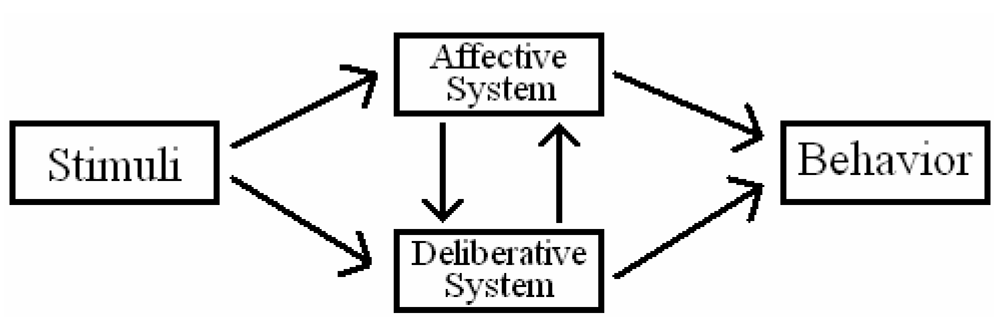



Figure 1 has four key features: (i) There are two qualitatively different systems, (ii) each system is potentially influenced by environmental stimuli, (iii) there are reciprocal interactions between the two systems, and (iv) behavior is the outcome of activity in both systems. We discuss each of these in turn.

**Two competing systems: affect and deliberation**

***Motivations for two systems***: Our central assumption is that there are two competing systems.[1] This perspective is of course not novel. Dual-process models of the human mind are ubiquitous in accounts of human behavior dating back to the ancient Greeks.[2] At an intermediate level of antiquity, for example, Adam Smith, in the Theory of Moral Sentiments, viewed human behavior as a struggle between the "passions" and an "impartial spectator" (see Ashraf, Camerer, & Loewenstein, 2005). The passions refer to immediate motivational forces and feeling states, such as hunger, thirst, anger, and sexual lust. The impartial spectator refers to the human ability to take a dispassionate view of one's own conduct — to evaluate one's own behavior as if through the eyes of another person who is unaffected by any passions of the moment. Smith viewed the ability to assume the perspective of an impartial spectator as a powerfully moderating force in human behavior, the source "of self-denial, of self-government, of that command of the passions which subjects all the movements of our nature to what our own dignity and honour, and the propriety of our own conduct, require" (1759:26). He recognized, however, that such perspective-taking has limits, and can be overcome by sufficiently intense passions:

> There are some situations which bear so hard upon human nature, that the greatest degree of self-government, which can belong to so imperfect a creature as man, is not able to stifle, altogether, the voice of human weakness, or reduce the violence of the passions to that pitch of moderation, in which the impartial spectator can entirely enter into them. (1759:29)

Dual-process models are also common in contemporary psychology (see Chaiken & Trope, 1999 for a broad introduction), and have also been applied to the domain of decision making. Most dual process models of decision making focus on two processes that differ (roughly) along one of two dimensions: (i) cognition vs. emotion (e.g., Epstein, 1994; Sloman, 1996; Slovic et al, 2002; Finucane, Peters, & Slovic, forthcoming), and (ii) controlled vs. automatic processes (e.g., Kahneman & Frederick, 2002). While these dimensions are clearly somewhat closely related, our model is motivated primarily by the former. In this group, the dual-process models that are closest in spirit to our own are Metcalfe & Mischel's hot/cool model (1999) and Fazio & Towles-Schwen's MODE model (1999).[3]

Metcalfe & Mischel (1999) distinguish between a “hot emotional system” that is simple, reactive, and fast, and a “cool cognitive system” that is complex, reflective, and slow (and “devoid of emotion”). The person’s behavior depends on which system is dominant at a particular moment. Metcalfe & Mischel use this model primarily to explain the diverse results obtained in studies based on Mischel’s delay-of-gratification paradigm (described in more detail below), and in particular how different strategies might be useful in helping the cool system to gain dominance (e.g., by obscuring a reward stimulus).

Fazio & Towles-Schwen’s MODE model (1999) distinguishes two types of attitude-to-behavior processes: spontaneous processing and deliberative processing. The former is automatic, fast, and does not require cognitive effort. Behavior is simply caused by accessible, and thereby automatically activated, attitudes relevant to the behavior. Deliberative processing, in contrast, is controlled, slow, and effortful. According to this view, implicit attitudes guide spontaneous behavior, whereas explicit attitudes are the bases for intentional actions (similar assumptions are made in the dual-attitude model proposed by Wilson, Lindsey, & Schooler, 2000).

Economists, too, have developed dual-process models of the human mind. In their seminal work, Shefrin and Thaler (Thaler & Shefrin 1981, Shefrin & Thaler 1988) propose a model in which a far-sighted “planner” interacts with a series of myopic “doers,” and they discuss the implications for intertemporal choice. More recently, Bernheim & Rangel (2003, 2004), Benhabib & Bisin (2005), and Fudenberg & Levine (2006) have all developed dual-process models. Like Shefrin and Thaler’s model, these models are narrowly developed to focus exclusively on intertemporal choice. In this paper, in contrast, we develop applications of our model not only to intertemporal choice, but also to choices between risky prospects or choices between social allocations. Moreover, these models are mostly silent on the determinants of the relative influence of the two systems, which is a major focus of ours.

In addition to being motivated by dual-process models of the human mind, we are also motivated by evidence from neuroscience on the functional specificity of different regions of the human brain. Neuroscientists view the human brain as a complex organ composed of many specialized systems, and view behavioral (and other) outcomes as determined by interactions between these systems. Indeed to many neuroscientists the notion of just two systems would seem naive. If one had to pick the most natural neurophysiologic division of the human brain, however, it would likely be that between the prefrontal cortex and evolutionarily older brain structures.

When human evolution departed from that of apes approximately 6 million years ago, our brains were not redesigned anew. Rather, new capabilities — perhaps most importantly, the ability to deliberate about the broader consequences of our actions — were gradually added to the underlying, more primitive brain systems. The evolutionarily older brain systems, which evolved to promote survival and reproduction, have changed little over the course of human evolution and continue to play the same role that they did for our predecessors and do for other mammals (MacLean, 1990). They incorporate motivational mechanisms — often operating outside of consciousness — designed to ensure that we eat when nutritionally deficient, take actions to maintain body temperature, have sex when the situation is propitious for reproduction, and so forth.

The seemingly unique human ability to focus on broader goals appears to rely heavily on the prefrontal cortex, which is also the region of the brain that expanded most dramatically in the course of human evolution (Manuck et al, 2003). As Massey (2002:15) comments:

> Emotionality clearly preceded rationality in evolutionary sequence, and as rationality developed it did not replace emotionality as a basis for human interaction. Rather, rational abilities were gradually *added* to preexisting and simultaneously developing emotional capacities. Indeed, the neural anatomy essential for full rationality—the prefrontal cortex—is a very recent evolutionary innovation, emerging only in the last 150,000 years of a 6-million-year existence, representing only about 2.5 percent of humanity's total time on earth.

Jonathan Cohen (2005:6) reaches a similar conclusion:

> "The very fact that economic theory, contrived by the human mind, can describe optimal behavior provides *prima facie* evidence that human beings can conceive of optimal behavior and, therefore, in principle are capable of it. There is good reason to believe that this capability— presumably dependent on higher cognitive

faculties such as reasoning, planning and problem solving—relies heavily on the function of a particular set of brain structures, including the prefrontal cortex... In other words, these structures may be a critical substrate for 'homo economicus.'"

The earliest, and perhaps still the best, evidence that the prefrontal cortex plays such a role comes from studies of people with damage to the prefrontal cortex (for an overview, see Damasio, 1994). In particular, patients with damage to the ventromedial prefrontal cortex exhibit impaired decision-making abilities. Such people often exhibit no overt limitations in their intellectual abilities, and are often quite able to predict and verbally describe the future consequences of different behaviors. However, they have trouble assessing the importance of — and their preferences about — those future consequences. Moreover, while some such patients do formulate plans (or take jobs), they usually lose focus and fail to implement those plans. Lhermitte (1986) found that, due to their inability to act on long-term goals, the behavior of patients with prefrontal lesions becomes largely a function of immediate contingencies of the environment, a pattern that he aptly describes as an "environmental dependency syndrome."[4]

***Motivations for Affect vs. Deliberation*:**

Our distinction between affect and deliberation is motivated by and roughly similar to many of the dichotomies described above — Adam Smith's passions vs. impartial spectator, Metcalfe & Mischel's hot emotional system vs. cool cognitive system, and the neurophysiologic division between the prefrontal cortex and evolutionarily older brain structures. Specifically, we assume that the deliberative system operates in a standard consequentialist fashion, choosing its desired course of action based on an assessment of likely consequences. Indeed, in our applications, we often assume as a natural starting point that the operation of the deliberative system is well described by the standard rational-choice model.[5] Thus, the more novel feature of our model is the affective system, which merits further discussion.

Our use of the term affect differs from many lay definitions, which tend to focus on the subjective feeling states associated with emotions. In our usage, the defining characteristic is that affects carry "action tendencies" (Frijda, 1986) — e.g., anger motivates us to aggress, pain to take steps to ease the pain, and fear to escape (or in some cases to freeze). This perspective is consistent with accounts from evolutionary psychologists (Cosmides & Tooby, 2004) wherein affects are 'superordinate programs' that orchestrate responses to recurrent situations of adaptive significance in our evolutionary past. In addition, affect, as we use the term, embodies not only standard emotions such as anger, fear, and jealousy, but also drive states such as hunger, thirst, and sexual desire, and motivational states such as physical pain, discomfort (e.g., nausea), and drug craving.[6] In contrast to lay definitions, then, affects can and typically do involve a wide array of physiological and psychological changes beyond subjective feeling states — fear, for example, produces shifts in perception and attention, changes in goals, physiological effects, and effects on memory. Moreover, many affective processes occur below the threshold of conscious awareness, and hence are not experienced as an "emotion" or "affect" in the lay sense (LeDoux, 1996; Winkielman & Berridge, 2004).

Our use of the term affect is related to the distinction between expected emotions and immediate emotions (Loewenstein et al, 2001; Loewenstein & Lerner, 2003; Rick & Loewenstein, forthcoming). Expected emotions are emotions that are anticipated to occur in the future (perhaps even the very near future) as a result of decisions. As such, expected emotions are part of the likely consequences of decisions and therefore incorporated into deliberation. Indeed, one interpretation of the standard consequentialist model of decision making is that people seek to create positive expected emotions and avoid negative expected emotions. Immediate emotions, in contrast, are experienced at the moment of decision, and might be completely unrelated to the decision at hand, in which case they are referred to as "incidental" (Bodenhausen, 1993). Perhaps most importantly, although they are experienced while making a decision, they are not affected by the choice that is made, and thus, under the usual rational-choice perspective, should be irrelevant to choices. But numerous studies have found that immediate emotions do influence decision making (Lerner & Keltner, 2000; Raghunathan & Pham, 1999; Ariely & Loewenstein, 2006; Wilson & Daly, 2003). A natural interpretation of our affective system is that it captures the influence of immediate emotions.

Our distinction between affect and deliberation is also similar to a distinction that Kent Berridge (1995) draws between "wanting" and "liking." Wanting refers to an immediate motivation to acquire something or engage in some activity. Liking, in contrast, refers to how much one actually ends up enjoying the good or activity. Under this interpretation, our affective system makes decisions based on wanting, whereas our deliberative system makes decisions based on liking. Berridge indeed finds that wanting and liking are mediated by different, albeit overlapping, neural systems.[7]

***Our Formalization:*** To formalize the existence of two systems, we assume that there are two "objective functions" operating simultaneously. Specifically, consider an individual who must choose an option *x* out of some choice set *X*. On one hand, **t**he affective system is motivated to engage in certain behaviors, and we capture these motivations with a motivational function *M(x,a)*. (The variable ***a*** captures the intensity of affective motivations, as we discuss below.) If the affective system alone were completely in charge of behavior, the affective system would "choose" $x^A \equiv \arg\max_{x \in X} M(x,a)$, which we refer to as the *affective optimum*. On the other hand, the deliberative system evaluates behavior with a broader and more goal-oriented perspective, and we capture the desirability of actions as perceived by the deliberative system with a utility function *U(x)*. If the deliberative system alone were completely in charge of behavior, the deliberative system would choose $x^D \equiv \arg\max_{x \in X} U(x)$, which we refer to as the *deliberative optimum*.

Notice that, as yet, we have assumed very little about the objectives of the two systems. In Sections III, IV, and V, we make specific assumptions about these objectives in order to derive specific predictions. Importantly, these assumptions will depend on the specific domain of behavior under consideration. Our goal is to begin with a general framework that can be applied quite broadly, then to show in subsequent sections how this general framework can be applied to almost any type of decision.

**Environmental Stimuli**

***Motivations***: A second key feature of Figure 1 is that each system is potentially influenced by environmental stimuli. In some — indeed probably most — cases, the two systems will respond to the same stimuli with similar motivational tendencies. For example, during a break at a conference, the availability of a snack might create a surge of hunger in the affective system and be perceived by the deliberative system as a welcome opportunity to recharge before the next session. However, because the two systems operate according to quite different principles, in other situations the same stimulus can influence the two systems differently. For example, if you are on a diet, the snack might still produce a surge of hunger, but also serve as a reminder that you are attempting to lose weight and had resolved not to snack.  The dual-process perspective is most useful for analyzing the latter type of situation.

Although both systems can be influenced by environmental stimuli, the effect is more pronounced for the affective system.  The deliberative system assesses the likely consequences of different actions with a broad, goal-oriented perspective.  While, as in the examples above, environmental stimuli might serve to remind the deliberative system of its current goals, such effects are likely small and, more importantly, relatively invariant to the exact nature of the stimuli.  The affective system, in contrast, is directly triggered by environmental stimuli, and hence is quite sensitive to the nature of those stimuli.

Indeed, evidence suggests a number of factors that influence the strength of affective motivations while affecting the goals of the deliberative system much less if at all.  Perhaps most important is the *temporal proximity* of reward and cost stimuli.  The affective system is highly responsive to temporal proximity; affective motivations are intense when rewards and punishments are immediate but much less intense when they are temporally remote. Deliberation is, in contrast, much less sensitive to immediacy.[8]  The importance of immediacy has been documented in countless studies employing diverse methods.  In one of the most recent and definitive studies, Berns and collaborators (Berns et al, 2006) scanned the brains of subjects as they were waiting to receive electric shocks of different intensities.  They found activation of several regions that are known to respond to the experience of pain when subjects were waiting to experience it.  Moreover, they found that activation of these regions increased dramatically as the shock approached in time.

In addition to temporal proximity, various forms of *non-temporal proximity* have similar effects (Lewin 1951; Trope & Liberman 2003).  Thus, for example, a tempting snack is more likely to evoke hunger to the extent that it is nearby, visible, or being consumed by someone else. Likewise, a person who makes you angry is more likely to evoke anger to the extent that he is geographically close (or likely to be soon) or visible.  At the same time, and similar to the pattern for temporal proximity, these features have much less effect on deliberation than they have on affect.

Early, and still compelling, evidence on the role of non-temporal proximity comes from a series of classic studies conducted by Walter Mischel and colleagues (see for instance Mischel, Ebbesen, & Zeiss, 1972; Mischel, Shoda, & Rodriguez, 1989; and Mischel, Ayduk, & Mendoza-

Denton, 2003). Young subjects (ages 4-13) were instructed by an experimenter that they could have a small snack immediately or a larger snack if they waited for the experimenter to return. The experimenter then measured how long the subjects were willing and able to wait for the larger snack (with a cap of fifteen minutes). In a baseline treatment, children had the larger delayed snack positioned in front of them as they waited for the experimenter. Relative to this baseline treatment, subjects were able to delay significantly longer when the larger snack was not present, or even when the larger snack was present but covered. Similarly, even when the larger snack is present and uncovered, subjects were able to delay longer if their attention was diverted away from the snack so that they spent less time looking at the larger snack.

