Neuroeconomics: How Neuroscience Can Inform Economics

COLIN CAMERER, GEORGE LOEWENSTEIN, *and* DRAZEN PRELEC*

Who knows what I want to do? Who knows what anyone wants to do? How can you be sure about something like that? Isn't it all a question of brain chemistry, signals going back and forth, electrical energy in the cortex? How do you know whether something is really what you want to do or just some kind of nerve impulse in the brain. Some minor little activity takes place somewhere in this unimportant place in one of the brain hemispheres and suddenly I want to go to Montana or I don't want to go to Montana. (*White Noise*, Don DeLillo)

1. *Introduction*
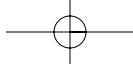
In the last two decades, following almost a century of separation, economics has begun to import insights from psychology. "Behavioral economics" is now a prominent fixture on the intellectual landscape and has spawned applications to topics in economics,

such as finance, game theory, labor economics, public finance, law, and macroeconomics (see Colin Camerer and George Loewenstein 2004). Behavioral economics has mostly been informed by a branch of psychology called "behavioral decision research," but other cognitive sciences are ripe for harvest. Some important insights will surely come from neuroscience, either directly or because neuroscience will reshape what is believed about psychology which in turn informs economics.

Neuroscience uses imaging of brain activity and other techniques to infer details about how the brain works. The brain is the ultimate "black box." The foundations of economic theory were constructed assuming that details about the functioning of the brain's black box would not be known. This pessimism was expressed by William Jevons in 1871:

> I hesitate to say that men will ever have the means of measuring directly the feelings of the human heart. It is from the quantitative effects of the feelings that we must estimate their comparative amounts.

Since feelings were meant to predict behavior but could only be assessed from behavior, economists realized that, without direct measurement, feelings were useless intervening constructs. In the 1940s, the concepts of ordinal utility and revealed preference eliminated the superfluous intermediate step of positing immeasurable feelings. Revealed preference theory simply equates unobserved preferences with observed choices. Circularity is avoided by assuming that people behave consistently, which makes the theory falsifiable; once they have revealed that they prefer A to B, people should not subsequently choose B over A. Later extensions—discounted, expected, and subjective expected utility, and Bayesian updating—provided similar "as if" tools which sidestepped psychological detail. The "as if" approach made good sense as long as the brain remained substantially a black box. The development of economics could not be held hostage to progress in other human sciences.
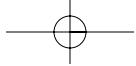
But now neuroscience has proved Jevons's pessimistic prediction wrong; the study of the brain and nervous system *is* beginning to allow direct measurement of thoughts and feelings. These measurements are, in turn, challenging our understanding of the relation between mind and action, leading to new theoretical constructs and calling old ones into question. How can the new findings of neuroscience, and the theories they have spawned, inform an economic theory that developed so impressively in their absence?

In thinking about the ways that neuroscience can inform economics, it is useful to distinguish two types of contributions, which we term *incremental* and *radical* approaches. In the incremental approach, neuroscience adds variables to conventional accounts of decision making or suggests specific functional forms to replace "as if" assumptions that have never been well supported empirically. For example, research on the neurobiology of addiction suggests how drug consumption limits pleasure from future consumption of other goods (dynamic cross-partial effects in utility for commodity bundles) and how environmental cues trigger unpleasant craving and increase demand. These effects can be approximated by extending standard theory and then applying conventional tools (see Douglas Bernheim and Antonio Rangel 2004; David Laibson 2001; Ted O'Donoghue and Matthew Rabin 1997).

The radical approach involves turning back the hands of time and asking how economics might have evolved differently if it had been informed from the start by insights and findings now available from neuroscience. Neuroscience, we will argue, points to an entirely new set of constructs to underlie economic decision making. The standard economic theory of constrained utility maximization is most naturally interpreted either as the result of learning based on consumption experiences (which is of little help when prices, income, and opportunity sets change), or careful deliberation—a balancing of the costs and benefits of different options—as might characterize complex decisions like planning for retirement, buying a house, or hammering out a contract. Although economists may privately acknowledge that actual flesh-and-blood human beings often choose without much deliberation, the economic models as written invariably represent decisions in a "deliberative equilibrium," i.e., that are at a stage where further deliberation, computation, reflection, etc. would not by itself alter the agent's choice. The variables that enter into the formulation of the decision problem—the preferences, information, and constraints—are precisely the variables that should affect the decision, if the person had unlimited time and computing ability.

While not denying that deliberation is part of human decision making, neuroscience points out two generic inadequacies of this approach—its inability to handle the crucial roles of automatic and emotional processing.

First, much of the brain implements "automatic" processes, which are faster than conscious deliberations and which occur with little or no awareness or feeling of effort (John Bargh et al. 1996; Bargh and Tanya Chartrand 1999; Walter Schneider and Richard Shiffrin 1977; Shiffrin and Schneider 1977). Because people have little or no introspective access to these processes, or volitional control over them, and these processes were evolved to solve problems of evolutionary importance rather than respect logical dicta, the behavior these processes generate need not follow normative axioms of inference and choice.

Second, our behavior is strongly influenced by finely tuned affective (emotion) systems whose basic design is common to humans and many animals (Joseph LeDoux 1996; Jaak Panksepp 1998; Edmund Rolls 1999). These systems are essential for daily functioning, and when they are damaged or perturbed, by brain injury, stress, imbalances in neurotransmitters, or the "heat of the moment," the logical-deliberative system—even if completely intact—cannot regulate behavior appropriately.

Human behavior thus requires a fluid interaction between controlled and automatic processes, and between cognitive and affective systems. However, many behaviors that emerge from this interplay are routinely and falsely interpreted as being the product of cognitive deliberation alone (George Wolford, Michael Miller, and Michael Gazzaniga 2000). These results (some of which are described below) suggest that introspective accounts of the basis for choice should be taken with a grain of salt. Because automatic processes are designed to keep behavior "off-line" and below consciousness, we have far more introspective access to controlled than to automatic processes. Since we see only the top of the automatic iceberg, we naturally tend to exaggerate the importance of control.
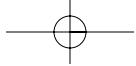
Neuroscience findings and methods will undoubtedly play an increasingly prominent role in economics and other social sciences (e.g., law, Terrence Chorvat, Kevin McCabe, and Vernon Smith 2004). Indeed, a new area of economics that has been branded "neuroeconomics" has already formed the basis of numerous academic gatherings that have brought neuroscientists and economists together (see also Paul Zak, in press).[1] Participating in the development of a shared intellectual enterprise will help us ensure that the neuroscience informs economic questions we care about. Stimulated by the authors' own participation in a number of such meetings, our goal in this paper is to describe what neuroscientists do and how their discoveries and views of human behavior might inform economic analysis. In the next section (2), we describe the diversity of tools that neuroscientists use. Section 3 introduces a simplified account of how cognition and affect on the one hand, and automatic and controlled processes on the other, work separately, and interact. Section 4 discusses general implications for economics. Section 5 goes into greater detail about implications of neuroeconomics for four topics in economics: intertemporal choice, decision making under risk, game theory, and labor-market discrimination. Through most of the paper, our focus is largely on how neuroscience can inform models of microfoundations of individual decision making. Section 6 discusses some broader macro implications and concludes.