A third important factor is the pattern of stimulus change over time. In particular, the affective system is highly attuned to changes in circumstances, but insensitive to levels of things; it adapts to ongoing or repeated stimuli (Frederick & Loewenstein, 1999). There are good reasons for the affective system to adapt to ongoing stimuli (e.g., Becker & Rayo, forthcoming). Affective processes serve important motivational and regulatory functions. Negative affective states alert us that something is wrong and motivate us to change. Positive affective states provide immediate reward when we rectify whatever is wrong. Thus, for example, when our body temperature drops, blood flow is progressively reduced to peripheral regions of the body, heart rate tends to decrease, and our brains instigate emotional programs designed to motivate us to seek out warmth. We experience discomfort, and almost anything that raises our body temperature becomes pleasurable — a process called “alliesthesia” (Cabanac, 1971). When an adverse situation remains unchanged for a prolonged period, however, our affective system takes its constancy as a cue that nothing can be done about the situation, and that it is not worth devoting further attention or motivation to it.

A fourth factor is the *vividness* of stimuli, by which we mean the ability to conjure the experience in mind. Researchers who study the impact of “incidental emotions” (Bodenhaus, 1993) have become increasingly expert at evoking emotion, and many of the manipulations play on vividness by, for instance, showing people movies of an emotion-evoking event (Lerner, Small, & Loewenstein, 2004), having people write essays in which they imagine themselves in a situation (Lerner & Keltner, 2000), playing music (Halberstadt & Niedenthal, 1997; Blood & Zatorre, 2001), or even through the artful use of odors (Ditto, Epstein, & Pizarro, 2004, Zald & Pardo, 1997).

***Our Formalization***: To incorporate such effects into our model, we introduce the variable *a* that captures the *intensity* of affective motivation. Formally, recall that affective motivations are captured by the motivational function *M(x,a)*. We further assume that the larger is *a*, the stronger will be the affective motivations. Of course, we must clarify what we mean by "stronger affective motivation." In general, there are two implications of having a stronger affective motivation. First, it might change the affective optimum $x^A$ — weak hunger motivates you to eat a modest amount, whereas strong hunger motivates you to eat anything in sight. More importantly, a stronger affective motivation changes the intensity of a person's motivation to move toward the affective optimum. These effects are a bit difficult to quantify for the general case because we have not specified the nature of the possible choices. But for many cases, it is possible to define the options such that *x* represents the extent to which a person engages in some behavior and affective intensity tends to increase the motivation to engage in *x* (e.g., in the realm of intertemporal choice, *x* might be the amount a person consumes now). In such cases, we can formally assume that the larger is *a*, the larger is $\partial M/\partial x$ (for all *x*).

Affective intensity will depend on two factors. First, it will be larger to the extent that environmental stimuli are tied to the behavior under consideration. Specific stimuli create motivations for specific behaviors: hunger creates a motivation to eat, sexual arousal creates a motivation for sex, and sympathy creates a motivation to help others. Second, and more importantly, affective intensity will depend on the factors discussed above: temporal proximity, non-temporal proximity, novelty/adaptation, and vividness.

Note that our formal analysis does not permit a corresponding influence of stimuli on deliberation. In principle, it would be straightforward to introduce such effects into our model by modifying the deliberative system's utility function — e.g., we might use *U(x,c)* where *c* represents cognitive states. However, because we do not explicitly focus on such considerations in this paper, for the sake of simplicity they are not incorporated into the mathematical model.

**Reciprocal interactions between affect and deliberation.**

***Motivations:*** A third feature of Figure 1 is that there are reciprocal interactions between the two systems. One reason to believe in the existence of such interactions is purely physiological: there are neural connections running from the evolutionarily older brain systems to the prefrontal cortex and also in the reverse direction. But other evidence also suggests such reciprocal interactions.

Perhaps most importantly, there is considerable evidence that emotional input from the affective system is required for sound deliberative thinking, because without this input the deliberative system has a difficult time assessing the value of future consequences. For example, Damasio's "somatic marker hypothesis" posits that normal decision making is guided by somatic reactions to deliberations about alternatives that provide information about their relative desirability. In support of this perspective, Damasio and colleagues (Damasio, 1994; Bechara et al, 1997) show that certain neurological abnormalities that block such somatic reactions but produce minimal cognitive deficits lead to significant impairments of risky decision making.[9] Other research, by Wilson and colleagues (e.g., Wilson & Schooler, 1991; Wilson et al, 1993) shows that the quality of decision making suffers when affective inputs are suppressed by having decision makers think systematically about the pros and cons of a decision.[10]

At the same time, there is also evidence that affective motivations and drives can undermine sound deliberative thinking. Research on "motivational" biases on judgment documents the diverse ways in which affect can bias cognitive deliberations. Thus, when people are powerfully motivated to believe something, they can usually find ways to do so (e.g. Kunda, 1990; Forgas, 1995). Similarly, research on "mood-congruent memory" (Bower, 1981; Wright & Bower, 1992) demonstrates that emotions (affect) can also bias what we think about and how we think about it. Indeed, in some situations, it seems that affect can virtually short-circuit sound deliberative thinking (e.g., Baumeister, Heatherton, & Tice, 1994; Loewenstein, 1996). As Adam Smith (1789:221) commented,

> When we are about to act, the eagerness of passion will seldom allow us to consider what we are doing with the candour of an indifferent person. The violent emotions which at that time agitate us, discolour our views of things, even when we are endeavouring to place ourselves in the situation of another, and to regard the objects that interest us in the light in which they will naturally appear to him.

There are also reasons to believe that the deliberative system can influence activity in the affective system. For instance, while the sight of a snake might trigger a visceral motivation to flee, the deliberative system can sometimes override this motivation – e.g., if the snake is behind glass, so the individual recognizes that there are objectively no adverse consequences to approach, the person is likely to suppress the impulse to flee.

***Our Formalization:*** In fact, our formal analysis will not focus much on these reciprocal interactions. In part, we do not focus on such interactions because they are not essential for any of the applications. But in part we do not model such interactions because it is a bit difficult to distinguish them from the question of which system has a stronger influence on behavior (which we address next). For instance, when the affective system takes dominant control of behavior, one interpretation is that the affective system is influencing the activity in the deliberative system, but another interpretation is that the affective system currently has a particularly strong direct influence over behavior. Mathematically, both interpretations are consistent with our mathematical formulation.

**Behavior is Influenced by Both Systems**

***Motivations***: The final key feature of Figure 1 is that behavior is the outcome of activity in both systems. The key question, then, is what determines their balance of power.

A variety of evidence suggests that the affective system holds a kind of primacy in determining behavior — that is, the affective system has default control of behavior, but the deliberative system can step in to exert its influence as well. One bit of evidence is anecdotal: people often seem to operate in an effectively preprogrammed fashion that is only interrupted by deliberation. As Adam Smith argued early on, if the deliberative system does not get activated — if it does not attend to a particular choice situation — then behavior will be driven entirely by affective motivations. (Anyone who has ever been put in front of a table of snacks and who has found himself eating without having engaged in deliberation can appreciate this notion.)

Better evidence comes from research demonstrating that affective reactions tend to occur first, temporally, with deliberations typically playing a secondary, corrective, role. For instance, Joseph LeDoux and his colleagues (summarized in LeDoux, 1996) have demonstrated that both the cortex and the lower brain structures play a role in fear responses in rats. Specifically, fear responses are influenced by two separate neural pathways from the sensory thalamus (a lower-brain structure that performs crude processing of external stimuli) to the amygdala (another lower-brain structure that plays a critical role in fear responses). One pathway goes directly from the sensory thalamus to the amygdala, and the second goes first from the sensory thalamus to the neocortex and from there to the amygdala. Moreover, they also discovered that the direct pathway is about twice as fast as the indirect pathway. As a result, rats can have an affective reaction to a stimulus before their cortex has had the chance to perform more refined processing of the stimulus. Such immediate affective responses provide organisms with a fast but crude assessment of the behavioral options they face which makes it possible to take rapid action. To use LeDoux's example, it is useful to have an immediate defensive reaction to a curved object rather than wait for the cortex to decide whether that object is a coiled snake or a curved stick.[11]

A similar pattern can be seen in humans. In a series of seminal papers with titles such as "Feeling and Thinking: Preferences Need No Inferences" (1980), and "On the Primacy of Affect" (1984), Zajonc presented the results of studies which showed that people can often identify their affective reaction to something — whether they like it or not — more rapidly than they can even say what it is, and that their memory for affective reactions can be dissociated from their memory for details of a situation, with the former often being better. People often remember whether we liked or disliked a particular person, book, or movie, without being able to remember any details other than our affective reaction (Bargh, 1984). Similarly, Gilbert & Gill (2000) propose that people are "momentary realists" who initially trust their emotional reactions and only correct them through a comparatively laborious and time-consuming cognitive process. Thus, if the car behind you honks after the light turns green, you are likely to respond with immediate anger, followed, perhaps, by the recognition that if it had been you stuck at a green light behind a driver who was eating a sandwich while talking on the cell phone, you might have reacted similarly. As Adam Smith (1759:136) expressed it, "We are angry, for a moment, even at the stone that hurts us. A child beats it, a dog barks at it, a choleric man is apt to curse it. The least reflection, indeed, corrects this sentiment, and we soon become sensible, that what has no feeling is a very improper object of revenge."

When deliberation gets involved, what determines the extent to which it influences behavior? There is, in fact, compelling evidence that deliberation does not easily take full control. Rather, when in conflict with affect, deliberative control, to the extent that it is possible at all, requires an expenditure of effort. The most important evidence along these lines comes from research by Baumeister and colleagues on willpower (for a summary see Baumeister & Vohs, 2003), by which they mean an inner exertion of effort required to implement some desired behavior. Their basic contention is that such willpower is a resource in limited supply (at least in the short run), and that depletion of this resource by recent use will reduce a person's ability to implement desired behaviors. To demonstrate this point, Baumeister's basic willpower paradigm involves having subjects carry out two successive, unrelated tasks that both (arguably) require willpower, and comparing the behavior on the second task to a control group which had not performed the first task. The general finding is that exerting willpower in one situation tends to undermine people's propensity to use it in a subsequent situation. Thus, in one study, subjects who sat in front of a bowl of cookies without partaking subsequently gave up trying to solve a difficult problem more quickly than did subjects who were not first tempted by the cookies.

Because the target behaviors in Baumeister's studies — e.g., not eating cookies or trying to solve a difficult puzzle — typically involve pursuit of broader goals whereas not doing these behaviors typically involves indulging affective motivations, we believe there is a natural interpretation of these results for our model. Specifically, it is attempts by the deliberative system to override affective motivations that require an inner exertion of effort or willpower. Hence, if a person's willpower is depleted by recent use, the deliberative system will have less influence over behavior. Consistent with our view, a related line of research (summarized in Baumeister & Vohs, 2003) shows that simply making decisions can undermine willpower. In this research, some subjects were asked to make a long series of choices between products while other subjects were simply asked to report on their usage of the same products. Afterward, in an ostensibly new study administered by a new experimenter, they were asked to consume a bad-tasting beverage. Subjects who had made choices drank a significantly smaller amount of the beverage than did those in the control group.

Hence, one situation in which affect will have more sway over behavior is when the deliberative system is "worn out" from past willpower use. A second, related, situation is when the deliberative system is "distracted" by unrelated cognitive tasks. Research has shown that having subjects perform simple cognitive tasks — an intervention labeled "cognitive load" — undermines efforts at self-control. In one innovative study, Shiv & Fedorikhin (1999) had some subjects memorize a 7-digit number (high cognitive load) and others memorize a 2-digit number (low cognitive load). Subjects in both groups were instructed to walk to another room in the building where they were to report the number they had memorized. On their way, they encountered a table at which they were presented with a choice between a highly caloric slice of cake and a bowl of fruit-salad. The researchers predicted that high cognitive load would undermine self-control leading to choice of the cake, and this is what they found; 59% chose the cake in the high-load condition, but only 37% in the low-load condition.

Although we will not incorporate it into our formal analysis, we briefly mention here a third variable which seems to undermine self-control: stress. Several studies have shown, for example, that stress often leads to relapse by abstinent addicts. In one of the most carefully crafted studies of this type, Shiffman & Waters (2004) had smokers who had quit carry palm pilots around which beeped at random intervals, then asked them questions. They were also instructed to enter information into the palm pilot if they smoked a cigarette. One of the most important findings from this study, which reinforces findings from numerous other studies employing different methods, was that relapse was often immediately preceded by stressful events. Similar findings have been obtained with people who are attempting to diet --- for instance, they increase food intake after receiving feedback that they had failed at an easy task and when anticipating that they would have to give a speech in front of an audience (Heatherton, Herman, & Polivy, 1991; see, also, Polivy, Herman, & McFarlane, 1994; Polivy & Herman, 1999)

***Our Formalization***: To formalize these ideas, we assume that the deliberative system makes the final choice, but it must make this choice subject to having to exert cognitive effort — or willpower — to control affective motivations. We capture this cognitive effort by assuming that, to induce some behavior $x \neq x^A$, the deliberative system must exert an effort cost, in utility units, of $h(W,\sigma)*[M(x^A,a)–M(x,a)]$. This formulation assumes that the further the deliberative system moves behavior away from the affective optimum — that is, the larger is $[M(x^A,a)–M(x,a)]$ — the more willpower is required. The factor $h(W,\sigma)$ represents the cost to the deliberative system of mobilizing willpower — i.e., the higher is $h(W,\sigma)$, the larger is the cognitive effort required to induce a given deviation from the affective optimum.[12]

Based on our discussion above, we incorporate two factors that make it more costly for the deliberative system to exert willpower. The first, and perhaps more important, is the person's current *willpower strength*, which we denote by *W*. This variable is meant to capture the current stock of willpower reserves; we assume that *h* is decreasing in *W*, so that, as one's willpower strength is depleted, the deliberative system finds it more difficult (more costly) to influence the affective system. Our analysis in this paper will focus on one particular implication with regard to willpower strength: the more willpower a person has used in the recent past, the more her current willpower strength will be depleted, and hence exerting willpower becomes more costly.[13]

Although none of the results we present in the paper depend on these specifics, there is a natural way to capture dynamic willpower effects. Let $w_t \equiv M(x_t^A, a_t) – M(x_t, a_t)$ denote the amount of willpower exerted in period *t*, and then let $W_{t+1} = f(w_t, W_t)$ where *f* is decreasing in $w_t$ and increasing in $W_t$. In words, willpower strength in period *t+1* is smaller if more willpower was used in period *t*, but larger if the person had more willpower strength in period *t*. It also seems natural to assume that willpower is replenished over time — e.g., that there is a replenishment rate $r > 0$ such that $f(w,W) > W$ when $w < r$ and $f(w,W) < W$ when $w > r$ — and that there is an upper bound $\overline{W}$ on the stock of the resource — $f(0,W) < \overline{W}$ for all *W*.[14]

The second factor that makes it more costly for the deliberative system to exert willpower is cognitive load, which we denote by $\sigma$. Our analysis in this paper will also focus on one particular implication with regard to cognitive load: if a person's deliberative system is distracted by unrelated cognitive tasks, exerting willpower becomes more costly.

**General Implications**:

We now put together the pieces of our formalization to derive the general implications of our model. To make the final choice, the deliberative system trades off the desirability of actions — as reflected by its utility function $U(x)$ — against the willpower effort required to implement actions. Hence, the deliberative system will choose the action $x \in X$ that maximizes:

$$V(x) \quad \equiv \quad U(x) \quad - \quad h(W,\sigma) * \left[ M(x^A, a) - M(x,a) \right]$$

Because the affective optimum $x^A$ is not affected by the person's actual choice, choosing *x* to maximize $V(x)$ boils down choosing *x* to maximize $U(x) + h(W,\sigma)M(x,a)$.[15] It follows that the deliberative system will choose an option that is somewhere in between the deliberative optimum and the affective optimum (either $x^D \geq x \geq x^A$ or $x^A \geq x \geq x^D$). Exactly where behavior falls will depend on the cost of mobilizing willpower as captured by $h(W,\sigma)$. As the cost of willpower becomes smaller, behavior will be closer to the deliberative optimum. In the limit as the cost of willpower approaches zero, behavior will converge to the deliberative optimum (the deliberative system will be in complete control). Analogously, as the cost of willpower becomes larger, behavior will be closer to the affective optimum. In the limit as the cost of willpower gets very large, behavior will converge to the affective optimum (the affective system will be in complete control).