2. *Neuroscience Methods*

Scientific technologies are not just tools scientists use to explore areas of interest. New tools also *define* new scientific fields and erase old boundaries. The telescope created astronomy by elevating the science from

[1] The first meeting was held at Carnegie Mellon in 1997. Later meetings were held in Arizona and Princeton, in 2001, in Minnesota in 2002, and on Martha's Vineyard in 2003 and Kiawah Island in 2004. Occasional sessions devoted to this rapidly growing topic are now common at large annual meetings in both economics and neuroscience.

pure cosmological speculation. The microscope made possible similar advances in biology. The same is true of economics. Its boundaries have been constantly reshaped by tools such as mathematical, econometric, and simulation methods. Likewise, the current surge of interest in neuroscience by psychologists emerged largely from new methods, and the methods may productively blur the boundaries of economics and psychology. This section reviews some of these methods.

2.1 *Brain Imaging*

Brain imaging is currently the most popular neuroscientific tool. Most brain imaging involves a comparison of people performing different tasks—an "experimental" task and a "control" task. The difference between images taken while subject is performing the two tasks provides a picture of regions of the brain that are differentially activated by the experimental task.

There are three basic imaging methods. The oldest, electro-encephalogram (or EEG) uses electrodes attached to the scalp to measure electrical activity synchronized to stimulus events or behavioral responses (known as Event Related Potentials or ERPs). Like EEG, positron emission topography (PET) scanning is an old technique in the rapidly changing time-frame of neuroscience, but is still a useful technique. PET measures blood flow in the brain, which is a reasonable proxy for neural activity, since neural activity in a region leads to increased blood flow to that region. The newest, and currently most popular, imaging method is functional magnetic resonance imaging (fMRI), which tracks blood flow in the brain using changes in magnetic properties due to blood oxygenation (the "BOLD signal"). Simultaneous direct recording of neural processing and fMRI responses confirms that the BOLD signal reflects input to neurons and their processing (Nikos Logothetis et al. 2001).
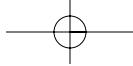
Although fMRI is increasingly becoming the method of choice, each method has pros and cons. EEG has excellent temporal resolution (on the order of one millisecond) and is the only method used with humans that directly monitors neural activity, as opposed to, e.g., blood flow. But spatial resolution is poor and it can only measure activity in the outer part of the brain. EEG resolution has, however, been improving through the use of ever-increasing numbers of electrodes. Interpolation methods, and the combined use of EEG and fMRI for measuring outer-brain signals and inner-brain signals at the same time, promise to create statistical methods which make reasonable inferences about activity throughout the brain from EEG signals. For economics, a major advantage of EEG is its relative unobtrusiveness and portability. Portability will eventually reach the point where it will be possible to take unobtrusive measurements from people as they go about their daily affairs. PET and fMRI provide better spatial resolution than EEG, but poorer temporal resolution because blood-flow to neurally active areas occurs with a stochastic lag from a few seconds (fMRI) to a minute (PET).

Brain imaging still provides only a crude snapshot of brain activity. Neural processes are thought to occur on a 0.1 millimeter scale in 100 milliseconds (msec), but the spatial and temporal resolution of a typical scanner is only 3 millimeters and several seconds. Multiple trials per subject can be averaged to form composite images, but doing so constrains experimental design. However, the technology has improved rapidly and will continue to improve. Hybrid techniques that combine the strengths of different methods are particularly promising. Techniques for simultaneously scanning multiple brains ("hyperscanning") have also been developed, which can be used to study multiple-brain-level differences in activity in games and markets (Read Montague et al. 2002).

2.2 *Single-Neuron Measurement*

Even the finest-grained brain imaging techniques only measure activity of "circuits" consisting of thousands of neurons. In

single neuron measurement, tiny electrodes are inserted into the brain, each measuring a single neuron's firing. As we discuss below, *single* neuron measurement studies have produced some striking findings that, we believe, are relevant to economics. A limitation of single neuron measurement is that, because insertion of the wires damages neurons, it is largely restricted to animals.

Studying animals is informative about humans because many brain structures and functions of non-human mammals are similar to those of humans (e.g., we are more genetically similar to many species of monkeys than those species are to other species). Neuroscientists commonly divide the brain into crude regions that reflect a combination of evolutionary development, functions, and physiology. The most common, *triune* division draws a distinction between the "reptilian brain," which is responsible for basic survival functions, such as breathing, sleeping, eating, the "mammalian brain," which encompasses neural units associated with social emotions, and the "hominid" brain, which is unique to humans and includes much of our oversized cortex—the thin, folded, layer covering the brain that is responsible for such "higher" functions as language, consciousness and long-term planning (Paul MacLean 1990). Because single neuron measurement is largely restricted to nonhuman animals, it has so far shed far more light on the basic emotional and motivational processes that humans share with other mammals than on higher-level processes such as language and consciousness.

2.3 *Electrical Brain Stimulation (EBS)*
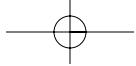
Electrical brain stimulation is another method that is largely restricted to animals. In 1954, two psychologists (James Olds and Peter Milner 1954) discovered that rats would learn and execute novel behaviors if rewarded by brief pulses of electrical brain stimulation (EBS) to certain sites in the brain. Rats (and many other vertebrates, including humans) will work hard for EBS.

For a big series of EBS pulses, rats will leap over hurdles, cross electrified grids, and forego their only daily opportunities to eat, drink, or mate. Animals also trade EBS off against smaller rewards in a sensible fashion—e.g., they demand more EBS to forego food when they are hungry. Unlike more naturalistic rewards, EBS does not satiate. And electrical brain stimulation at specific sites often elicits behaviors such as eating, drinking (Joseph Mendelson 1967), or copulation (Anthony Caggiula and Bartley Hoebel 1966). Many abused drugs, such as cocaine, amphetamine, heroin, cannabis, and nicotine, lower the threshold at which animals will lever-press for EBS (Roy Wise 1996). Despite its obvious applications to economics, only a few studies have explored substitutability of EBS and other reinforcers (Steven Hursh and B. H. Natelson; Leonard Green and Howard Rachlin 1991; Peter Shizgal 1999).

2.4 *Psychopathology and Brain Damage in Humans*

Chronic mental illnesses (e.g., schizophrenia), developmental disorders (e.g., autism), degenerative diseases of the nervous system, and accidents and strokes that damage localized brain regions help us understand how the brain works (e.g., Antonio Damasio 1994). When patients with known damage to an area X perform a special task more poorly than "normal" patients, and do other tasks equally well, one can infer that area X is used to do the special task. Patients who have undergone neurosurgical procedures such as lobotomy (used in the past to treat depression) or radical bisection of the brain (an extreme remedy for epilepsy, now rarely used) have also provided valuable data (see Walter Freeman and James Watts 1942; Gazzaniga and LeDoux 1978).

Finally, a relatively new method called transcranial magnetic stimulation (TMS) uses pulsed magnetic fields to *temporarily* disrupt brain function in specific regions. The difference in cognitive and behavioral

functioning that results from such disruptions provides clues about which regions control which neural functions. The theoretical advantage of TMS over brain imaging is that TMS directly leads to causal inferences about brain functioning rather than the purely associational evidence provided by imaging techniques. Unfortunately, the use of TMS is currently limited to the cortex (it is particularly useful for studying visual processes in the occipital lobe, in the back of the head). It is also controversial because it can cause seizures and may have other bad long-run effects.

2.5 *Psychophysical Measurement*

An old and simple technique is measurement of psychophysiological indicators like heart rate, blood pressure, galvanic skin response (GSR, sweating in the palms), and pupil dilation (pupils dilate in response to arousal, including monetary reward). These measurements are easy, not too obtrusive, portable, and very rapid in time. The drawback is that measurements can fluctuate for many reasons (e.g., body movement) and many different combinations of emotions lead to similar psychophysiological output, just as being pulled over by a cop and meeting a blind date may produce very similar emotional anxiety responses. Often these measurements are useful in combination with other techniques or in lesion patients who are likely to have very different physiological reactions (e.g., sociopaths do not show normal GSR fear reactions before a possible monetary loss). Facial musculature can also be measured by attaching small electrodes to smiling muscles (on the cheekbones) and frowning muscles (between the eyebrows).

2.6 *Diffusion Tensor Imaging (DTI)*
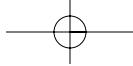
Diffusion tensor imaging (DTI) is a new technique which exploits the fact that water flows rapidly though myleinated (sheathed) neural axons (Dennis Bihan et al. 2001). Imaging the water flow can therefore reveal the trajectories which project from one neural region to others (like watching patterns of car traffic from a helicopter can tell you roughly the ebb and flow of economic and social activity). Knowing where a region's neurons project is extremely useful in understanding neural circuitry and is an important complement to simply imaging activity in multiple regions (using fMRI) with little ability to pin down which activity occurs earliest. Furthermore, the technique can be used after autopsies, which is an obvious advantage.

2.7 *Is Neuroscience Just About Where Things Happen in the Brain?*

Neuroscience is sometimes criticized as providing little more than a picture of "where things happen in the brain" or, more cynically, as simply showing that behavior is caused by action in the nervous system (which was never in doubt). Indeed, some neuroscientists who are purely interested in the functionality of different brain regions would endorse such a characterization unapologetically.

However, the long-run goal of neuroscience is to provide more than a map of the mind. By tracking what parts of the brain are activated by different tasks, and especially by looking for overlap between diverse tasks, neuroscientists are gaining an understanding of what different parts of the brain do, how the parts interact in "circuitry," and, hence, how the brain solves different types of problems. For example, because different parts of the brain are more or less associated with affective or cognitive processing (a distinction we define more precisely below), imaging people while they are doing different types of economic tasks provides important clues about the mix of affective and cognitive processes in those tasks. For example, Sanfey et al. (2003) find that the insula cortex—a region in the temporal lobe that encodes bodily sensations like pain and odor disgust—is active when people receive

low offers in an ultimatum bargaining game. This means that even if rejecting a low offer is done because of an adapted instinct to build up a reputation for toughness (in order to get more in the future), the circuitry that encodes this instinct clearly has an affective component, which is not purely cognitive.

For neuroeconomists, knowing more about functional specialization, and how regions collaborate in different tasks, could substitute familiar distinctions between categories of economic behavior (sometimes established arbitrarily by suggestions which become modeling conventions) with new ones grounded in neural detail. For example, the insula activity noted by Sanfey et al. in bargaining is also present when subjects playing matrix games are asked to guess what other subjects think *they* will do (second-order beliefs) (see Meghana Bhatt and Camerer, in press). This suggests that maybe the subjects who receive low offers in the ultimatum study are not disgusted, they are simply evaluating a second-order belief about what proposers expect them to do, as an input into an emotional evaluation. This is just a speculation, of course, but it shows how direct understanding of neural circuitry can inspire theorizing and the search for new data.

In light of the long list of methods reviewed earlier in this section, it is also worth emphasizing that neuroscience *isn't only about brain imaging*. Brain lesion studies (as well as TMS) allow one to examine the impact of disabling specific parts of the brain. They have often provided clearer evidence on functionality of specific brain regions—even with tiny sample sizes from rare types of damage—than imaging methods have. Single neuron measurement provides information not just about what parts of the brain "light up," but about the specific conditions that cause specific neurons to fire at accelerated or decelerated rates. And electrical brain stimulation, though it is largely off limits to studies involving

humans, provides a level of experimental control that economists conducting field research should envy.

As always, "the proof is in the pudding." Here, our goal is not to review the many ways in which neuroscience will rapidly change economic theory, because that is not where we are yet. Our goal is to showcase some key findings in neuroscience, and stimulate the reader's curiosity about what these findings might mean for economics.

3. *Basic Lessons from Neuroscience*

Because most of these techniques involve localization of brain activity, this can easily foster a misperception that neuroscience is merely developing a "geography of the brain," a map of which brain bits do what part of the job. If that were indeed so, then there would be little reason for economists to pay attention. In reality, however, neuroscience is beginning to use regional activity differences and other clues to elucidate the principles of brain organization and functioning, which in turn, is radically changing our understanding of how the brain works. Our goal in this section is to highlight some of the findings from neuroscience that may prove relevant to economics.

3.1 *A Two-Dimensional Theoretical Framework*

Our organizing theme, depicted in Table 1, emphasizes the two distinctions mentioned in the introduction, between *controlled* and *automatic* processes, and between *cognition* and *affect*.

3.1.1 *Automatic and Controlled Processes*

The distinction between automatic and controlled processes was first proposed by Schneider and Shiffrin (1977). Many others have developed similar two-system models since then, with different labels: rule-based and associative (Steven Sloman 1996); rational and experiential systems (Lee Kirkpatrick and Seymour Epstein 1992);
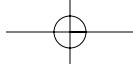
TABLE 1
TWO DIMENSIONS OF NEURAL FUNCTIONING

	Cognitive	Affective
Controlled Processes ■ serial ■ effortful ■ evoked deliberately ■ good introspective access	**I**	**II**
Automatic Processes ■ parallel ■ effortless ■ reflexive ■ no introspective access	**III**	**IV**

reflective and reflexive (Matthew Lieberman et al. 2002); deliberative and implementive systems (Peter Gollwitzer, Kentaro Fujita, and Gabriele Oettingen 2004); assessment and locomotion (Arie Kruglanski et al. 2000), and type I and type II processes (Daniel Kahneman and Shane Frederick 2002).

Controlled processes, as described by the two rows of Table 1, are serial (they use step-by-step logic or computations), tend to be invoked deliberately by the agent when her or she encounters a challenge or surprise (Reid Hastie 1984), and are often associated with a subjective feeling of effort. People can typically provide a good introspective account of controlled processes. Thus, if asked how they solved a math problem or choose a new car, they can often recall the considerations and the steps leading up to the choice.[2] Standard tools of economics, such as decision trees and dynamic programming, can be viewed as stylized representations of controlled processes.

Automatic processes are the opposite of controlled processes on each of these dimensions— they operate in parallel, are not accessible to consciousness, and are relatively effortless. Parallelism facilitates rapid response, allows for massive multitasking, and gives the brain remarkable power when it comes to certain types of tasks, such as visual identification. Parallelism also provides redundancy that decreases the brain's vulnerability to injury. When neurons are progressively destroyed in a region, the consequences are typically gradual rather than sudden ("graceful degradation").[3] "Connectionist" neural network models formulated by cognitive psychologists (David Rumelhart and James McClelland 1986) capture these features and have been applied to many domains,

---

[2] Elaborate methods have been developed that maximize the validity of such "verbal protocols" (see, e.g., Herbert Simon).