Although we interpret our model as reflecting that the deliberative system is choosing behavior subject to willpower costs, there is a second interpretation of our model that is perhaps conceptually more consistent with Figure 1. Because the deliberative optimum $x^D$, like the affective optimum $x^A$, is not affected by the person's actual choice, choosing $x$ to maximize $V(x)$ is equivalent to choosing $x$ to *minimize*:

$$\left[U(x^D) - U(x)\right] \quad + \quad h(W,\sigma) * \left[M(x^A, a) - M(x,a)\right]$$

Hence, our model is equivalent to thinking of behavior as coming from the minimization of a weighted sum of two costs: a cost to the deliberative system from not getting its optimum $x^D$, and a cost to the affective system from not getting its optimum $x^A$. In this interpretation, $h(W,\sigma)$ captures the relative weights of the two systems.

Although the two interpretations are formally equivalent, we prefer the former. At some level, the deliberative system is making a "decision" in terms of how much willpower to exert, while the affective system is more reflexive. Indeed, the prefrontal cortex is sometimes referred to as performing an "executive function" (Shallice & Burgess, 1998); much as a chief executive needs to work through the existing structure and culture of the firm to implement her plans, the deliberative system has to work through the affective system to influence behavior. Hence, we shall adopt the first perspective in the remainder of the paper.

While we have motivated our model as a dual-system approach, in the end, behavior is determined by a single "objective" function $V(x)$. What is the value, then, of the dual-system approach? One way in which the dual-system approach is useful is that it provides a natural interpretation of many behavioral outcomes. There are many situations in which people are "of two minds" — they cognitively believe they ought to be doing one thing, but then end up doing another. When evaluating risky prospects, people might cognitively believe that they should weight probabilities linearly, but then makes choices that reflect an insensitivity to probabilities; and when weighing some intertemporal indulgence such as a tasty but highly caloric morsel or a willing but forbidden sexual partner, people might cognitively think that the indulgence is not worth the future costs, but then indulge nonetheless. Our model provides a natural interpretation: people's cognitive beliefs for what they ought to do reflect only the objectives of the deliberative system, whereas actual behavior is influenced by affective motivations as well. In other words, many deviations from the standard "rational-choice" paradigm can be interpreted as coming from the motivations of the affective system.

A second way in which the dual-system approach is useful is that it provides a template for how to interpret research from neuroscience. Recent research in neuroscience, particularly in the subdiscipline of neuroeconomics, often focuses on where we see brain activity when people make decisions. And, while neuroscientists are often interested in more fine partitions, a frequent focus is on the extent to which activity occurs in the prefrontal cortex or in evolutionarily older brain systems. To the extent that our deliberative system is roughly meant to capture activity in the prefrontal cortex whereas our affective system is roughly meant to capture activity in the evolutionarily older brain systems, according to our model such research can be used to shed insight on the different objectives of the two systems. Indeed, we have already discussed some neuroscientific research in this way, and do so further in the discussion of specific applications.

But perhaps the most important value of the dual-system approach is that it generates testable implications:

Implication #1: Willpower depletion (decreased $W$) or cognitive load (increased $\sigma$) increases the cost of exerting willpower (increase $h(W,\sigma)$), which reduces the influence of the deliberative system and therefore shifts behavior further from the deliberative optimum.

Implication #2: If affective intensity (increased $a$) increases the conflict between the two systems (shifts $x^A$ further from $x^D$), then it will shift behavior further from the deliberative optimum; in contrast, if it decreases the conflict between the two systems (shifts $x^A$ closer to $x^D$), then it will shift behavior closer to the deliberative optimum.

Notice that, whereas the effects of willpower depletion and cognitive load are straightforward, the effects of affective intensity are more nuanced because it might increase or decrease the conflict between the systems. In many applications, it will be natural to assume the former --- that affective motivations diverge from deliberative goals and the stronger the intensity of affective motivations the larger the divergence. In some applications, however, it is more natural to assume that at low levels affective intensity further aligns the two systems while at higher levels it creates a divergence. Indeed, our application to social preferences will have this feature.

In the next three sections, we apply our model to three specific domains: intertemporal choice, risky decision making, and social preferences. In each domain, we make specific assumptions about the objectives of the two systems and use these to derive specific testable predictions of our model. In some cases we find existing evidence that supports these predictions, but in others we propose them as testable, but as yet untested, implications of the model.

## III. Intertemporal Choice

In this section, we describe the most straightforward application of our model, to intertemporal choices — decisions that involve tradeoffs between current and future costs and benefits. People are often of two minds when it comes to intertemporal choice; they are powerfully motivated to take myopic actions, such as eating highly caloric foods, imbibing addictive drugs, eschewing contraception, "flaming" on email, and so on, while recognizing simultaneously that these activities are not in their self-interest. As Adam Smith (1759:227) wrote, seemingly referring to an act of sexual misconduct,

> At the very time of acting, at the moment in which passion mounts the highest, he hesitates and trembles at the thought of what he is about to do: he is secretly conscious to himself that he is breaking through those measures of conduct which, in all his cool hours, he had resolved never to infringe, which he had never seen infringed by others without the highest disapprobation, and the infringement of which, his own mind forebodes, must soon render him the object of the same disagreeable sentiments.

In order to apply our model, we must make assumptions about what drives the two systems --- that is, about the two objective functions. In the realm of time preference, there is a natural starting point: The affective system is driven primarily by short-term payoffs, whereas the deliberative system cares about both short-term and longer-term payoffs. In this section, we investigate the implications of our model given these assumptions.

There is direct evidence in support of this perspective. On the deliberative side, Frederick (2003) asks subjects how they believe they should respond to outcomes occurring at different times, and most people generally believe that time discounting is not normatively justified --- that outcomes should receive the same weight regardless of when they occur. To the extent that people's beliefs for how they ought to behave reflect the goals of the deliberative system, this evidence suggests that the deliberative system takes a far-sighted perspective.

On the affective side, when animals are presented with intertemporal choices, they are extremely myopic. There is a long literature that demonstrates extreme myopia in pigeons and rats. Indeed, even some species of monkeys tend to exhibit extreme myopia --- for example, Stevens, Hallinan, & Hauser (2005) find that two species of New World monkeys, cotton-top tamarins and common marmosets, are willing to wait only 8 seconds and 14 seconds, respectively, for a food reward that is three times as large. Monkeys that are closer, evolutionarily, to humans show less myopia --- for example, Tobin et al (1996) present similar trade-offs to two cynomolgus monkeys, and they find that one is able to wait 32 seconds while the other prefers to wait for the larger food reward even at a delay of 46 seconds, the maximum they allow. Even so, to the extent that animal behavior can be used to shed insight on the motivations of humans' affective system, this evidence suggests that the affective system is myopic, and that concern for longer-term rewards and costs come from the deliberative system.

More convincing direct evidence comes from neuroscience. As we discuss in Section II, there is evidence that affective regions of the brain are sensitive to the temporal proximity of rewards and costs whereas deliberative regions are not. Even more direct support comes from a study by McClure et al (2004). Subjects' brains were scanned using fMRI while they made choices between smaller-sooner rewards vs. larger-later rewards. Some of the choices were between money amounts that would be received that day vs. larger amounts that would be received in two or four weeks. Others were between two money amounts that would both be delayed --- e.g., \$5 in two weeks vs. \$7 in four weeks.[16] All of these intertemporal choices produced activation in prefrontal regions associated with calculation and deliberation (relative to a baseline resting state). However, when one of the options involved an immediate reward, brain regions associated with affective processing also became activated. Moreover, in situations where an immediate reward was one of the options, the relative activation of the two regions was a significant predictor of choice. In particular, the larger the relative activation of the affective regions, the more likely the subject was to choose the immediate reward.

Consider a simple model of intertemporal choice that incorporates these assumptions about the objectives of the two systems. Suppose that each option $x$ in the choice set $X$ is associated with a stream of payoffs $x_1, x_2, \ldots, x_T$, where payoff $x_t$ is received in period $t$. The myopic affective system cares only about the immediate payoff, and so the affective system's motivational function is $M(x,a) = ax_1$. This formulation incorporates that having more intense affective motivations --- for any of the reasons discussed in Section II --- will lead to a stronger affective motivation.[17] The more far-sighted deliberative system values both immediate and future payoffs. As a natural starting point, we assume that deliberative system has standard exponential discounting, and so its utility function is $U(x) = x_1 + \delta x_2 + \ldots + \delta^T x_T$.

With this formulation, the affective optimum $x^A$ will be the option that yields the largest immediate payoff, which we'll denote by $x_1^A$. According to our model, the person will choose the option $x$ that maximizes:

$$V(x) = [x_1 + \delta x_2 + \ldots + \delta^T x_T] - h(W,\sigma)*[ax_1^A - ax_1].$$

Since $x_1^A$ is exogenous to the person's choice, this is equivalent to choosing the $x$ that maximizes:

$$\tilde{V}(x) = x_1 + [1/(1+h(W,\sigma)a)][\delta x_2 + \ldots + \delta^T x_T]$$

$$= x_1 + \beta\delta x_2 + \ldots + \beta\delta^T x_T,$$

where $\beta = 1/(1+h(W,\sigma)a) < 1$. Hence, our model provides a natural interpretation --- or re-interpretation --- of hyperbolic discounting. Specifically, even if the deliberative system discounts exponentially, because behavior is also influenced by a more myopic affective system, people will be more impatient when facing now vs. near-future trade-offs than they will be when facing future vs. further-future trade-offs --- which is the essence of hyperbolic discounting.[18]

Note that our model does not generate the literal hyperbolic functional form that often appears in the psychology literature. Rather, it generates the simplified “quasi-hyperbolic” functional form that has been used quite broadly in the economics literature (e.g., Laibson, 1997; O’Donoghue & Rabin, 1999). Even so, this form is consistent with much of the evidence for hyperbolic discounting. It is consistent with intertemporal preference reversals (e.g., Ainslie, 1975; Kirby, 1997). For instance, according to our model, a person might choose \$100 now over \$120 next week but also choose \$120 in six weeks over \$100 in five weeks. In particular, the person will choose \$100 now over \$120 next week if $100 > \beta\delta 110$; the person will choose \$120 in six weeks over \$100 in five weeks if $\beta\delta^6 120 > \beta\delta^5 100$ or $\delta 120 > 100$; and given that $\beta < 1$, it is easy for both conditions to hold. In addition, this form is consistent with declining (average) discount rates (e.g., Thaler, 1981; Benzion et al, 1989). Finally, it is consistent with recent evidence (e.g., Frederick, Loewenstein, & O’Donoghue, 2002) suggesting that most of the action in discounting revolves around the distinction between now and the future --- and in particular, that people exhibit nearly constant discounting when facing two future trade-offs.

Beyond providing an alternative account of hyperbolic time discounting, our model also generates testable predictions by applying the two general implications from Section II:

Implication #1: Willpower depletion or cognitive load will lead to more myopic behavior when people face trade-offs between immediate payoffs and future payoffs, but will not have much effect when people face trade-offs between payoffs that are all in the future.

Implication #2: Any factor that increases the intensity of the affective motivation for the immediate payoff will lead to more myopic behavior.

There is some existing evidence on these predictions. To our knowledge there is no direct evidence on the effects of willpower depletion on how people choose between a smaller-sooner reward and a larger-later reward. Even so, some studies in the literature on willpower depletion investigate how willpower depletion affects choices that arguably involve making intertemporal trade-offs. For instance, Vohs & Heatherton (2000) investigate how willpower depletion affects the amount of ice cream people eat when asked to taste and rate three flavors. To the extent that eating ice cream involves immediate benefits and future costs, eating more ice cream can be taken to reflect increased myopia. In support of Implication #1, they find that, among dieters, willpower depletion leads subjects to eat more ice cream. However, they find no effect among non-dieters. In addition, Vohs & Faber (2007) find that willpower depletion leads to increased impulse buying. Of course, impulse buying, per se, need have nothing to do with making intertemporal trade-offs --- a person can be prone to make hasty (and bad) buying decisions without making any trade-offs between immediate and future payoffs. In one of their experiments, however, the goods available for purchase could be classified as healthy food vs. unhealthy food, and so the mix of products purchased might reflect intertemporal trade-offs --- e.g., buying more unhealthy goods might reflect increased myopia. In fact, they did not find a significant relationship between willpower depletion and the mix of products purchased (although the sign of the relationship is consistent with Implication #1).

The Shiv & Fedorikhin (1999) study reported in Section II provides indirect support for the cognitive-load prediction --- specifically, cognitive load makes subjects more prone to choose cake over fruit, and choosing cake over fruit seems to reflect increased myopia. Benjamin, Brown, & Shapiro (2006) provide more direct evidence. They ask Chilean high school juniors to make a series of short-term trade-offs (now vs. next week) and long-term trade-offs (four weeks vs. five weeks) for monetary payoffs, and they classify people as patient if they behave in a way consistent with wealth maximization. Relative to control subjects, subjects who answer these questions while under cognitive load (holding a seven-digit number in memory) showed non-trivial reductions in short-term patience, although the difference was not statistically significant. In contrast, cognitive load had no effect on long-term patience.

Implication #2 generates a host of predictions based on the different factors that we discuss in Section II that increase affective intensity. Most straightforwardly, our model predicts that non-temporal proximity of immediate outcomes should play a large role in elicited discount rates. Thus, for example, the extent that an immediate reward can be seen or smelled (assuming that the appearance and smell are attractive) will affect the magnitude of discount rates that people’s behavior reveals, which is consistent with the research by Mischel and colleagues described in Section II. In addition, our model predicts that people who have particularly strong affective reactions to stimuli will exhibit more myopic behavior. In fact, direct support for this prediction comes from research by Hariri et al (2006), who find that people who exhibit larger affective reactions to random monetary gains and losses at one experimental session (as measured by neural activation in the ventral striatum) also show increased myopia when trading off immediate vs. future monetary payoffs at a different experimental session (that took place 10 to 63 weeks earlier). More generally, this prediction of our model helps to explain why impulsive actions are so often associated with strong emotions (e.g., road rage) or drives (e.g., impulsive eating). Finally, our model predicts stimulus-specific discounting. The sight of food might lead to increased discounting for food but not for sex, while the sight of an attractive potential sexual partner might lead to increased discounting for sex but not for food.

## IV. Risky Decision Making

A second natural application of our model is to risk preferences. Much as for time preferences, people are often of two minds when it comes to risks. We drive — or wish we were driving as we grip the airplane seat-divider with white-knuckles — even when we know at a cognitive level that it is safer to fly. We fear terrorism even when we know red meat poses a much greater risk of mortality. Perhaps the most dramatic illustration, however, comes from the phobias in which people are unable to face risks that they recognize, objectively, to be harmless. Indeed, the fact that people pay for therapy to deal with their fears, or take drugs (including alcohol) to overcome them, suggests that people's deliberative selves are not at peace with their affective reactions to risks.

To apply our two-system approach to risk preferences, we must make assumptions about how the two systems respond to risks. For the deliberative system, a natural assumption is that risks are evaluated according to their expected utility (or perhaps expected value). Indeed, most researchers, as well as knowledgeable lay people, agree that expected-utility theory is the appropriate prescriptive theory to use for evaluating risks. It is less obvious what drives the affective system, but we suggest that non-linear probability weighting and loss aversion --- two prominent features in many descriptive theories of risk preferences (for a recent review see Starmer, 2000) --- derive from the affective system. In this section, we investigate the implications of our model given these assumptions.