[3] The brain's ability to recover from environmental damage is also facilitated by a property called plasticity. In one study that illustrates the power of plasticity, the optic nerves of ferrets (which are born when they are still at a relatively immature state of development when the brain is still highly plastic) were disconnected at birth and reconnected to the auditory cortex (the portion of the brain that processes sound). The ferrets learned to "see" using auditory cortex, and some neurons in their auditory cortex actually took on the physical characteristics of neurons in the visual cortex (Lauire von Melchner, Sarah Pallas, and Mriganka Sur 2000).
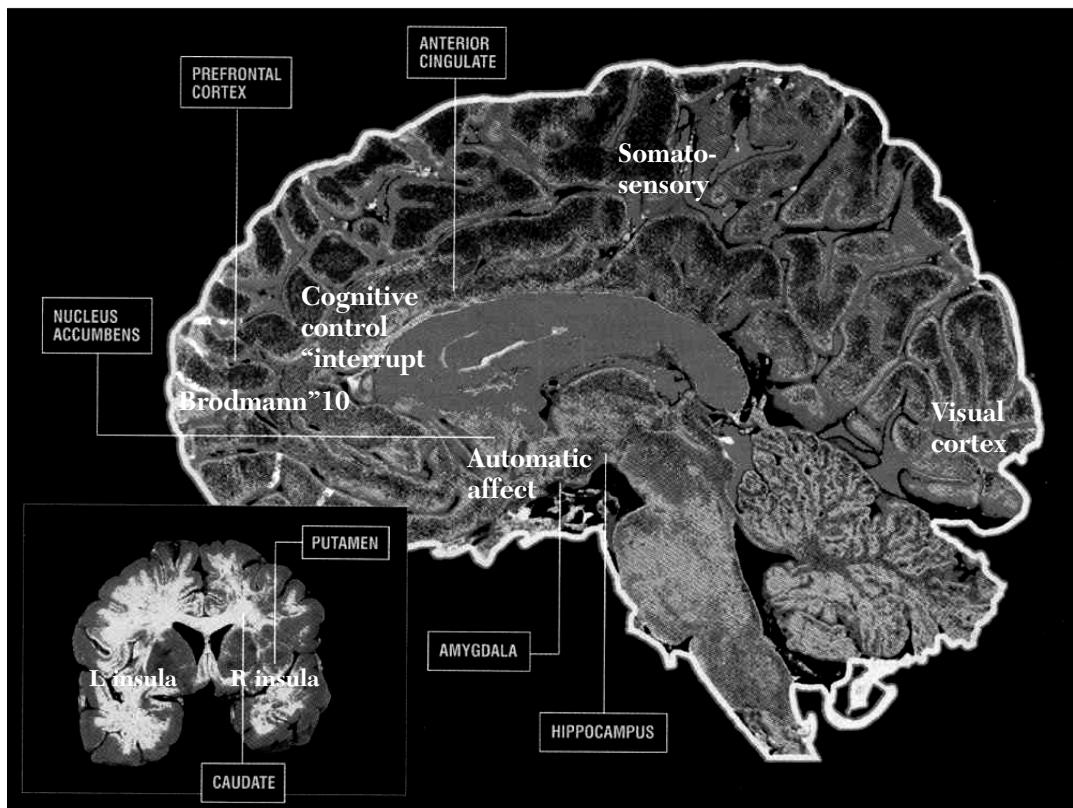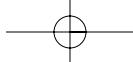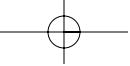
Figure 1. The human brain with some economically relevant areas marked.

including commercial ones. Models of this type have a very different structure from the systems of equations that economists typically work with. Unlike systems of equations, they are "black-box"; it is hard to intuit what they are doing by looking at individual parameters.

Because automatic processes are not accessible to consciousness, people often have surprisingly little introspective insight into why automatic choices or judgments were made. A face is perceived as "attractive" or a verbal remark as "sarcastic" automatically and effortlessly. It is only later that the controlled system may reflect on the judgment and attempt to substantiate it logically, and when it does, it often does so spuriously (e.g., Timothy Wilson, Samuel Lindsey, and Tonya Schooler 2000).

Automatic and controlled processes can be roughly distinguished by where they occur in the brain (Lieberman et al. 2002). Regions that support cognitive automatic activity are concentrated in the back (occipital), top (parietal), and side (temporal) parts of the brain (see Figure 1). The amygdala, buried below the cortex, is responsible for many important automatic affective responses, especially fear. Controlled processes occur mainly in the front (orbital and prefrontal) parts of the brain. The prefrontal cortex (pFC) is sometimes called the "executive" region, because it draws inputs from almost all other regions, integrates them to form near and long-term goals, and plans actions that take these goals into account (Timothy Shallice and Paul Burgess 1996). The prefrontal area is the

region that has grown the most in the course of human evolution and which, therefore, most sharply differentiates us from our closest primate relatives (Stephen Manuck et al. 2003).

Automatic processes—whether cognitive or affective—are the default mode of brain operation. They whir along all the time, even when we dream, constituting most of the electro-chemical activity in the brain. Controlled processes occur at special moments when automatic processes become "interrupted," which happens when a person encounters unexpected events, experiences strong visceral states, or is presented with some kind of explicit challenge in the form of a novel decision or other type of problem. To the degree that controlled processes are well described by economic calculation but parallel processes are not, one could say that economics is about the "interrupt" or "override." [4]

3.1.2 *Affective and Cognitive Processes*

The second distinction, represented by the two columns of table 1, is between *affective* and *cognitive* processes. Such a distinction is pervasive in contemporary psychology (e.g., Robert Zajonc 1980, 1984, 1998; Zajonc and Daniel McIntosh 1992) and neuroscience (Damasio 1994; LeDoux 1996; Panksepp 1998), and has an historical lineage going back to the ancient Greeks and earlier (Plato described people as driving a chariot drawn by two horses, reason and passions).

The distinguishing features of affective processing are somewhat counterintuitive. Most people undoubtedly associate affect with feeling states, and indeed most affect states do produce feeling states when they reach a threshold level of intensity. However, most affect probably operates below the threshold of conscious awareness
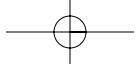
(LeDoux 1996; Piotr Winkielman and Kent Berridge 2004). As Rita Carter (1999) comments, "the conscious appreciation of emotion is looking more and more like one quite small, and sometimes inessential, element of a system of survival mechanisms that mainly operate—even in adults—at an unconscious level."

For most affect researchers, the central feature of affect is not the feeling states associated with it, but its role in human motivation. All affects have "valence"—they are either positive or negative (though some complex emotions, such as "bittersweet," can combine more basic emotions that have opposing valences). Many also carry "action tendencies" (Nico Frijda 1986; Leonard Berkowitz 1999)—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—as well as diverse other effects on sensory perception, memory, preferences, and so on (see, e.g., Leda Cosmides and John Tooby 2004). Affective processes, according to Zajonc's (1998) definition, are those that address "go/no-go" questions— that motivate approach or avoidance behavior. Cognitive processes, in contrast, are those that answer true/false questions.[5] Though it is not essential to our overall argument, our view is that cognition by itself cannot produce action; to influence behavior, the cognitive system must operate via the affective system.

Affect, as we use the term, embodies not only 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. Ross Buck (1999) refers to these latter influences as "biological affects," which he distinguishes

---

[4] As David Laibson and Andrew Caplin, respectively, have aptly expressed it.

[5] By this definition, neural processes that don't have valence are not affects. There are also neural processes that produce actions that are not best defined as affects – e.g., the reflexes that cause you to draw away from a hot object or an electric shock.

from the more traditional "social affects." Affect, thus, coincides closely with the historical concept of *the passions*. 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, a recent study showed that hurt feelings activated the same brain regions activated by broken bones or other physical injuries (Naomi Eisenberger et al. 2003).[6]

3.1.3 *The Quadrants in Action: An Illustration*

These two dimensions, in combination, define four quadrants (labeled I to IV in Table 1). Quadrant I is in charge when you deliberate about whether to refinance your house, poring over present-value calculations; quadrant II is undoubtedly the rarest in pure form. It is used by "method actors" who imagine previous emotional experiences so as to actually experience those emotions during a performance; quadrant III governs the movement of your hand as you return serve; and quadrant IV makes you jump when somebody says "Boo!"

Most behavior results from the interaction of all four quadrants. A natural instinct of economists trained in parsimonious modeling is to think that cognition is typically controlled, and affect is automatic, so there are really only two dimensions (quadrants I and IV) rather than four. But a lot of cognitive processing is automatic as well—e.g., visual perception or language processing. Research on "emotional regulation" shows many ways that cognition influences emotion, which implies the capacity for controlling emotion.[7]

The four-quadrant model is just a way to remind readers that the cognitive–affective and controlled–automatic dimensions are not perfectly correlated, and to provide a broad view to guide exploratory research. For some purposes, reducing the two dimensions to one, or the four quadrants to two, will certainly be useful. Furthermore, noting all four cells is not a claim that all are equally important. It is just a suggestion that leaving out one of the combinations would lead to a model which is incomplete for some purposes.

Consider what happens when a party host approaches you with a plate of sushi.

Quadrant III: Your first task is to figure out what is on the plate. The occipital cortex in the back of the brain is the first on the scene, drawing in signals from your eyes via your optic nerves. It decodes the sushi into primitive patterns such as lines and corners then uses a "cascading process" to discern larger shapes (Stephen Kosslyn 1994). Further downstream, in the inferior temporal visual cortex (ITVC), this information becomes integrated with stored representations of objects, which permits you to recognize the objects on the plate as sushi. This latter process is extraordinarily complicated (and has proved difficult for artificial intelligence researchers to recreate in computers) because objects can take so many forms, orientations, and sizes.

Quadrant IV: This is where affect enters the picture. Neurons in the inferior temporal visual cortex are sensitive only to the

---

[6] 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. Subjects thought they were playing a game that involved throwing a ball back and forth with two other people, but in fact the computer controlled the two animated figures that they saw on the screen. After a period of three-way play, the two other "players" began to exclude the subjects by throwing the ball back and forth between themselves. The social snub triggered neural activity in a part of the brain called the anterior cingulate cortex, which also processes physical pain, and the insula, which is active during physical and social discomfort.

[7] For example, when people are shown unpleasant photographs and told to interpret the photos so that they don't experience negative emotions (e.g., imagine a picture of women crying as having been taken at a wedding), there is less activity in the insula and amygdala, emotional areas, and in medial orbitofrontal cortex, and area which the insula projects to (Kevin Ochsner et al., in press).

identity of an object; they don't tell you whether it will taste good. Outputs of the inferior temporal visual cortex as well as outputs from other sensory systems feed into the orbitofrontal cortex to determine the "reward value of the recognized object." This is a highly particular representation. In economic terms, what is represented is neither pure information (i.e., that this is sushi) nor pure utility (i.e., that it is something I like) but rather a fusion of information *and* utility. It is as if certain neurons in the orbitofrontal cortex are saying "this is sushi and I want it."

The reward value of sushi depends in turn on many factors. First, there is your personal history with sushi. If you got sick on sushi in the past, you will have an unconscious and automatic aversion to it ("taste aversion conditioning"). The amygdala seems to play a critical role in this kind of long-term learning (LeDoux 1996). Second, the reward value of the sushi will depend on your current level of hunger; people can eat almost anything—grass, bugs, human flesh—if they are hungry enough. The orbitofrontal cortex and a subcortical region called the hypothalamus are sensitive to your level of hunger (Rolls 1999). Neurons in these regions fire more rapidly at the sight or taste of food when you are hungry, and fire less rapidly when you are not hungry.

Quadrants I and II: Processing often ends before quadrants I and II go to work. If you are hungry, and like sushi, your motor cortex will guide your arm to reach for the sushi and eat it, drawing on automatic quadrant III (reaching) and IV (taste and enjoyment) processes. Under some circumstances, however, higher level processing may enter the picture. If you saw a recent documentary on the risks of eating raw fish, you may recoil; or if you dislike sushi but anticipate disappointment in the eyes of your proud host who made the sushi herself, you'll eat it anyway (or pick it up and discreetly hide it in a napkin
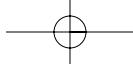
when she turns to serve other guests).[8] These explicit thoughts involve anticipated feelings (your own and the host's) and draw on explicit memories from a part of the brain called the hippocampus (see figure 1), inputs from the affective system (sometimes referred to as the "limbic system"), and anticipation (planning) from the prefrontal cortex.

Because standard economics is best described by the controlled, cognitive, processes of quadrant 1, in the remainder of this section we focus on the other half of each dichotomy—automatic and affective processes—providing further details of their functioning.

3.2 *Automatic Processes*

Here, we review some key principles of neural functioning that characterize automatic processes. Our short-list includes: *parallelism, specialization, and coordination*. Unpacking this a bit, we would say that: (1) much of the brain's processing involves processes that unfold in parallel and are not

---

[8] Paul Romer (2000) uses the example of peanut tastes as an illustration of how understanding the cause of revealed preference matters. One person loves the taste of peanuts, but is allergic to them and knows that the consequences of eating would be disastrous. When she is hungry, her visceral system motivates her to eat peanuts, but her deliberative system, with its ability to consider delayed consequences, inhibits her from eating them. The other person developed a "taste aversion" to peanuts many years ago, as a result of having gotten sick right after eating them. She knows at a cognitive level that the peanuts were not the cause of her sickness, but her visceral system overrules cognitive awareness. Revealed preference theory would stop at the conclusion that both women have disutility for peanuts. But the fact that the mechanisms underlying their preferences are different leads to predictable differences in other kinds of behavior. For example, the taste-averse woman will have a higher price elasticity (she'll eat peanuts if paid enough) and she will learn to enjoy peanuts after eating them a few times (her taste-aversion can be "extinguished"). The allergic woman will also like the smell of peanuts, which the taste-averter won't. Treatments also differ: cognitive therapies for treating harmless phobias and taste-aversion train people to use their conscious quadrant I processing to overrule visceral quadrant IV impulses, whereas treatment of the allergic peanut-avoider will concentrate on a medical cure.

accessible to consciousness; (2) the brain utilizes multiple systems specialized to perform specific functions; and (3) it figures out how to use existing specialized systems to accomplish new tasks efficiently, whatever functions they originally evolved to perform.

3.2.1 *Parallelism*

The brain performs a huge number of different computations in parallel. Because of the massively interconnected "network" architecture of neural systems, computations done in one part of the brain have the potential to influence any other computation, even when there is no logical connection between the two.
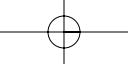
Recent psychological research on automatic processing provides many striking examples of such spurious interactions. In one particularly clever study, Gary Wells and Richard Petty (1980) had subjects listen to an editorial delivered over headphones while shaking their head either up–down, or right–left (subjects were led to believe that shaking was a legitimate part of the "product test"). Those who were instructed to shake up–down reported agreeing with the editorial more than those instructed to shake left–right, presumably because in our culture up–down head movement is associated with an attitude of acceptance, and left–right with an attitude of rejection. A similar impact on preference was also observed in a study in which subjects were asked to evaluate cartoons while holding a pen either clamped horizontally between their teeth or grasping it with pursed lips as if puffing on a cigarette (Fritz Strack, Leonard Martin, and Sabine Stepper 1988). The horizontal clamp forces the mouth into a smile, which enhances ratings, while "puffing" forces the mouth into a pursed expression, which lowers ratings. What the brain seems to be doing, in these examples, is seeking a "global equilibrium" that would reconcile the forced action (e.g., forced smile) with the response and the perceived attributes of the evaluated object.