Whereas expected-utility theory assumes that probabilities are weighed linearly, many successful descriptive theories of risk preferences assume that people transform the probabilities into decision weights. The most common form of probability weighting is the S-shaped probability-weighting function, wherein low probabilities are overweighted and high probabilities are underweighted, and, perhaps most importantly, people are somewhat insensitive to changes in the probabilities (except at the extremes). There is, in fact, direct evidence that supports our contention that such insensitivity to probabilities derives from the affective system. Studies that measure fear by means of physiological responses such as changes in heart rate and skin conductance — which primarily reflect activity in the affective system — find that reactions to an uncertain impending shock depend on the expected intensity of the shock but not the likelihood of receiving it (except if it is zero) (Deane, 1969; Bankart & Elliott, 1974; Elliott, 1975; Monat, Averill, & Lazarus, 1972; Snortum & Wilding, 1971). Other evidence shows that emotional responses result largely from *mental images* of outcomes (Damasio, 1994). Because such images are largely invariant with respect to probability — one's mental image of winning a lottery, for example, depends a lot on how much one wins but not that much on one's chance of winning — emotional responses tend to be insensitive to probabilities.

Consider a simple model of risk preferences that incorporates a non-linear response to probability by the affective system.[19] Suppose that each option $x$ in the choice set $X$ is a lottery $x \equiv (x_1, p_1;...;x_N, p_N)$. The deliberative system evaluates options according to their expected value, and so its utility function is $U(x) = \sum p_i x_i$.[20] The affective system, in contrast, is less sensitive to the probabilities. For our exposition, we assume that the affective system evaluates options using the probability-weighting function suggested by Prelec (1998). Specifically, the affective system's motivational function is $M(x,a) = \sum w(p_i, a)\, x_i$, where $w(p_i,a) = \exp(-[-\ln p_i]^{1-a})$.[21] By assuming that an increase in $a$ makes the probability-weighting function more S-shaped, this formulation incorporates that having more intense affective motivations --- for any of the reasons discussed in Section II --- makes the affective system more insensitive to probabilities.

According to our model, the person will choose the option $x$ that maximizes:

$$V(x) = \sum p_i x_i + h(W,\sigma)* \left[M(x^A, a) - \sum \left(\exp(-[-\ln p_i]^{1-a})\right) x_i\right]$$

where $x^A$ is the best option from the perspective of the affective system. Because this affective optimum is exogenous to the person's choice, this is equivalent to choosing the $x$ that maximizes $\tilde{V}(x) = \sum \tilde{w}(p_i, a)x_i$ where:

$$\tilde{w}(p_i, a) = \frac{p_i + h(W,\sigma)\exp(-[-\ln p_i]^{1-a})}{1 + h(W,\sigma)}.$$

This formulation yields the straightforward conclusion that, if the deliberative system weights probabilities linearly while the affective system has an S-shaped probability-weighting function, the combined effect will be an S-shaped probability-weighting function that is less S-shaped than that for the affective system. More importantly, this formulation generates testable predictions by applying the two general implications from Section II.

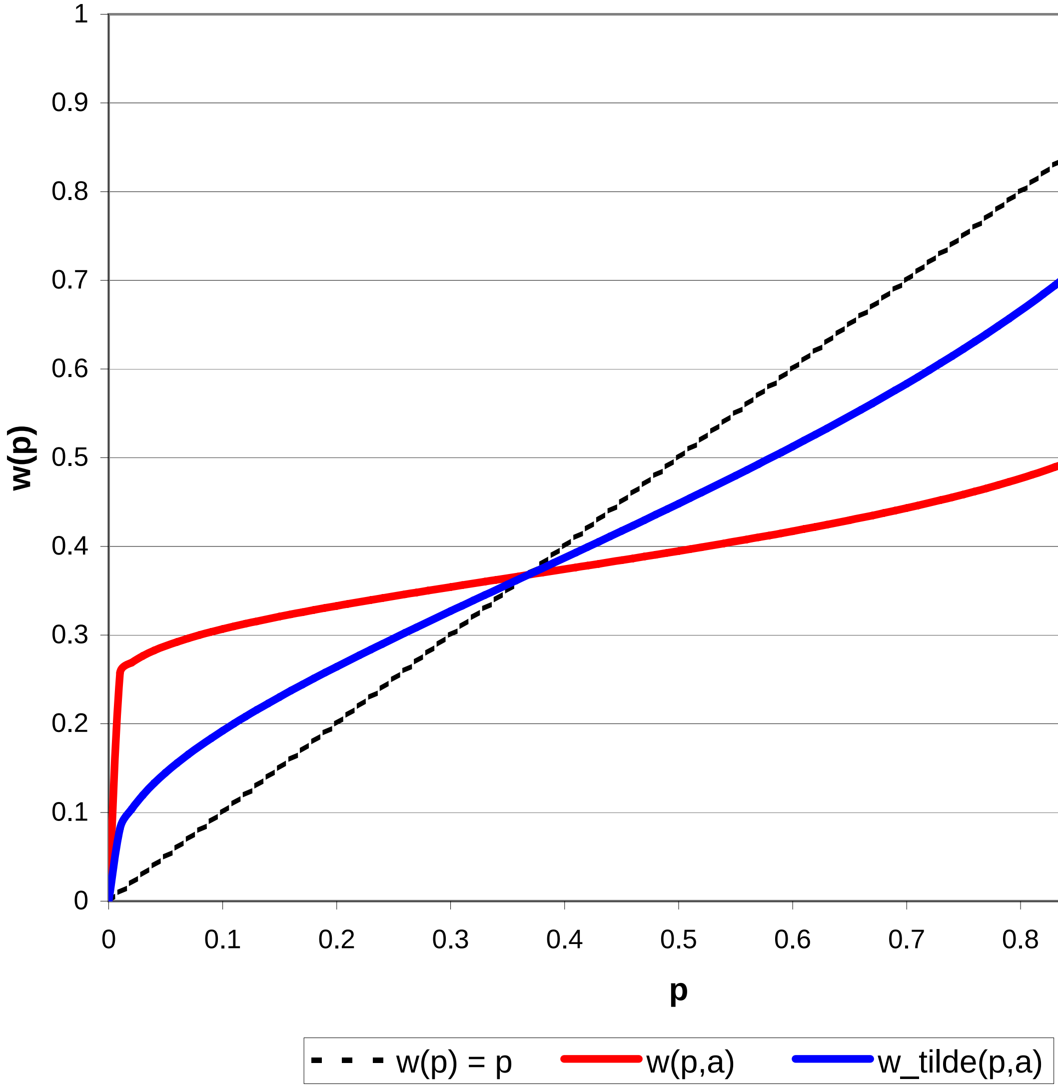
Implication #1: Willpower depletion or cognitive load will increase $h(W,\sigma)$ and thus make $\tilde{w}(p_i, a)$ more S-shaped.

Implication #2: Any factor that increases the intensity of the affective motivation will increase $a$ and thus make $\tilde{w}(p_i, a)$ more S-shaped.

These implications are illustrated in Figure 2.[22] In particular, Figure 2 illustrates how $\tilde{w}(p_i, a)$ will be in between $w(p_i,a)$ and linear weighting. If either willpower depletion or cognitive load increase $h(W,\sigma)$, then $\tilde{w}(p_i, a)$ will be closer to $w(p_i,a)$ and thus more S-shaped (Implication #1). If affective intensity increases, $w(p_i,a)$ will become more S-shaped, which implies, holding $h(W,\sigma)$ constant, that $\tilde{w}(p_i, a)$ will become more S-shaped as well.

While we know of no existing evidence on the effects of willpower depletion or cognitive load on probability weighting, there several supportive pieces of evidence on the effects of affective intensity. In one study (Ditto, Epstein, & Pizarro, 2004), 80 undergraduates played a game in which they could win chocolate chip cookies at the risk of being required to work on a boring task in the lab for an extra 30 minutes. For half of the participants, there was a high chance of winning the cookies (80%), and for the other half, there was a lower chance (60%). Crossed with this manipulation, half of the participants were only told about the cookies,

**Figure 2:  Probability-Weighting Function**



whereas for the other half the cookies were freshly baked in the lab and placed in front of participants as they made their decision. Participants for whom the cookies were merely described were less likely to play the game under high-risk than under low-risk conditions (45% vs. 95%), but participants who could see and smell the cookies, in contrast, were about as likely to play the game under high-risk as under low-risk conditions (80% vs. 85%). Hence, exactly as our model predicts, people are less sensitive to probabilities when affective intensity --- desire for cookies --- is high. Another set of studies compares "affect-rich" decisions to "affect-neutral" decisions. Again consistent with the prediction from our model, Rottenstreich & Hsee (2001) find that probability-weighting for affect-rich outcomes such as kisses, electric shocks, and vacations is more S-shaped than probability-weighting of affect-poor outcomes such as money. More generally, Sunstein (2002; 2003) documents a variety of ways in which strong emotions can lead to insensitivity to significant variations in probabilities and discusses the implication of such "probability neglect" for a wide range of legal and policy applications.

A second prominent feature in many descriptive theories of risk preferences is loss aversion, which is the tendency to weight losses more heavily than gains. Much as for non-linear probability weighting, we suggest that loss aversion derives from the affective system, and investigate the implications of this perspective. Again, there is direct evidence in support of this perspective. For instance, Chen et al (2006) introduce a currency into a colony of capuchin monkeys, and then assess how their behavior corresponds to economics models. When they present the monkeys with gambles, the capuchin monkeys display loss aversion. To the extent that animal behavior reflects humans' affective system, this result suggests that loss aversion indeed derives from the affective system.

There is also neuroscientific evidence. Tom et al (2007) collected fMRI data while subjects decided whether to accept or reject 50-50 gambles to win a sum of money vs. lose another sum of money. The gambles differed in the magnitudes of the gains and losses, and the researchers found that dopaminergic midbrain regions (i.e., affective regions) react to these changes. Moreover, these regions display a kind of neural loss aversion: the increase in activity when the gain amount increases is smaller than the decrease in activity when the loss amount increases. Another piece of direct evidence comes from a study by Shiv et al (2003), who compared normal people, patients with brain lesions in regions related to emotional processing (they were normal on most cognitive tests, including tests of intelligence), and patients with lesions in regions unrelated to emotion. Subjects made 20 rounds of investment decisions, where in each round they were given a dollar and made a choice between keeping it or wagering it on a 50-50 chance of losing it or winning $2.50. Patients with emotion-related lesions invested more often than other subjects — that is, they exhibited less loss aversion — and ultimately earned more money. Moreover, whereas normal people and patients with lesions unrelated to emotion were influenced by their outcomes in previous rounds, patients with emotion-related lesions were not.

Consider a simple model of risk preferences in which the affective system exhibits loss aversion. As above, suppose that each option $x$ in the choice set $X$ is a lottery $x \equiv (x_1, p_1;...;x_N, p_N)$, and that the deliberative system's utility function is $U(x) = \sum p_i x_i$. The affective system, in contrast, weighs losses more heavily than gains (and, for simplicity, weights probabilities linearly). Specifically, the affective system's motivational function is $M(x,a) = \sum p_i v(x_i, a)$ where

$$v(x_i, a) = \begin{cases} x_i & if\ \ x_i \geq 0 \\ \lambda(a) x_i & if\ \ x_i \leq 0. \end{cases}$$

We incorporate that more intense affective motivations imply more loss aversion by assuming that $\lambda(0)=1$ and that an increase in $a$ implies an increase in $\lambda$.

According to our model, the person will choose the option $x$ that maximizes:

$$V(x) = \sum p_i x_i + h(W,\sigma)* \left[M(x^A, a) - \sum p_i v(x_i, a)\right]$$

where $x^A$ is the best option from the perspective of the affective system. Because this affective optimum is exogenous to the person's choice, this is equivalent to choosing the $x$ that maximizes $\tilde{V}(x) = \sum_{i=1}^{N} p_i \tilde{v}(x_i, a)$ where

$$\tilde{v}(x_i, a) = \begin{cases} x_i & if\ \ x_i \geq 0 \\ \tilde{\lambda}(a) x_i & if\ \ x_i \leq 0. \end{cases}$$

and $\tilde{\lambda}(a) = \dfrac{1 + h(W,\sigma)\lambda(a)}{1 + h(W,\sigma)}$.

Much as for probability weighting, this formulation yields the straightforward conclusion that, if the deliberative system weights gains and losses equally while the affective system has loss aversion, the combined effect will be loss aversion that is less strong than that for the affective system. But again, we are more interested in the testable predictions generated by applying the two general implications from Section II.

Implication #1: Willpower depletion or cognitive load will increase $h(W,\sigma)$ and thus increase $\tilde{\lambda}(a)$.

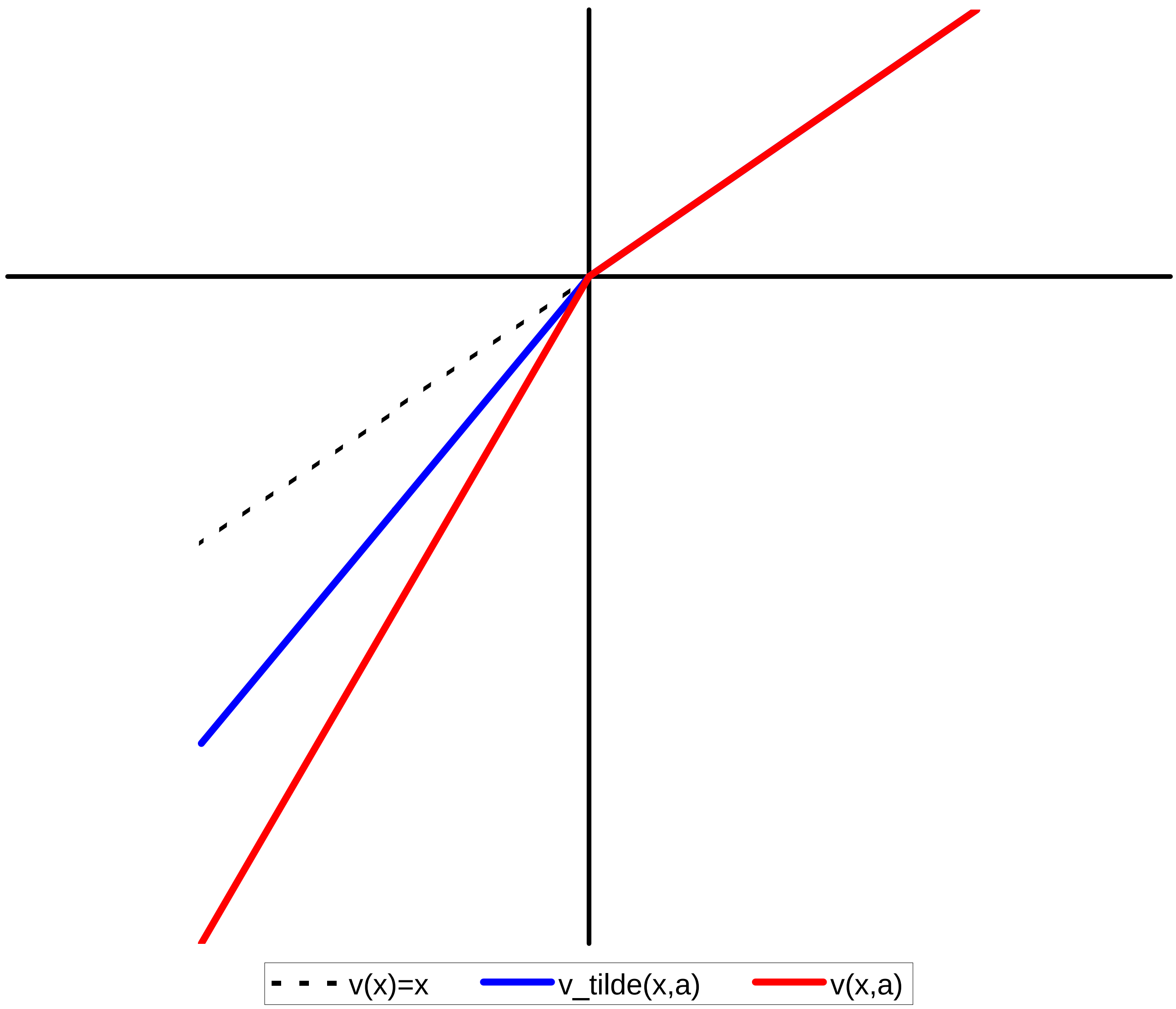
Implication #2: Any factor that increases the intensity of the affective motivation will increase $a$ and thus increase $\tilde{\lambda}(a)$.