Given that people are capable of deliberation, why doesn't quadrant I thinking correct the automatic activity produced by quadrant III and IV processes when they are wrong? Indeed sometimes it does: pilots learn to trust their instrument panels, even when they conflict with strong sensory intuitions about where their plane is headed, but such cognitive override is more the exception than the rule. To override automatic processes, quadrant I has to (a) recognize that an initial impression is wrong (which requires self-awareness about behavior in the other quadrants), and then (b) deliberately correct that impression. But when sense making works outside of consciousness it will not generate alarm bells to trigger the recognition required in (a). This is surely the case in the Epley and Gilovich studies. Even when the external influence is obvious and inappropriate, or the subject is warned ahead of time, the deliberation required to correct the first impression is hard work, competing for mental resources and attention with all the other work that needs to be done at the same moment (Daniel Gilbert 2002). The struggle between rapid unconscious pattern-detection processes and their slow, effortful modulation by deliberation is not a fair contest; so automatic impressions will influence behavior much of the time.

3.2.2 *Specialization*

Neurons in different parts of the brain have different shapes and structures, different functional properties, and operate in coordination as systems that are functionally specialized. Progress in neuroscience often involves tracing well-known psychological functions to circumscribed brain areas. For example, Broca's and Wernicke's areas are involved, respectively, in the production and comprehension of language. Patients with Wernicke damage cannot understand spoken words. While they can produce words themselves, they can't monitor their own speech, which results in sentences of correctly articulated words strung together into unintelligible gibberish. People with damage to Broca's area, in contrast, can understand what is said to them and typically know what

they want to say, but have difficulty articulating it, sometimes to the point of not being able to generate any words at all.

Beyond uncovering the nature of these specialized systems, neuroscience has led to the discovery of new functional systems, some of which are quite surprising. For example, surgeons conducting brain surgery on an epileptic patient discovered a small region of her brain which, when stimulated, caused her to laugh (Itzhak Fried 1998), hinting at the existence of a "humor system." Neuroscientists have also located an area in the temporal lobe that, when stimulated electrically, produces intense religious feelings—e.g., the sense of a holy presence or even explicit visions of God or Christ, even in otherwise unreligious people (Michael Persinger and Faye Healey 2002).

More generally, neuroscience has begun to change our classification of functional processes in the brain, in some cases identifying distinct brain processes that serve the same function, and in others drawing connections between processes that have been commonly viewed as distinct.

An example of the former is memory. Studies of patients with localized brain damage have confirmed the existence of distinct memory systems, which can be selectively "knocked out." Anterograde amnesiacs, for example, are commonly able to recall information acquired before injury, and they are able to acquire implicit information, including perceptual-motor skills (like the ability to read text in a mirror), but they are unable to recall new *explicit* information for more than a minute.[9] Anterograde amnesiacs can also acquire new emotional associations without the explicit memories necessary to make sense of them. In one famous example, a doctor introduced himself with a tack concealed in his hand that pricked a patient when they shook hands. Later the doctor returned and, although he did not remember the doctor, he nevertheless responded negatively to the doctor's arrival and refused to shake his hand.

An example of the latter involves the experience of fear, and recognition of fear in others. Experiencing an emotion and recognizing the emotion when displayed intuitively seems to involve entirely different processes. However, recent findings suggest that this is not the case. Numerous studies have implicated the amygdala—a small "organ" in the brain that is also intimately linked to the sense of smell—to fear processing. Lesions to the amygdala of rats and other animals disrupt or even eliminate the animals' fear-responses. Humans with stroke damage in the amygdala show similar deficits in reacting to threatening stimuli.

The same damage that disrupts fear responses in humans also cripples a person's ability to recognize facial expressions of fear in others, and to represent such expressions in pictures. Figure 2 shows how a patient with amygdala damage expressed different emotions when asked to draw them in the form of facial expressions. The patient was able to render most emotions with remarkable artistic talent. But when it came to drawing an expression of *fear* she felt clueless. She didn't even try to draw an adult face; instead she drew a picture of a crawling infant looking apprehensive.

The finding that people who don't experience fear also can't recognize it or represent it pictorially, suggests that two phenomena that had been viewed as distinct—experiencing and representing fear—in fact have important commonalities. Beyond this, they point to the intriguing notion that to recognize an emotion in others, one needs to be able to experience it oneself (see Alvin Goldman 2003).

The idea that people have specialized systems that are invoked in specific situations could have dramatic consequences for economics. The standard model of economic behavior assumes that people have a unitary

---

[9] Retrograde amnesia is the more familiar kind, in which people forget their past but can form new memories. The 2000 movie "Memento" drew an accurate portrait of anterograde amnesia and—interestingly for economists— its vulnerability to exploitation by people who know you are forgetting what they did to you.
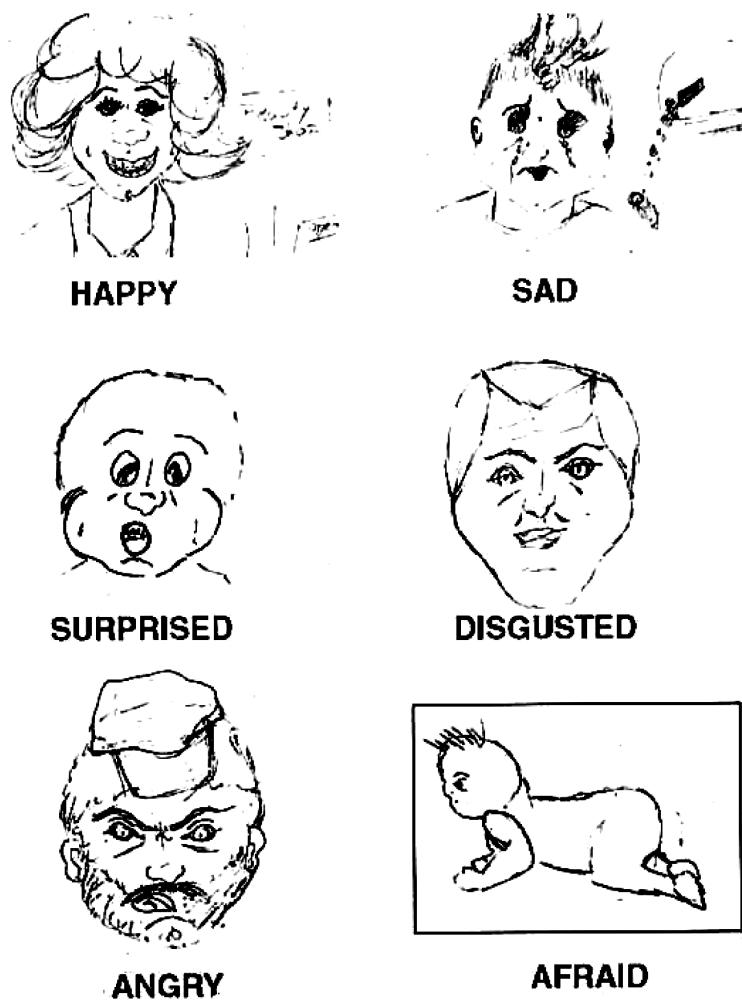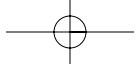
Figure 2. Representations of different emotions drawn by a patient with amygdala damage
(Adolphs et al. 1995).

set of preferences which they seek to satisfy, and economists often criticize psychology for lacking such a unified perspective. The existence of such selectively—invoked, specialized, systems, however, raises the question of whether a unified account of behavior is likely to do a very good job of capturing the complexities of human behavior. As Jonathan Cohen (personal communication) aptly expressed it at a recent conference on neuroeconomics, "Economics has one theory, psychologists many—perhaps because

brains have different systems that they use to solve different problems."

3.2.3 *Coordination*

In a process that is not well understood, the brain figures out how to do the tasks it is assigned efficiently, using the specialized systems it has at its disposal. When the brain is confronted with a new problem it initially draws heavily on diverse regions, including, often, the prefrontal cortex (where controlled
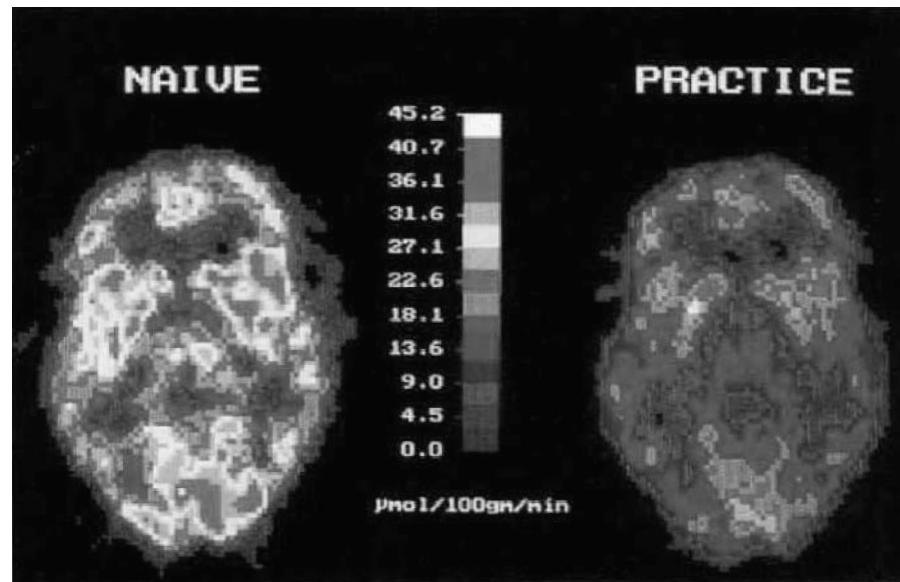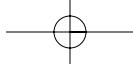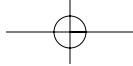
Figure 3. Regions of brain activation when first playing Tetris (left) and after several weeks of practice (right) (Haier et al. 1992).

processes are concentrated). But over time, activity becomes more streamlined, concentrating in regions that specialized in processing relevant to the task. In one study (Richard Haier et al. 1992), subjects' brains were imaged at different points in time as they gained experience with the computer game Tetris, which requires rapid hand–eye coordination and spatial reasoning. When subjects began playing, they were highly aroused and many parts of the brain were active (Figure 3, left panel). However, as they got better at the game, overall blood flow to the brain decreased markedly, and activity became localized in only a few brain regions (Figure 3, right panel).

Much as an economy ideally adjusts to the introduction of a new product by gradually shifting production to the firms that can produce the best goods most cheaply, with experience at a task or problem, the brain seems to gradually shift processing toward brain regions and specialized systems that can solve problems automatically and efficiently with low effort.

Andrew Lo and Dmitry Repin (2002) observe a similar result in a remarkable study of professional foreign-exchange and derivatives traders who were wired for psychophysiological measurements while they traded. Less-experienced traders showed significant physiological reactions to about half of the market events (e.g., trend reversals). More experienced traders reacted much less to the same events. Years of automatizing apparently enabled seasonal traders to react calmly to dramatic events that send a novice trader on an emotional roller coaster.

Given the severe limitations of controlled processes, the brain is constantly in the process of automating the processing of tasks—i.e., executing them using automatic rather than controlled processes. Indeed, one of the hallmarks of expertise in an area is the use of automatic processes such as visual imagery and categorization. In one now-famous study, Fernand Gobet and Simon (1996) tested memory for configurations of chess pieces positioned on a chessboard.

They found that expert chess players were able to store the positions of players almost instantly—but only if they were in positions corresponding to a plausible game. For randomly arranged chess pieces, the experts were not much better than novices. Further investigations have found that chess grandmasters store roughly 10,000 different possible board setups in memory which they can recognize almost instantly and respond to. More recent research in decision making suggests that this is a more general phenomenon, because decision making often takes the form of pattern matching rather than of an explicit weighing of costs and benefits (e.g., Robyn Leboeuf 2002; Doug Medin and Max Bazerman 1999).

In some cases, repeated use of particular specialized systems can produce physically recognizable changes. Studies have found that, for example, violinists who finger violin strings with their left hand show enlarged development of cortical regions which correspond to fingers on the left hand (Thomas Elbert et al. 1995), and the brain regions responsible for navigation and spatial memory (the hippocampus) of London taxi drivers are larger than comparable areas in non-taxi drivers (Eleanor Maguire et al. 2000). While the exact causality is difficult to determine (perhaps preexisting differences in size of brain region affect people's talents and occupational choices), a causal connection in the posited direction has been demonstrated in songbirds, whose brains show observable differences as a function of whether or not they have been exposed to the song that is characteristic of their species (Carol Whaling et al. 1997).

3.2.4 *The Winner Take All Nature of Neural Processing*
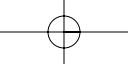
Another feature that is reminiscent of the operation of an economy is the "winner-take-all" nature of neural information processing (M. James Nichols and Bill Newsome 2002). While there is a lot that we do not know about the way in which the huge amount of neural activity is distilled into a categorical percept, or a decision, we do know that the brain doesn't invariably integrate (i.e., average) over the signals carried by individual neurons. In particular, when two distinct groups of neurons convey different information about the external world, the resulting perceptual judgment often adopts the information of one neuronal group and entirely suppresses the information carried by the other. This is referred to as a "winner-take-all" principle of neural signal-extraction. As a result, many brain processes are fundamentally categorical, yielding well-defined perceptions and thoughts even if the incoming information is highly ambiguous.[10] The advantages of this principle are clear if the ultimate job of the brain is to initiate a discrete action, or categorize an object as one of several discrete types. The disadvantage is that updating of beliefs in response to new information can procede in fits and starts, with beliefs remaining static as long as the new information does not lead to recategorization, but changing abruptly and dramatically when the accumulation of evidence results in a change of categorization (see Sendhil Mullainathan 2002).

3.3 *Affective Processes*

The way the brain evolved is critical to understanding human behavior. In many domains, such as eating, drinking, sex, and even drug use, human behavior resembles that of our close mammalian relatives, which is not surprising because we share many of the neural mechanisms that are largely responsible for these behaviors. Many of the

---

[10] Alternatively, one could imagine that perceptions or beliefs aggregate all relevant neural information (which would also be more in tune with Bayesian updating). So far as we can tell, the brain can use both principles of aggregation: when different neural populations carry "similar" information, then the overall perceptual judgment is an average of all information; when they carry very different information, then the overall judgment follows the "winner-take-all" rule (Nichols and Newsome 2002). For a simple model of a neural network that exhibits these properties, see Richard Hahnloser et al. 2000.

processes that occur in these systems are affective rather than cognitive; they are directly concerned with motivation. This might not matter for economics were it not for the fact that the principles that guide the affective system—the way that it operates— is so much at variance with the standard economic account of behavior.