These implications are illustrated in Figure 3. Figure 3 illustrates that $\tilde{v}(x_i, a)$ will lie between $v(x_i,a)$ and a linear value function. If either willpower depletion or cognitive load increase $h(W,\sigma)$, then $\tilde{v}(x_i, a)$ will be closer to $v(x_i,a)$ and thus exhibit more loss aversion (Implication #1). If affective intensity increases, $v(x_i,a)$ will become more kinked, which implies, holding $h(W,\sigma)$ constant, that $\tilde{v}(x_i, a)$ will become more kinked as well.

Although we are not aware of any evidence on willpower depletion, Benjamin, Brown, & Shapiro (2006) provide some evidence on the effects of cognitive load. In addition to asking the time preference questions described in Section III, they also asked their subjects to make a series of choices between a sure payoff and a risky payoff (safe vs. risky) and a series of choices between a risky payoff and a riskier payoff (risky vs. risky), and they classify people as risk-neutral if their answers are consistent with maximizing expected value. Relative to control subjects, subjects who answer these questions while under cognitive load showed substantial reductions in the propensity to be risk-neutral, although this difference was statistically significant only for the risky-vs.-risky choices. To the extent that small-stakes risk aversion derives from loss aversion (as suggested by Rabin (2000) and Rabin & Thaler (2000)), these results are consistent with cognitive load leading to increased loss aversion.

There is also a variety of support for the role of affect. One type of such evidence examines the role of affect for the "endowment effect" — the tendency to value an object more highly when one owns it, a tendency that is commonly attributed to loss aversion (e.g., Tversky & Kahneman, 1991). Considerable research suggests that the endowment effect is more pronounced for emotional outcomes such as changes in health status (see for instance Thaler,

**Figure 3: Value Function**



1980). In one meta-analysis (Horowitz & McConnell, 2002), whereas the mean ratio of willingness to accept relative to willingness to pay for ordinary private goods was found to be about 2.9, the mean ratio for goods involving health and safety was 10.1.

Another source of increased affective intensity is the temporal proximity of uncertain outcomes. In particular, our model predicts that as uncertain outcomes approach in time and thus affective intensity increases, people ought to become more loss averse. There is a great deal of evidence that temporal proximity is an important determinant of fear responses. As the prospect of an uncertain aversive event approaches in time, fear tends to increase even when cognitive assessments of the probability or likely severity of the event remain constant (Loewenstein, 1987; Roth et al, 1996). Similarly, after the moment of peak risk recedes into the past (e.g., after a near-accident), fear lingers for some period, but dissipates over time.[23] Evidence that temporal proximity can influence risk behaviors comes from studies that document “chickening out” wherein people initially agree to do various embarrassing things in front of other people (mime, sing, tell jokes, or dance) in exchange for payment, but then later change their minds (Van Boven et al, 2002). Moreover, consistent with changes in the affective state of fear being the cause, subjects who were shown a film clip designed to induce fear (from Kubrick's “The Shining”) right before they made their initial decision were much less likely to choose to perform, and hence much less likely to change their minds when the time came.[24]

Finally, we briefly mention a third feature that sometimes appears in descriptive theories of risk preferences (most notably, prospect theory): people show diminishing sensitivity to both gains and losses. Diminishing sensitivity generates an asymmetry in how people respond to gains vs. losses wherein people are risk-averse over gains and risk-seeking over losses. A natural extension of our analysis above would posit that this departure from risk neutrality also derives from the affective system. In fact, there is some recent neuroscientific evidence that supports this assumption (De Martino et al, 2006). The implications of this assumption would be analogous to those above.

In conclusion, taking account of interactions between affect and deliberation helps to make sense of several of the most important behavioral effects in the literature on decision making under risk. It also leads to novel predictions about factors that will affect the strength of these effects – that will determine the extent to which people weigh probabilities in a non-linear fashion and/or overweight losses relative to gains.

## V. Social Preferences

It cannot be controversial to anyone but a Vulcan that social preferences are powerfully influenced by affect. Humans experience a wide range of social emotions, from powerful empathic responses such as sympathy and sadness to more negative emotions such as anger and envy. To give a flavor for how our two-system perspective can be applied to social preferences, in this section we apply our model to one specific social motive — altruism — and its associated affect — sympathy. The perspective we suggest is that the deliberative system has a stable concern for others driven by moral and ethical principles for how one ought to behave. The affective system, in contrast, is driven toward anything between pure self-interest and extreme altruism depending on the degree of sympathy that is triggered.[25]

One motivation for this perspective comes from studies of other-regarding behavior in animals, which, again, we take as evidence for what drives the affective system. Animals, including monkeys and rats, can be powerfully moved by the plight of others (for an overview, see Preston & de Waal, 2002). For example, rats who view a distressed fellow rat suspended in air by a harness will press a bar to lower the rat back to safe ground (Rice & Gainer, 1962). A more recent study demonstrates that rats can have such powerful sympathetic reactions to others that they become debilitated — specifically, when another rat is administered electric shocks, the focal rat may retreat to a corner and crouch there motionless (Preston & de Waal, 2002). In another remarkable study (Masserman, Wechkin, & Terris, 1964), hungry rhesus monkeys were trained to pull two chains, one of which delivered half as much food as the other. The situation was then altered so that pulling the chain with the larger reward caused a monkey in sight of the subject to receive an electric shock. After witnessing such a shock, two-thirds of the monkeys preferred the non-shock chain and, of the remaining third, one monkey stopped pulling either chain for 5 days and another for 12 days after witnessing another being shocked.

At the same time, other-regarding behavior is not always observed in animals. In the primate studies, for instance, self-starvation to avoid shocking another animal was induced more by visual than auditory cues (i.e., seeing as opposed to hearing the distress of the other animal), was more likely in animals that had experienced shock themselves, was enhanced by familiarity with the shocked individual, was less when the shock recipient was an albino, and was nonexistent when it was a different species of animal. Perhaps stretching the terminology we introduced in Section II, we can interpret these findings as decreased proximity reducing the concern for others.

Humans, like animals, are capable of remarkable depths of sympathy toward others in some circumstances, and remarkable indifference in other circumstances. In the Theory of Moral Sentiments, Adam Smith (1759:192-3) provides a chilling account of the latter:

> Let us suppose that the great empire of China, with all its myriads of inhabitants, was suddenly swallowed up by an earthquake, and let us consider how a man of humanity in Europe, who had no sort of connection with that part of the world, would be affected upon receiving intelligence of this dreadful calamity. He would, I imagine, first of all express very strongly his sorrow for the misfortune of that unhappy people, he would make many melancholy reflections upon the precariousness of human life, and the vanity of all the labours of man, which could thus be annihilated in a moment…. And when all this fine philosophy was over…, he would pursue his business or his pleasure, take his repose or his diversion, with the same ease and tranquility as if no such accident had happened. The most frivolous disaster which could befall himself would occasion a more real disturbance. If he was to lose his little finger to morrow, he would not sleep to-night; but, provided he never saw them, he will snore with the most profound security over the ruin of a hundred millions of his brethren.

If people based their behavior toward other people solely on their affective reactions to them, then, sympathetic beggars would be millionaires and United Way would go out of business. Given this remarkable lack of connection between our sympathy toward others and the gravity of their plight or need for assistance, how is it that humans behave in an at all sensible way toward fellow humans? According to Adam Smith, the answer is “reason, principle, conscience, the inhabitant of the breast, the man within, the great judge and arbiter of our conduct” (1759:194). In other words, it is the fact that the deliberative system moderates the sympathetic reactions of the affective system.

There is also some direct evidence for the objectives of the deliberative system. People seem to have well-defined notions of what would be a fair or reasonable allocation of resources between two unknown people (Yaari & Bar-Hillel, 1984). Moreover, these well-defined notions also influence people’s choices in simple allocation (dictator) games in which they allocate resources between themselves and anonymous others — situations that should evoke relatively little sympathy (Andreoni & Miller, 2002; Charness & Rabin, 2002). Hence, it seems likely that the goals of the deliberative system reflect some combination of moral and ethical principles for how one ought to behave. Indeed, philosophers have long discussed how people's behavior can be influenced by sophisticated reasoning about ethical principles (e.g., Kant, 1785).[26]

Consider a simple model of social preferences that formalizes our perspective. Suppose that each option $x$ in the choice set $X$ is a pair of payoffs $x = (x_S, x_O)$, where $x_S$ is a payoff for oneself and $x_O$ is a payoff for another person. The deliberative system puts some stable weight $\phi$ on the other person's payoff, and so its utility function is $U(x) = x_S + \phi x_O$. The affective system, in contrast, puts a variable weight on the other person's payoff that depends on the degree of sympathy that the person currently feels toward the other. Since the degree of sympathy is naturally interpreted as the intensity of affect, we assume that the affective system's motivational function is $M(x,a) = x_S + a\, x_O$.

According to our model, the person will choose the option $x$ that maximizes:

$$V(x) = [x_S + \phi x_O] - h(W,\sigma)[M(x^A,a) - ( x_S + a\, x_O)]$$

where $x^A$ is the best option from the perspective of the affective system. Once again, this affective optimum is independent of the person's choice, and so this is equivalent to choosing $x$ to maximize $\tilde{V}(x) = x_S + \tilde{\phi}(a)\, x_O$ where

$$\tilde{\phi}(a) = \frac{\phi + h(W,\sigma)a}{1 + h(W,\sigma)}.$$

Unlike for time preferences and risk preferences, where the affective system moves behavior away from the deliberative optimum in one systematic direction, here the affective system can push behavior towards more or less concern for others relative to the deliberative optimum. In situations where there is very little sympathy triggered in the affective system, the affective system will push behavior closer to pure self-interest — as reflected by $a < \phi$ implying $\tilde{\phi}(a) < \phi$. In contrast, in situations where there are very high levels of sympathy triggered in the affective system, the affective system will push behavior towards more altruism — as reflected by $a > \phi$ implying $\tilde{\phi}(a) > \phi$. When a person passes a sympathetic beggar on the street, the person may find herself giving money to the beggar when she thinks she ought to give that money to the United Way. At the same time, when a person is at home and not experiencing any sympathetic reactions, she may find herself not giving to the United Way when she thinks she ought to (it often seems to require "effort" to write that check to the United Way).

Once again, we focus on testable predictions derived from applying the general implications from Section II.

Implication #1: Willpower depletion or cognitive load will increase $h(W,\sigma)$, which will increase $\tilde{\phi}(a)$ when affective intensity is high ($a > \phi$) and decrease $\tilde{\phi}(a)$ when affective intensity is low ($a < \phi$).

Implication #2: Any factor that increases the intensity of the affective motivation will increase $a$ and thus increase $\tilde{\phi}(a)$.

Implication #1 reflects that our model yields more nuanced predictions in the realm of social preferences than for time preferences or risk preferences because the implications of willpower depletion or cognitive load depend on the degree of sympathy triggered. Specifically, when a person experiences little or no sympathy, as when deciding whether to donate to the United Way, our model predicts that willpower depletion or cognitive load should reduce the likelihood of the act. In contrast, when a person experiences high sympathy, as when deciding whether to pay for a sympathetic beggar's dinner, our model predicts that willpower depletion or cognitive load should increase the likelihood of the act. In fact, there is some initial, albeit limited, evidence for both willpower depletion and cognitive load.

A study by Gailliot et al (2007) provides limited support for the effects of willpower depletion. They provide a series of studies designed to demonstrate that self-control tasks reduce blood glucose levels, and that consuming a glucose drink between two self-control tasks (as in the usual willpower paradigm) would eliminate the effect of willpower depletion. In one study, subjects first endogenously decide how much self-control to exert on a first task --- specifically, the first task was taking an actual examination that counted for course credit, and the measure of self-control exerted was the order in which the exams were turned in (previous research suggests exam performance and intelligence are not related to exam completion order). Subjects were then randomly assigned to consume a glucose drink or a placebo. Finally, after a short delay to allow their bodies to process the glucose, subjects answered two questions on helping behavior, how much they would donate to a charity, and how much they would help a stranger from their class who had been evicted from his or her apartment. Consistent with their glucose hypothesis, among subjects who consumed the glucose drink, there was no significant relationship between exam completion order and helping behavior. More relevant for our purposes, among subjects who consumed the placebo, the later they turned in their exams (more willpower depletion), the less likely they were to engage in helping behavior. Because donating to a charity and helping a stranger would seem to be low-sympathy situations, this result is consistent with our prediction that, when sympathy is low, willpower depletion leads to less concern for others.

A study by Skitka et al (2002) provides limited support for the effects of cognitive load. Subjects were shown a number case studies of people who had contracted AIDS in different ways, and different case studies made the victim appear more or less responsible (e.g., sexual contact versus a blood transfusion). For each case study, subjects were asked whether the individual should be given subsidized access to drug treatment, and filled out measures of blame and responsibility. In addition, subjects were asked their political orientations. The key manipulation for our perspective is that half of the subjects made their judgments and allocation decisions while also engaged in a tone-tracking task that has been commonly used to induce cognitive load. The study found that subjects were less likely to advocate subsidized treatment under conditions of high load, which we would interpret as evidence that deliberative reactions are more sympathetic than affective reactions to AIDS victims. More interestingly, under conditions of high load, both liberals and conservatives were less likely to provide subsidized treatment to those deemed responsible (relative to those deemed not responsible), whereas under conditions of low load, liberals treated both groups equally whereas conservatives continued to favor groups who were seen as less responsible for contracting the disease. These findings are consistent with our framework if affective and deliberative reactions were consistent for conservatives — so cognitive load has no effect — but conflicting for liberals.

There is also evidence on the effects of affective intensity (Implication #2). Perhaps the most direct evidence comes from a study by Batson et al (1995) on empathy-induced altruism. They manipulate subjects' empathy toward a target individual by having them read a short description of that individual's need while taking an objective perspective (low empathy) or while trying to imagine how that individual feels (high empathy). They then give subjects the opportunity to help the target despite the fact that doing so would violate some moral principle of justice such as random allocations or allocation based on need. Consistent with Implication #2, they found that subjects in the high-empathy treatment were much more likely to help the target individual.

Implication #2 also helps to explain why people treat statistical deaths differently than identifiable ones, since foreknowledge of who will die (or which group deaths will come from) creates a more concrete — and evocative — image of the consequences (see, Schelling, 1968; Bohnet & Frey, 1999; Slovic, forthcoming; and Small & Loewenstein, 2003 for an experimental demonstration). The impact of identifiability on affect may help to explain an anomalous tendency of altruists to contribute more to specific instances of a problem than to appeals addressing the entire problem, and more to specific victims than to multiple victims, even when the latter dominates the former in terms of total help rendered (Kogut & Ritov, 2003). Requests for donations to medical research which base their requests on the testimony of a single "poster child" rather than general descriptions of the affliction or its prevalence seem to exploit this phenomenon.[27]

A recent study by Small, Loewenstein, & Slovic (2007) provides further support for our dual-process perspective on the role of identifiability. They provide subjects with the opportunity to donate to a real charity (*Save the Children*), but they manipulate whether subjects are shown an identifiable victim (a picture and a description of a little girl) or a statistical victim (factual information about the overall problem). In addition, they also manipulate the extent to which people are primed to think more deliberatively. They find that deliberative thought tends to decrease donations to the identifiable victim, but does not affect donations to the statistical victim. Under the plausible assumption that the affective system is playing a major role in donations to the identifiable victim and very little role in donations to the statistical victim, these results are exactly what our model (Implication #2) predicts.

While we have focused our analysis solely on the simple social motive of altruism, researchers have discussed other social motives as well. For instance, there is a large literature that focuses on people's concerns for relative payoffs --- that is, how their payoffs compare to others' payoffs (Messick & Sentis, 1985; Loewenstein, Thompson, & Bazerman, 1989; Fehr & Schmidt, 1999; Bolton & Ockenfels, 1999). There is another literature that focuses on reciprocity --- that is, a desire to be kind to those who are kind to you and unkind to those who are unkind (Rabin, 1993). In principle, our model could be applied to these social concerns as well; however, because both concerns would seem to have both a deliberative and an affective component, it is not entirely obvious what to assume about the motives of the two systems (although see Knoch et al, 2006 for some suggestive neural evidence on reciprocity). Hence, we leave an analysis of these motives for future research.

## VI. Implications for the Evaluation of Well-being

A major new thrust of behavioral decision research and behavioral economics is to use insights from psychology to devise policies that help people to make better decisions. In particular, a variety of evidence documents systematic ways in which people fail to maximize their own well-being, and thus suggests scope for policy to help counteract these failures. Initial research on this issue has taken a cautious approach, attempting to identify policies that create minor changes to incentives or to how choices are presented that, for the most part, do not restrict the autonomy action --- don't tell people what to do --- while at the same time have a beneficial influence for people who fail to maximize their own well-being (see, Camerer et al, 2003; Thaler & Sunstein, 2003). But as this literature continues to develop, there is a major open question that needs to be addressed: What is the appropriate measure of well-being that should be used to judge (a) whether people are indeed failing to maximize their own well-being and (b) whether the benefits of proposed policies outweigh the costs.

In economics, the standard approach to the evaluation of well-being is the *revealed-preference approach*: we infer a person's preferences from observing her choices, and we then use those preferences to evaluate her well-being. Because this approach merely assumes *a priori* that whatever a person does must correspond to what maximizes her well-being, it is inconsistent with the very notion of people failing to maximize their well-being. In contrast, our model provides some guidelines for how to measure well-being when it cannot be directly inferred from behavior.

In particular, our model suggests two natural candidates for measuring well-being: (i) the deliberative system's utility function $U$, which reflects the desirability of actions as perceived by the deliberative system; and (ii) the deliberative system's objective function $V$, which reflects both the desirability of actions as perceived by the deliberative system and the willpower effort required to implement actions.  While each has its advantages, unfortunately, neither candidate seems entirely appropriate.  (A third possibility is the affective system's motivational function *M*, but we think everyone would agree that this is inappropriate.)

The main argument in favor of using the deliberative system's utility function *U* as the criterion for evaluating well-being is that it represents how people would "like" to behave, both from a removed perspective and even in the moment.  A major problem with this approach arises from the fact that it ignores reciprocal interactions between deliberation and affect.  One manifestation of these interactions is that affect might influence the broader goals of the deliberative system.  In other words, if one were to elicit from a person how she would like to behave, the answer is likely not invariant to the current activity in the affective system.  A possible response is that the proper way to elicit broader goals is to first put people in an affectively neutral state.  But even then, to the extent that deliberative system needs affective inputs to evaluate different options (as suggested by Damasio's research on patients with damage to the prefrontal cortex), we still might not get a "true" measure of the person's broader goals.[28]

A second major problem with this approach arises from the failure to take any account of willpower effort.  Even holding constant behavioral outcomes, if implementing those outcomes required willpower effort, and if that effort was unpleasant, then it seems inappropriate to ignore that unpleasantness in the analysis of well-being.  Put more concretely, it would seem that a policy that did not affect behavioral outcomes but reduced the willpower effort required to implement those outcomes would make people better off, but we would conclude otherwise if we use *U* as the criterion for well-being.  Indeed, a large body of research on "hot-cold empathy gaps" (e.g., Van Boven and Loewenstein, 2003; Loewenstein and Angner, 2003) suggests that people tend to give little if any weight to affective states they are not currently experiencing, even when such affective states (arguably) confer real utility and disutility and hence should normatively be taken into account.  It is easy to commit to a diet when one is not currently hungry or to decide to go cold turkey right after satiating oneself on a drug, but it is likely that resolutions of this type pay insufficient heed to the miseries that would actually be involved in implementing the decision.

So should we instead use the deliberative system's objective function *V* as the criterion for well-being?  Doing so would be more in line with the standard revealed-preference approach, because *V* represents the "preferences" that would be inferred from the person's behavior.  The major problem with this approach, however, is that it corresponds to a belief that actual behavior maximizes well-being, and there are reasons to believe that actual behavior often reflects an excessive influence of affect. One reason to be wary is that affective states are relatively transient, and, as demonstrated by the research on incidental affect, can often influence behavior even when they are manifestly irrelevant to the decision at hand. An even bigger reason to be wary is that affective states can often be easily manipulated for strategic purposes that promote the interests of the manipulator to the detriment of the interests of the decision-maker. Finally, for intertemporal decisions there is evidence that people under-appreciate the effects of future affect.

Hence, neither candidate seems entirely appropriate. Ideally we would like something in-between that recognizes the value of policies that reduce the willpower required to implement outcomes, but also something that takes limited account of affective states, particularly those states that are transient, easily manipulated, and tangential to decisions. At this point, there is no clear measure of well-being that has these properties.

## VII. Discussion

There is a great deal of evidence that people's decisions (and judgments and attitudes) are influenced by both affective and deliberative processes. Whereas standard rational-choice models focus, for the most part, on deliberative processes, our main contribution in this paper has been to develop a formal framework to incorporate affective processes. We conclude by discussing some broader implications of our framework.

While we have demonstrated the usefulness of our framework in three domains, it can be useful for a much broader set of applications. For instance, affective processes would seem to be important for understanding advertising, especially advertising that conveys no evident information; behavior in negotiations, and in particular why bargaining can break down into a mutually destructive, affect-driven morass (e.g., nasty divorces); behavior in financial markets and especially reactions to news events; and political preferences, and specifically how people seem to respond to political candidates, political parties at an affective level and why aptly named 'hot-button' issues such as gay marriage hold such sway over the electorate. Moreover, even in the domains we have discussed, a more detailed application of our framework would have the potential to explain many complex real-world behaviors.

There are a number of directions in which to further expand upon our framework. Perhaps the most important is to more fully explore the dynamics of willpower. While we have described the effects of short-term changes in willpower strength, the long-term dynamics of willpower may be more important. For instance, our model suggests an alternative explanation for why poor people might be more prone to engage in risky behaviors such as smoking, unsafe sex, and so forth. Existing explanations usually take the form of poor people's benefits being larger than rich people's, their costs being lower (poor people have less to live for), or the assumption that they are more impatient (short-sighted). Our model suggests that if poor people are constantly required to exert willpower to live within their means (i.e., to constantly forgo enticing purchases), then they will have relatively little willpower strength remaining to resist inexpensive temptations like cigarettes or a willing sexual partner.

The dynamics of willpower might be more important in making sense of the complicated patterns of self-control behavior (or lack thereof) that have been documented in the literatures on addiction, dieting, and sexual risk-taking. Drug addicts, for instance, binge, go cold turkey, enter rehab programs, flush their drugs down the toilet, and relapse. Dieters exhibit frequent bouts of overeating that can be triggered by such disparate events as having eaten something that "breaks" one's diet; thinking that one has done this (whether the food eaten was truly high calorie or not); feeling anxious, depressed, or otherwise dysphoric; drinking alcohol; eating with someone else who overeats; smelling and thinking about attractive foods; or being deprived of a favorite food (for an overview, see Herman and Polivy, 2003). To understand these behavioral complexities, we need to incorporate the dynamics of interactions between the two systems.

There are even more nuanced willpower dynamics. For instance, some, albeit preliminary, studies have found support for the idea that, in addition to being depleted in the short-term by exertion, willpower, like a muscle, may become strengthened in the long-term through repeated use (Muraven, Baumeister, and Tice, 1999). More importantly, people's behavior might also reflect their attempts to manage their use of willpower. There is in fact experimental evidence in (a modified version of) the Baumeister paradigm that people do have some awareness of the dynamic properties of willpower and take these into account in a strategic fashion (Muraven, 1998). Specifically, people who were aware that there would be multiple willpower tasks seemed to conserve willpower on the earlier task (relative to subjects who were unaware), and in fact those subjects were able to exert more willpower on the later task.[29]

A second direction in which to expand our framework is to people's assessments of their own behaviors. Because such assessments are an inherently cognitive task, they will naturally tend to exaggerate the role played by deliberation. In effect, one could say that the deliberative self egocentrically views itself as in control and commensurately underestimates the influence of affect (see Wegner and Wheatley, 1999).[30] This failure to appreciate the role of affect in behavior can have a negative impact on efforts at self-control. Perhaps the most important form of self-control is not willpower, but rather "self-management" — the ability to avoid getting into, or to remove oneself from, a situation that is likely to engender self-destructive behaviors. Dieters can steer clear of banquets, drug addicts of places and persons associated with drug-taking, smokers of smoky bars, and alcoholics of bars and parties. To the extent that people are unaware of, or underappreciate, the impact of affect on their own behavior, they are likely to underutilize such strategies.

The failure to appreciate the role of affect in behavior can also lead people to take decisions that they later reverse. Real-world instances abound of people making some decision and later reversing it, often even at a cost. Such behaviors arise naturally in our framework, because "preferences" change as environmental stimuli and the costs of willpower change. One might be tempted to view such instances as instances of the deliberative system overturning earlier "bad" decisions driven by the affective system — as when a person gets married in the heat of passion and shortly thereafter has the marriage annulled. However, our model implies no such asymmetry. It could just as well be that a person's deliberative system chooses some "desirable" course of action when affective motivations are low, but then later reverses this decision after stronger affective motivations are triggered.

A third implication of failing to appreciate the role of affect is that people will exaggerate the importance of willpower as a determinant of self-control. People who are thin often believe they are thin due to willpower, and that those who are less fortunate exhibit a lack of willpower. However, it is far more likely that those who are thin are blessed (at least in times of plentiful food) with a high metabolism or a well-functioning ventromedial hypothalamus (which regulates hunger and satiation). Indeed, obese people who go to the extraordinary length of stapling their stomach to lose weight often report that they have a sudden experience of "willpower" despite the obvious fact that stapling one's stomach affects hunger rather than willpower (Gawande, 2001). It is easy and natural for those who lack drives and impulses for drugs, food, and sex to condemn, and hence to be excessively judgmental and punitive, toward those who are subject to them — to assume that these behaviors result from a generalized character deficit, a deficiency in willpower. Similarly, the rich, who are not confronted with the constant task of reigning in their desires, are likely to judge the short-sighted behaviors of the poor too harshly.

More speculatively, the deliberative system may have an ability to “train” the affective system over time to experience certain emotions — notably guilt and satisfaction — in a way that serves long-run goals. In particular, the deliberative system might train the affective system to experience guilt reactions in response to certain undesirable behaviors, and to experience feelings of satisfaction in response to certain desirable behaviors. For instance, many people seem to train themselves to experience satisfaction whenever they transfer money into their savings account and guilt whenever they transfer money out of that account; and many academics train themselves to experience guilt when they aren’t working. Of course, such training can have undesirable long-run effects. Someone who successfully creates affective reactions to savings behavior may find that those affective reactions persist even when it is logically time to start consuming those savings. Indeed, when these emotions become sufficiently intense and ingrained, they can actually drive behavior farther than the deliberative system wanted, producing disorders of excessive preoccupation with the future such as workaholism, tightwaddism, and, even more extremely, anorexia.[31]

After decades of domination by a cognitive perspective, in recent decades affect has come to the fore as a topic of great interest among psychologists. In this paper we attempt to integrate many of the findings from research conducted by psychologists and decision researchers interested in affect by proposing a formal model of interactions between affect and deliberation that can both explain existing findings and that also generates testable, but as yet untested, predictions. If further testing substantiates these predictions, and hence the model, this could constitute the first step toward a formal theoretical perspective that integrates two major sides of human judgment and behavior.

# References

Ainslie G (1975) Specious reward: a behavioral theory of impulsiveness and impulse control. Psychol Bull 82:463-496.

Ameriks, John, Andrew Caplin, John Leahy, and Tom Tyler (2003). "Measuring Self Control." Mimeo, New York University.

Andreoni, James and John Miller (2002). "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, **70**, 737-753.

Ariely, D. & Loewenstein, G. (2006). The heat of the moment: The effect of sexual arousal on sexual decision making. Journal of Behavioral Decision Making, 19(2), 87-98.

Armony JL, Servan Schreiber D, Cohen JD, & LeDoux JE (1995). An anatomically constrained neural network model of fear conditioning. Behavioral Neuroscience, 109, 246 256.

Ashraf, Nava, Colin Camerer, and George Loewenstein (2005). "Adam Smith: Behavioral Economist." *Journal of Economic Perspectives*, 19(3), 131-145.

Bankart, C. Peter and Rogers Elliott (1974). "Heart Rate and Skin Conductance in Anticipation of Shocks with Varying Probability of Occurrence." *Psychophysiology*, **11**, 160-174.

Bargh, John A. (1984). "Automatic and Conscious Processing of Social Information," in R. S. Wyer and T. K. Srull, eds., *Handbook of Social Cognition (Vol 3)*. Hillsdale, NJ: Erlbaum, 1-43.

Barlow, David H. (1988). *Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic*. New York, NY: Guilford Press.

Batson, C. Daniel, Tricia R. Klein, Lori Highberger, and Laura L. Shaw (1995). "Immorality From Empathy-Induced Altruism: When Compassion and Justice Conflict." Journal of Personality and Social Psychology, 68(6), 1042-1054.

Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1994). Losing Control: How and Why People Fail at Self-Regulation. San Diego: Academic Press.

Baumeister, Roy F. and Kathleen D. Vohs (2003). "Willpower, Choice, and Self-Control," in George Loewenstein, Daniel Read and Roy F. Baumeister, eds., *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. New York: Russell Sage Foundation, 201-216.

Bechara, A., H. Damasio, et al. (1997). "Deciding advantageously before knowing the advantageous strategy." Science 275(5304): 1293-1294.

Becker, Gary & Rayo, Luis (forthcoming). Evolutionary Efficiency and Mean Reversion in Happiness. Journal of Political Economy.

Benhabib, Jess and Alberto Bisin (2005). "Modelling Internal Commitment Mechanisms and Self-Control: A Neuroeconomics Approach to Consumption-Saving Decisions." *Games and Economic Behavior*, 52(2), 460-492.

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro (2006). "Who is 'Behavioral'? Cognitive Ability and Anomalous Preferences." Mimeo, Harvard University.

Ben-Zion, U., A. Rapoport, et al. (1989). "Discount rates inferred from decisions: An experimental study." Management Science 35: 270-284.

Bernheim, B. Douglas and Antonio Rangel (2003). "Emotions, Cognition, and Savings: Theory and Policy." Mimeo, Stanford University.

Bernheim, B. Douglas and Antonio Rangel (2004). "Addiction and Cue-Triggered Decision Processes." *American Economic Review*, **94**(5), 1558-1590.

Berns, Gregory S., Jonathan Chappelow, Milos Cekic, Caroline F. Zink, Giuseppe Pagnoni, Megan E. Martin-Skurski (2006). Neurobiological Substrates of Dread Science 312, 754 - 758.

Berridge, Kent C. (1995). "Food Reward: Brain Substrates of Wanting and Liking." *Neuroscience and Biobehavioral Reviews*, **20**(1), 1-25.

Blood, Anne J. & Zatorre, Robert J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proceedings of the National Academy of Sciences, 98, 11818-11823.

Bodenhausen, Galen V. (1993). "Emotions, Arousal, and Stereotypic Judgments: A Heuristic Model of Affect and Stereotyping," in Mackie, D. M. and D. L. Hamilton, eds., *Affect, Cognition, and Stereotyping: Interactive Processes in Group Perception*. San Diego, CA: Academic Press, 13-37.

Bohnet, Iris and Bruno Frey (1999). "The Sound of Silence in Prisoner's Dilemma and Dictator Games." *Journal of Economic Behavior and Organization*, **38**, 43-57.

Bolton, G. and A. Ockenfels (2000). "ERC: A Theory of Equity, Reciprocity, and Competition," American Economic Review, 90, 166-193.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*, 129-148.

Brasil-Neto, Joaquim Pereira, Alavaro Pascual-Leone, J. Valls-Solé, L.G. Cohen, and M. Hallett (1992). "Focal Transcranial Magnetic Stimulation and Response Bias in a Forced-Choice Task." *Journal of Neurology, Neurosurgery and Psychiatry*, **55**(10), 964-966.

Buck, R. (1984). *The Communication of Emotion*. New York: Guilford Press.

Cabanac, Michel (1971). "Physiological Role of Pleasure." Science, 173 (September 17), 1103-1107.

Camerer, Colin, Issacharoff, Samuel, Loewenstein, George, O'Donoghue, Ted, Rabin, Matthew. 2003. Regulation for conservatives: Behavioral economics and the case for "asymmetric paternalism." *University of Pennsylvania Law Review* 151 (Jan.): 1211-1254.

Chaiken, Shelly and Yacov Trope, Eds. (1999). *Dual-Pocess Theories in Social Psychology*. New York: Guilford Press.

Charness, Gary and Matthew Rabin (2002). "Social Preferences: Some Simple Tests and a New Model." *Quarterly Journal of Economics*, **117**, 817-869.

Chen, M. Keith, Venkat Lakshminarayanan, and Laurie R. Santos (2006). "How Basic are Behavioral Biases? Evidence from Capuchin Monkey Trading Behavior." Journal of Political Economy, 114(3), 517-537.

Cleckley, Hervey Milton (1976). *The Mask of Sanity: An Attempt to Clarify Some Issues about the So-Called Psychopathic Personality*. St. Louis: Mosby.

Cohen, Jonathan D. (2005). "The Vulcanization of the Human Brain: A Neural Perspective on Interactions Between Cognition and Emotion and Optimality in Decision Making." Working Paper, Department of Psychology, Princeton University.

Cosmides, Leda and John Tooby, 2004. "Evolutionary Psychology and the Emotions" In Handbook of Emotions, 2nd Edition M. Lewis & J. M. Haviland-Jones, Editors. NY: Guilford.

Damasio, Antonio R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: G.P. Putnam.

Deane, G. E. (1969). "Cardiac Activity During Experimentally Induced Anxiety." *Psychophysiology*, **6**, 17-30.

DeBecker, G. (1997). The gift of fear: Survival signals that protect us from violence. Boston: Little, Brown, and Company.

De Martino, Benedetto, Dharshan Kumaran, Ben Seymour, and Raymond J. Dolan (2006). "Frames, Biases, and Rational Decision-Making in the Human Brain." Science, 313 (8/4/06), 684-687.

Ditto, Peter H., Eden Epstein, and David Pizarro (2004). "Motivational Myopia: Visceral Influences on Risk Taking Behavior." Poster presented at the Fourth Annual Meeting for the Society for Personality and Social Psychology, Austin, Texas.

Dunn, Barnaby D., Tim Dalgleish, Andrew D. Lawrence The somatic marker hypothesis: A critical evaluation Neuroscience and Biobehavioral Reviews xx (2005) 1–33

Eisenberger, Naomi I., Matthew D. Lieberman, and Kipling D. Williams (2003). "Does Rejection Hurt? An fMRI Study of Social Exclusion." *Science*, **302** (October 10, 2003), 290-292.

Elliott, Rogers (1975). "Heart Rate in Anticipation of Shocks Which have Different Probabilities of Occurrences." Psychological Reports, 36, 923–931.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. American Psychologist, 49, 709-724.

Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), Dual-process theories in social psychology . New York, NY: Guilford Press.

Fehr, E. and K. Schmidt (1999). "A Theory of Fairness, Competition, and Cooperation," Quarterly Journal of Economics, 114, 817-868.

Finucane, M. L., Peters, E., & Slovic, P. (in press). Judgment and decision making: The dance of affect and reason. In S. L. Schneider & J. Shanteau (Eds.), Emerging perspectives on judgment and decision research. New York: Cambridge University Press.

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). Psychological Bulletin, 117, 39-66.

Frederick, Shane 2003 "Measuring intergenerational time preference: Are future lives valued less?" Journal of Risk and Uncertainty. 26(1), 39-53.

Frederick, S. & Loewenstein, G. (1999). Hedonic Adaptation: In Well Being: The Foundations of Hedonic Psychology. Daniel Kahneman, Edward Diener, and Norbert Schwarz (Eds.). (pp. 302-329). New York: Russell Sage Foundation Press.

Frederick, S., G. Loewenstein, and T. O'Donoghue (2002), "Time Discounting and Time Preference: A Critical Review," Journal of Economic Literature, 40, 351-401.

Freud, Sigmund (1924). The Ego and the Id, translated by Joan Riviere. New York: W. W. Norton, 1962.

Frijda, Nico H. (1986). *The Emotions*. Cambridge: Cambridge University Press.

Fudenberg, Drew and David K. Levine (2004). "A Dual Self Model of Impulse Control." *American Economic Review*, 96(5), 1449-1476.

Gailliot, Matthew T., Roy F. Baumeister, C. Nathan DeWall, Jon K. Maner, E. Ashby Plant, Dianne M. Tice, Lauren E. Brewer, and Brandon J. Schmeichel (2007). "Self-Control

Relies on Glucose as a Limited Energy Source: Willpower Is More Than a Metaphor." Journal of Personality and Social Psychology, forthcoming.

Gawande, Atul (2001). "The Man Who Couldn't Stop Eating." *New York Times Magazine*, July 9, 2001.

Gilbert, Daniel T. and Michael Gill. 2000. "The Momentary Realist." *Psychological Science*, **11**, 394-398.

Gul, Faruk and Wolfgang Pesendorfer (2001). "Temptation and Self-Control." *Econometrica*, **69**, 1403-1435.

Halberstadt, J.B., & Niedenthal, P.M. (1997). Emotional state and the use of stimulus dimensions in judgment. Journal of Personality and Social Psychology, 72, 1017-1033.

Hariri, Ahmad R., Sarah M. Brown, Douglas E. Williamson, Janine D. Flory, Harriet de Wit, and Stephen B. Manuck (2006). "Preference for Immediate over Delayed Rewards is Associated with Magnitude of Ventral Striatal Activity." The Journal of Neuroscience, 26(51), 13213-13217.

Heatherton, T.F., C. Peter Herman, and Janet Polivy (1991). "Effects of physical threat and ego threat on eating behavior." *Journal of Personality and Social Psychology*, **60**, 138-143.

Herman, C. Peter and Janet Polivy (2003). "Dieting as an Exercise in Behavioral Economics," in George Loewenstein, Daniel Read and Roy F. Baumeister, eds., *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. New York: Russell Sage Foundation, 459-489.

Horowitz, John and Kenneth McConnell (2002). "A Review of WTA-WTP Studies." *Journal of Environmental Economics and Management*, **44** (3), 426-447

Kahneman, D. and Frederick, S. 2002. Representativeness revisited: Attribute substitution in intuitive judgment. pp. 49-81 in T. Gilovich, D. Griffin, and D. Kahneman [eds]. Heuristics & Biases: The Psychology of Intuitive Judgment. New York. Cambridge University Press.

Kant, Immanuel (1785). *Groundwork of the Metaphysics of Morals*, translated and edited by Mary Gregor. New York: Cambridge University Press, 1997.

Kirby K (1997) Bidding on the future: evidence against normative discounting of delayed rewards. Journal of Experimental Psychology: General 126:54-70.

Knoch, Daria, Alvaro Pascual-Leone, Kaspar Meyer, Valerie Treyer, and Ernst Fehr (2006). "Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex." Science 314 (3 November 2006), 912-915.

Kogut, Tehila and Ilana Ritov (2003). "The 'Identified Victim' Effect: An Identified Group, or Just a Single Individual?" Paper presented at the 19th Biannual Conference on Subjective Probability, Utility, and Decision-Making, Zurich, Switzerland, 25–27 August 2003.

Kunda, Ziva (1990). "The Case for Motivated Reasoning." *Psychological Bulletin*, **108**, 480-498.

Laibson, David (1997). "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, **112**, 443-477.

LeDoux, Joseph E. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York, NY: Simon and Schuster.

Lerner, J. S. and Keltner, D. (2000). Beyond valence: Toward a model of emotion-specific influences on judgment and choice. Cognition and Emotion, 14, 473-493.

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. Journal of Personality and Social Psychology. 81, 146-159.

Lerner, Jennifer S., Deborah A. Small, and George Loewenstein (2004). "Heart Strings and Purse Strings: Carry-Over Effects of Emotions on Economic Transactions." *Psychological Science*.

Lewin, Kurt (1951), Field Theory in Social Science. New York: Harper & Borthers.

Lhermitte, F. (1986). "Human Autonomy and the Frontal Lobes. 2. Patient Behavior in Complex and Social Situations — the Environmental Dependency Syndrome." *Annals of Neurology*, **19** (4), 335-343.

Loewenstein, George (1987). "Anticipation and the Valuation of Delayed Consumption." *Economic Journal* **97**, 666-684.

Loewenstein, George (1996). "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, **65,** 272-292.

Loewenstein, George and Erik Angner (2003). "Predicting and Indulging Changing Preferences," in George Loewenstein, Daniel Read, and Roy F. Baumeister, eds., *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. New York: Russell Sage Foundation, 351-391.

Loewenstein, George and Jennifer Lerner (2003). "The Role of Emotion in Decision Making," in R. J. Davidson, H. H Goldsmith, and K. R. Scherer, eds., *Handbook of Affective Science*. Oxford, England: Oxford University Press.

Loewenstein, George and Deborah A. Small (2007). "The Scarecrow and the Tin Man: The Vicissitudes of Human Sympathy and Caring." Review of General Psychology, forthcoming.

Loewenstein, G., Thompson, L., & Bazerman, M. (1989). Social utility and decision making in interpersonal contexts. Journal of Personality and Social Psychology, 57, 426-441.

Loewenstein, G., Weber, E., Hsee, C. & Welch, N. (2001). Risk as Feelings. Psychological Bulletin. 127, 267–286.

Lykken, David T. (1995). *The Antisocial Personalities*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

MacLean, Paul D. (1990). *The Triune Brain in Evolution: Role in Paleocerebral Function*. New York: Plenum.

Maia, Tiago V. and James L. McClelland A reexamination of the evidence for the somatic marker hypothesis: What participants really know in the Iowa gambling task Proceedings of the National Academy of Sciences PNAS 2004;101;16075-16080;

Manuck, Stephen B., Janine D. Flory, Matthew F. Muldoon, and Robert E. Ferrell (2003). "A Neurobiology of Intertemporal Choice," in George Loewenstein, Daniel Read and Roy F. Baumeister, eds., *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. New York: Russell Sage Foundation, 139-172.

Masserman, Jules Hyman, Stanley Wechkin, and William Terris (1964). "'Altruistic' Behavior in Rhesus Monkeys." *American Journal of Psychiatry* **121**, 584-585.

Massey, Douglas S. (2002) "A Brief History of Human Society: The Origin and Role of Emotion in Social Life." *American Sociological Review*, **67**, (1), 1-29.

McClure, Samuel M., David Laibson, George Loewenstein, and Jonathan D. Cohen (2004). "Separate Neural Systems Value Immediate and Delayed Monetary Rewards." *Science*, **306**(Oct 15), 503-507.

McClure, Samuel M. Keith M. Ericson, David I. Laibson, George Loewenstein, and Jonathan D. Cohen. (2006) "Time Discounting for Primary Rewards." Working Paper, Center for the Study of Brain, Mind, & Behavior and Department of Psychology, Princeton University, Princeton, NJ.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. Psychological Science, 8(6), 423-429.

Messick, D. M. & Sentis, K.P. (1985). Estimating social and nonsocial utility functions from ordinal data. European Journal of Social Psychology, 15(4): 389-399.

Metcalfe, Janet and Walter Mischel (1999). "A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower." *Psychological Review* **106**(1), 3-19.

Mischel, Walter, Ebbe B. Ebbesen, and Antonette Zeiss (1972). "Cognitive and Attentional Mechanisms in Delay of Gratification." *Journal of Personality and Social Pychology,* **21**(2), 204-218.

Mischel, Walter, Ozlem Ayduk, and Rodolfo Mendoza-Denton (2003). "Sustaining Delay of Gratification over Time: A Hot-Cool Systems Perspective," in George Loewenstein, Daniel Read and Roy F. Baumeister, eds., *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. New York: Russell Sage Foundation, 175-200.

Mischel, Walter, Yuichi Shoda, and Monica L. Rodriguez (1989). "Delay of Gratification in Children." *Science*, **244**(4907), 933-938.

Monat, Alan, James R. Averill, and Richard S. Lazarus (1972). "Anticipatory Stress and Coping Reactions Under Various Conditions of Uncertainty." *Journal of Personality and Social Psychology* **24**, 237-253.

Muraven, Mark (1998). *Mechanisms of Self-Control Failure: Motivation and Limited Resource*. Ph.D. diss., Case Western Reserve University.

Muraven, Mark, Roy F. Baumeister, and Dianne M. Tice (1999). "Longitudinal Improvement of Self-Regulation Through Practice: Building Self-Control Strength Through Repeated Exercise." *Journal of Social Psychology* **139,** 446-57.

Nolen-Hoeksema, Susan (1990). *Sex Differences in Depression*. Stanford, CA: Stanford University Press.

O'Donoghue, Ted and Matthew Rabin (1999). "Doing it Now or Later." American Economic Review, 89, 103-124.

Ozdenoren, Emre, Steve Salant, & Dan Silverman (2006). "Willpower and the Optimal Control of Visceral Urges." Mimeo, University of Michigan.

Panksepp, J. (1998). Affective Neuroscience. Oxford, Oxford University Press.

Peters, E., & Slovic, P. (2000). The springs of action: Affective and analytical information processing in choice. Personality & Social Psychology Bulletin, 26(12), 1465-1475.

Pham, M. T. (1998). Representativeness, relevance, and the use of feelings in decision making. Journal of Consumer Research, 25(2), 144-159.

Plato, Republic, translated by G.M.A. Grube, revised by C.D.C. Reeve. Indianapolis: Hackett Publishing Company, 1992.

Polivy, Janet, C. Peter Herman, and T. McFarlane (1994). "Effects of anxiety on eating: Does palatability moderate distress-induced overeating in dieters?" *Journal of Abnormal Psychology*, **103**, 505-510.

Polivy, Janet and C. Peter Herman (1999). "Distress and eating: Why do dieters overeat?" *International Journal of Eating Disorders*, **26**, 153-164.

Prelec (1998). "The Probability Weighting Function." Econometrica 66: 497-527.

Prelec, Drazen, George Loewenstein, and Ofer Zellermayer (1997). "Closet Tightwads: Compulsive Reluctance to Spend." Paper presented at the Association for Consumer Research Conference, Denver.

Preston, Stephanie D. and Frans B. M. de Waal (2002). "Empathy: Its Ultimate and Proximate Bases." *Behavioral and Brain Sciences*, **25**(1), 1-71.

Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics," American Economic Review, 83, 1281-1302.

Rabin, M. (2000). "Risk aversion and expected-utility theory: A calibration theorem." Econometrica 68(5): 1281-1292.

Rabin, Matthew and Richard Thaler (2001). "Anomalies: Risk aversion." Journal of Economic Perspectives, 15, 219-32.

Raghunathan, R., & Pham, M. T. (1999). All negative moods are not equal: Motivational influences of anxiety and sadness on decision making. Organizational Behavior & Human Decision Processes, 79(1), 56-77.

Rice, George E. and Priscilla Gainer (1962). "'Altruism' in the Albino Rat." *Journal of Comparative and Physiological Psychology*, **55**(1), 123-125.

Rick, S., and Loewenstein, G. (forthcoming). The Role of Emotion in Economic Behavior. In Lewis, M., Haviland-Jones, J. M., & Barrett, L. F. (Eds.). The handbook of emotion, 3rd Edition. New York: Guilford.

Rolls, E. T. (1999). The brain and emotion. New York, Oxford University Press.

Roth, Walton T., Gunnar Breivik, Per Erik Jorgensen, and Stefan Hofmann (1996). "Activation in Novice and Expert Parachutists While Jumping." *Psychophysiology* **33**, 63-72.

Rottenstreich, Yuval and Christopher K. Hsee (2001). "Money, Kisses, and Electric Shocks: On the Affective Psychology of Risk." *Psychological Science*, **12**, 185-190.

Sapolsky, Robert M. (1994). *Why Zebras Don't Get Ulcers*. New York: W. H. Freeman and Co.

Schelling, Thomas C. (1968). “The Life You Save may be Your Own,” in S. B. Chase, ed., *Problems in Public Expenditure Analysis.* Washington, DC: The Brookings Institute.

Shallice, Tim and Paul Burgess (1998). “The Domain of Supervisory Processes and the Temporal Organization of Behaviour,” in A. C. Roberts, T. W. Robbins and L. Weiskrantz, eds., *The Prefrontal Cortex: Executive and Cognitive Functions*. Oxford University Press.

Shefrin, Hersh M. and Richard H. Thaler (1988). “The Behavioral Life-Cycle Hypothesis.” *Economic Inquiry* **26**, 609-643.

Shiffman, Saul and Andrew J. Waters (2004). “Negative Affect and Smoking Lapses: A Prospective Analysis.” *Journal of Consulting and Clinical Psychology*, **72**(2), 192-201.

Shiv, Baba and Alexander Fedorikhin (1999). “Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making.” *Journal of Consumer Research*, **26**, 278-292.

Shiv, Baba, George Loewenstein, Antoine Bechara, Hanna Damasio, and Antonio Damasio (2003). “Investment Behavior and the Dark Side of Emotion.” Mimeo, University of Iowa.

Simon, H. A. (1967). Motivational and Emotional Controls of Cognition. Psychological Review, 74(1), 29-39.

Skitka, Linda J., Elizabeth Mullen, Thomas Griffin, S. Hutchinson, and B. Chamberlin (2002). “Dispositions, Ideological Scripts, or Motivated Correction? Understanding Ideological Differences in Attributions for Social Problems.” *Journal of Personality and Social Psychology*, **83**, 470-487.

Sloman, S.A. (1996). The empirical case for two systems of reasoning. Psychological Bulletin, 119, 3-22.

Slovic, Paul (forthcoming). “If I Look at the Mass I Will Never Act: Psychic Numbing and Genocide" *American Psychologist*.

Slovic, Paul, Melissa Finucane, Ellen Peters, and Donald G. MacGregor (2002). “The Affect Heuristic.” In T. Gilovich, D. Griffin, & D. Kahneman, (Eds.), Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge University Press, pp. 397-420.

Small, Deborah A. and George Loewenstein (2003). “Helping *The* Victim or Helping *a* Victim: Altruism and Identifiability.” *Journal of Risk and Uncertainty*, **26**(1), 5-16.

Small, Deborah. A, George Loewenstein, and Paul Slovic (2007). “Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims.” Organization Behavior and Human Decision Processes, 102, 143-153.

Smith, Adam (1759). *The Theory of Moral Sentiments*. Amherst, NY: Prometheus Books, 2000.

Snortum, John R. and Frederick W. Wilding (1971). "Temporal Estimation and Heart Rate as a Function of Repression-Sensitization Score and Probability of Shock." *Journal of Consulting and Clinical Psycholog*y, **37**(3), 417–422.

Spence, Sean A. and Chris D. Frith (1999). "Towards a Functional Anatomy of Volition." *Journal of Consciousness Studies*, **6**(8-9, Aug-Sep 1999), 11-29.

Starmer, Christopher (2000). "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk." *Journal Of Economic Literature*, **38**, 332-382.

Stevens, Jeffrey R., Elizabeth V. Hallinan, and Marc D. Hauser (2005). "The Ecology and Evolution of Patience in Two New World Monkeys." Biology Letters, 1, 223-226.

Strack, Fritz and Roland Deutsch (2004). "Reflective and Impulsive Determinants of Social Behavior." Personality and Social Psychology Review, 8(3), 220-247.

Sunstein, Cass R. (2002). "Probability Neglect: Emotions, Worst Cases, and Law." *Yale Law Journal*, **112**(1), 61-108.

Sunstein, Cass R. (2003). "Terrorism and Probability Neglect." *Journal of Risk and Uncertainty*, ***26***, 121-136.

Tesser, A., Martin, L. & Mendolia, M. (1995). The impact of thought on attitude extremity and attitude-behavior consistency. In R. E. Petty & J. A. Krosnick (Eds.) Attitude Strength: Antecedents and Consequences. Hillsdale, NJ: Lawrence Erlbaum (p. 73-92).

Thaler, Richard H. (1980). "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, **1**, 39-60.

Thaler, R. H. (1981). "Some empirical evidence on dynamic inconsistency." Economic Letters 8: 201-207.

Thaler, Richard H. and Hersh M. Shefrin (1981). "An Economic Theory of Self-Control." *Journal of Political Economy*, 89(2), 392-406.

Thaler, Richard H. & Sunstein, Cass R., 2003. Libertarian Paternalism. The *American Economic Review* 93 (May): 175-179.

Tobin, Henry, A.W. Logue, John J. Chelonis, and Kimberly T. Ackerman (1996). "Self-Control in the Monkey Macaca Fascicularis." Animal Learning & Behavior, 24(2), 168-174.

Tom, Sabrina M., Craig R. Fox, Christopher Trepel, and Russell A. Poldrack (2007). “The Heural Basis of Loss Aversion in Decision-Making Under Risk.” Science, 315 (1/26/07), 515-518.

Trope, Yaacov and Nira Liberman (2003), "Temporal construal," *Psychological Review*, 110 (3), 403-21.

Tversky, Amos and Daniel Kahneman (1991). “Loss Aversion in Riskless Choice: A Reference Dependent Model.” *Quarterly Journal of Economics*, **106**, 1039-1061.

Van Boven, Leaf and George Loewenstein (2003). “Social Projection of Transient Visceral Feelings.” *Personality and Social Psychology Bulletin*, **29**(09), 1159-1168.

Van Boven, Leaf, George Loewenstein, Edward Welch, and David Dunning (2002). “The Illusion of Courage: Underestimating the Impact of Fear of Embarrassment on the Self.” Working paper, Dept. of Social and Decision Sciences, CMU.

Vohs, Kathleen D. and Ronald J. Faber (2007). “Spent Resources: Self-Regulatory Resource Availability Affects Impulse Buying.” Journal of Consumer Research, 33, 537-547.

Vohs, Kathleen D. and Todd F. Heatherton (2000). “Self-Regulatory Failure: A Resource-Depletion Approach.” Psychological Science, 11(3), 249-254.

Wegner, Daniel M. and Thalia Wheatley (1999). “Apparent Mental Causation: Sources of the Experience of Will.” *American Psychologist*, **54**(7): 480-492.

Wilson M, Daly M (2003) Do pretty women inspire men to discount the future? Biology Letters Proceedings: Biological Sciences, 4, 177-179.

Wilson, Timothy. D., D. J. Lisle, Jonathan W. Schooler, Sara D. Hodges, K.J. Klaaren, and Suzanne J. LaFleur (1993). “Introspecting about Reasons can Reduce Post-Choice Satisfaction.” *Personality and Social Psychology Bulletin*, **19**, 331-339.

Wilson, Timothy D. and Jonathan W. Schooler (1991). “Thinking Too Much: Introspection can Reduce the Quality of Preferences and Decisions.” *Journal of Personality and Social Psychology*, **60**, 181-192.

Wilson, Timothy D., Samuel Lindsey, and Tonya Y. Schooler (2000), "A model of dual attitudes," *Psychological Review*, 107 (1), 101-26.

Winkielman, Piotr and Kent C. Berridge (2004). “Unconscious Emotion.” *Current Directions in Psychological Science*, **13**(3), 120-123.

Wright, W. F., & Bower, G. H. (1992). Mood effects on subjective probability assessment. *Organizational Behavior and Human Decision Processes*, *52*, 276-291.

Yaari, Menachem and Maya Bar-Hillel (1984). “On Dividing Justly.” *Social Choice and Welfare*, **1**, 1-24.

Zajonc, Robert B. (1980). “Feeling and Thinking: Preferences Need No Inferences.” *American Psychologist*, **35**(2), 151-175.

Zajonc, Robert B. (1984). “On the Primacy of Affect.” *American Psychologist*, **39**(2), 117-123.

Zald, David H. and Pardo, José V. (1997) Emotion, olfaction, and the human amygdala: Amygdala activation during aversive olfactory stimulation. Proceedings of the National Academy of Sciences.  94, 4119-4124.

**ENDNOTES**

[1] Affect and deliberation are probably better thought of as two different types of processes, or modes of information processing. We use the "systems" terminology because it is more natural to think of systems interacting and competing than processes.
[2] In the Republic, for example, Plato contrasts the immediacy of desires against the broader scope of reason, whose function in the human soul is to “rule with wisdom and forethought on behalf of the entire soul” (Plato, Republic 441e).
[3] For a more psychologically detailed dual-system model of behavior, see Strack & Deutsch (2004).
[4] In one clever study that illustrates the role of the prefrontal cortex in deliberative behavior, Chris Frith and colleagues (see Spence & Frith, 1999) scanned subjects' brains while they moved a finger they had been instructed to move in response to a noise. The brain scan revealed activation in the auditory cortex (which does crude processing of sounds) and the motor cortex (the area that controls movement). To localize where free-willed activity happens in the brain, they then added a new component to the task; instead of telling the subjects which finger to lift they left it to them to decide which one to move. With the addition of this new aspect to the task, the prefrontal cortex became activated as well.
[5] An assumption we think is partly supported by the fact that the rational-choice model was conceived of by the deliberative systems of scholars.
[6] Buck (1984) refers to these latter influences as “biological affects,” which he distinguishes from the more traditional “social affects.” Although emotions such as anger and fear might seem qualitatively different than the biological affects, they have more in common that might be supposed. Thus, for example, a recent study showed that hurt feelings activated the same brain regions as would broken bones or other physical injuries (Eisenberger et al 2003). The researchers scanned the brains of subjects using fMRI (functional magnetic resonance imaging) as they played a video game designed to produce a feeling of social rejection. The social snub triggered nerve activity in a part of the brain called the anterior cingulate cortex, which also processes physical pain.
[7] Our framework also bears a resemblance to Freud’s (1924) distinction between the id and the ego. The id, which represents biological forces and is governed by the “pleasure principle,” is similar to our affective system. The ego, governed by the “reality principle,” is similar to our deliberative system.
[8] The passage of time is also important retrospectively. Just as the affective system tends to react disproportionately to imminent outcomes, it also reacts especially strongly to events that have happened in the very recent past. Hence, in the immediate aftermath of an emotion-evoking situation, people often lose perspective and overreact to things that are unimportant, but as events recede into the past, the passions of the moment diminish and people gain a more evenhanded perspective.
[9] However, for recent critiques of these findings, see Maia & McClelland (2004) and Dunn, Dalgleish, & Lawrence (2005).
[10] This may be especially true for decisions that are affective in quality (see, e.g., Tesser, Martin, & Mendolia, 1995).
[11] Affective reactions can also redirect cognitive processes, as first proposed by Simon (1967), who posited that emotions serve as "cognitive interrupts" that prioritize the application of processing resources (see, more recently, Armony et al, 1995; DeBecker, 1997).

[12] Under the natural assumption that the function *M* is concave in *x*, the willpower required to shift behavior is an accelerating function of the magnitude of the departure.

[13] More generally, if a person has been forced to repeatedly make choices, and therefore repeatedly expend willpower, her current behavior will be further from the deliberative optimum. Hence, the frequency of choice can play a critical role in people's behavior — e.g., addictions to substances that are readily available, such as cigarettes, alcohol, and coffee, are especially hard to break.

[14] Also motivated by Baumeister's research, Ozdenoren, Salant, & Silverman (2006) propose an economic model in which a person's behavior is influenced by the need to manage a stock of willpower. Specifically, they assume that a person has some initial stock of willpower, and that to control one's behavior one must use this willpower. Unlike our model, however, they assume that willpower usage has no direct impact on the person's well-being, but rather it is merely a constraint (much like a budget constraint).

[15] Throughout, we assume that the model is well behaved in the sense that the functions *U*, *M*, and *V* all have a unique local maximum in *x*. This property holds if we assume that the functions *U* and *M* are concave in *x* and that $h(W,\sigma)$ is always positive. Weaker assumptions can also work.

[16] Subsequent research (McClure et al, 2006) replicates these findings with a primary reward – juice and/or water delivered to thirsty subjects.

[17] More generally, the affective system might deviate from this simple form in two ways. First, there might be a satiation point --- after eating enough ice cream, even the affective system eventually finds more ice cream unpleasant. Second, the affective system might have some concern for future payoffs, although still less concern than the deliberative system. As will become clear, the former is not important for our results; but see the next footnote for more on the latter.

[18] In fact, this conclusion does not rely on the affective system being completely myopic. For instance, if the affective system also discounts exponentially, but by more than the deliberative system (as discussed by McClure et al 2006), our model still generates that people will be more impatient when facing now vs. near-future trade-offs than they will be when facing future vs. further-future trade-offs. Moreover, our predictions below also continue to hold.

[19] This possibility was first suggested by Loewenstein et al (2001).

[20] Although our assumption that the deliberative system is risk-neutral is mostly for simplicity --- our qualitative conclusions hold even if it is risk-averse --- we are in part motivated Rabin (2000), who demonstrates that, if we take expected utility seriously as a coherent theory of risk preferences, then we must assume approximate risk neutrality for small- and even moderate-stakes gambles.

[21] Inevitably, this formulation is an extreme simplification, as the influence of affect on risky decision making is certainly more nuanced than can be captured by a simple term. The limitation of our approach is highlighted by the burgeoning literature on the impact of specific emotions on risk-taking, which finds that different types of emotions exert qualitatively and quantitatively different types of influences on risk perception and risk-taking (e.g., Raghunathan & Pham, 1999; Lerner & Keltner, 2001).

[22] In fact, these qualitative conclusions hold for any S-shaped probability-weighting function that the affective system might have, although the third requires that, within the affective system, an increase in the intensity of affective motivations leads to increased insensitivity to probabilities.

[23] Such a temporal pattern of fear is highly adaptive; an organism that felt similar fear toward distant and immediate risks would be unlikely to survive long in a hostile environment. Indeed, one of the characteristics of certain types of stress disorders that clinical psychologists treat is the tendency to ruminate over risks that are remote in time (e.g., Nolen-Hoechsema, 1990; Sapolsky, 1994) or to continue to experience fear toward no-longer threatening events that happened in the past (e.g., Barlow, 1988).

[24] The example of chickening out illustrates how a dynamic inconsistency can arise in risk preferences due to changing affective states over time. Here, as temporal proximity changes, so do risk attitudes.

[25] For a similar account, see Loewenstein & Small (2007).

[26] Further information about the interactions between the affective and deliberative systems in altruistic behavior comes from considering certain abnormal populations. For instance, psychopaths and

sociopaths, who tend not experience empathy, are purely Machiavellian and self-interested (Cleckley, 1976; Lykken, 1995).

[27] Identifiability is important for other social emotions besides altruism. For instance, identified perpetrators of criminal acts also elicit more intense emotional reactions than unidentified, statistical, perpetrators.

[28] In Berridge's (1995) research on multiple reward systems in the brain (on wanting vs. liking), he argues that people are most likely not consciously aware of what generates pleasure. If so, it seems we ought to be even less willing to take only the deliberative system into account when measuring well-being.

[29] Again, see Ozdenoren, Salant, & Silverman (2006) for a model that focuses on willpower management.

[30] There are a number of studies in which subjects are "manipulated" into behaving in certain ways and then asked to explain that behavior, and people invariably come up with plausible reasons for why the behavior was purposeful (see for instance Brasil-Neto et al, 1992).

[31] In a casual survey of visitors to an airport, Prelec, Loewenstein, and Zellermayer (1997) found that the majority of people perceived that spending too little rather than spending too much was their greater problem. Similarly, from survey data of TIAA-CREF participants, Ameriks et al (2003) find that, while many people perceive themselves to have a problem of over-consumption, many other people perceive their problem to be under-consumption.