3.3.1 *Primacy of Affect*
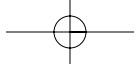
In contrast to the intuitive view of human behavior as driven by deliberations about costs and benefits, it does not do a terrible injustice to the field of psychology to say that a growing consensus has developed around the view that affect is *primary* in the sense that it is "first on the scene" and plays a dominant role in behavior. Indeed, as we discuss below (section 3.4.2), the conscious brain often erroneously interprets behavior that emerges from automatic, affective, processes as the outcome of cognitive deliberations.

In a series of seminal papers, Zajonc (1980, 1984, 1998) presented the results of studies which showed, first, 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, second, that affective reactions to things can be dissociated from memory for details of those things, with the former often being better. For example, we often remember whether we liked or disliked a particular person, book or movie, without being able to remember any other details (Bargh 1984). Subsequent research in social psychology takes Zajonc's initial research a step further by showing that the human brain affectively tags virtually all objects and concepts, and that these affective tags are brought to mind effortlessly and automatically when those objects and concepts are evoked (e.g., Russell Fazio et al. 1986; Anthony Greenwald, Mark Klinger, and Thomas Liu 1989; Greenwald 1992; Bargh, Shelly Chaiken, Paula Raymond, and Charles Hymes 1996; Jan De Houwer, Dirk Hermans, and Paul Eelen 1998; David Houston and Fazio 1989)

LeDoux and his colleagues (summarized in LeDoux 1996) have arrived at similar conclusions about the primacy of affect using very different research methods and subjects. Based on studies conducted on rats, they discovered that there are direct neural projections from the sensory thalamus (which performs crude signal-processing) to the amygdala (which is widely believed to play a critical role in the processing of affective stimuli) that are not channeled through the neocortex. As a result, animals can have an affective reaction to stimuli before their cortex has had the chance to perform more refined processing of the stimuli—they are literally afraid before knowing whether they should be. 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. They also provide a mechanism for interrupting and refocusing attention (Simon 1967), typically shifting processing from automatic to controlled processing. As Jorge Armony et al. (1995) note, "A threatening signal, such as one indicating the presence of a predator, arising from outside of the focus of attention will have a reduced representation in the cortex. Thus, if the amygdala relied solely on the cortical pathway to receive sensory information, it would not be capable of processing, and responding to, those danger signals that are not within the focus of attention" (see also, Armony et al. 1997; Gavin DeBecker 1997).[11]

A similar pattern to the one LeDoux observed in rats can also be seen in humans. Gilbert and Michael Gill (2000) propose that people are "momentary realists" who trust their immediate emotional reactions and only correct them through a comparatively laborious cognitive process. If the car behind

[11] The amygdala has excellent properties for performing such a function. Recent research in which amygdalae of subjects were scanned while threatening stimuli were flashed at various positions in the visual field found that amygdala activation is just as rapid and just as pronounced when such stimuli are presented outside as when they are presented inside the region of conscious awareness (Adam Anderson et al. 2003).

you honks after your light turns green, you are likely to respond with immediate anger, followed, perhaps, by a sheepish acknowledgment that maybe the honking person behind you had a point, as you were distracted when the light turned green.

It is important to note that while emotions may be fleeting, they can have a large economic impact if they create irreversible rash decisions (as in "crimes of passion"). Even emotions which could be transitory, such as embarrassment, can have long-run effects if they are kept alive by memory and social reminders. For example, Dora Costa and Matthew Kahn (2004) found that deserters in the Union Army during the Civil War, who were allowed to return to their hometowns without explicit penalty, often moved away because of the shame and social ostracism of being known to their neighbors as a deserter.

3.3.2 *Homeostasis*

To understand how the affective system operates, one needs to recall that humans did not evolve to be happy, but to survive and reproduce. An important process by which the body attempts to achieve these goals is called *homeostasis*. Homeostasis involves detectors that monitor when a system departs from a "set-point,"[12] and mechanisms that restore equilibrium when such departures are detected. Some—indeed most—of these mechanisms do not involve deliberate action. Thus, when the core body temperature falls below the 98.6° F set-point, blood tends to be withdrawn from extremities, and when it rises above the set-point one begins to sweat. But other processes do involve deliberate action—e.g., putting on one's jacket when

cold or turning on the air conditioner when hot. The brain motivates one to take such actions using both a "carrot" and a "stick." The stick reflects the fact that departing from a set-point usually feels bad—e.g., it feels bad to be either excessively hot or cold—and this negative feeling motivates one to take actions that move one back toward the set-point. The carrot is a process called "alliesthesia" (Michael Cabanac 1979) whereby actions that move one toward the set-point tend to feel pleasurable. When the body temperature falls below 98.6°, for example, almost anything that raises body temperature (such as placing one's hand in warm water) feels good, and conversely when the body temperature is elevated almost anything that lowers body temperature feels good.

As economists, we are used to thinking of preferences as the starting point for human behavior and behavior as the ending point. A neuroscience perspective, in contrast, views explicit behavior as only one of many mechanisms that the brain uses to maintain homeostasis, and preferences as transient state variables that ensure survival and reproduction. The traditional economic account of behavior, which assumes that humans act so as to maximally satisfy their preferences, starts in the middle (or perhaps even toward the end) of the neuroscience account. Rather than viewing pleasure as the goal of human behavior, a more realistic account would view pleasure as a homeostatic cue—an informational signal.[13,14]

---

[12] The mechanisms that monitor the body's state can be exquisitely complex, and can include both internal and external cues. In the case of nutritional regulation, for example, internal cues include gastric distension (James Gibbs et al. 1981), receptors sensitive to the chemical composition of the food draining from the stomach D. Greenberg, Smith and Gibbs (1990); Arthur L. Campfield and Smith 1990). External cues include time of day, estimated time till the next feeding, and the sight or smell of food.

[13] Even deliberative behavior generated by quadrant I is typically organized in a fashion that resembles homeostasis (Miller, Eugene Gallanter, and Karl Pribram 1960; Loewenstein 1999). Rather than simply maximizing preferences in a limitless fashion, people set goals for themselves, monitor their progress toward those goals, and adjust behavior when they fall short of their goals.

[14] Of course, even the neuroscience account begins, in some sense, in the middle. A more complete understanding of behavior would also ask how these different mechanisms evolved over time. Since evolution selects for genes that survive and reproduce, the result of evolution is unlikely to maximize pleasure or minimize pain. (See Gary Becker and Luis Rayo 2004.)

An important feature of many homeostatic systems is that they are highly attuned to *changes* in stimuli rather than their *levels*. A dramatic demonstration of such sensitivity to change came from single-neuron studies of monkeys responding to juice rewards (see Wolfram Schultz and Anthony Dickinson 2000). These studies measured the firing of dopamine neurons in the animal's ventral striatum, which is known to play a powerful role in motivation and action. In their paradigm, a tone was sounded, and two seconds later a juice reward was squirted into the monkey's mouth. Initially, the neurons did not fire until the juice was delivered. Once the animal learned that the tone forecasted the arrival of juice two seconds later, however, the same neurons fired at the sound of the tone, but *did not* fire when the juice reward arrived. These neurons were not responding to reward, or its absence . . . they were responding to deviations from expectations. (They are sometimes called "prediction neurons.") When the juice was expected from the tone, but was not delivered, the neurons fired at a very low rate, as if expressing disappointment. The same pattern can be observed at a behavioral level in animals, who will work harder (temporarily) when a reinforcement is suddenly increased and go "on strike" when reinforcement falls.[15] Neural sensitivity to change is probably important in explaining why the evaluation of risky gambles depends on a reference point which
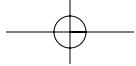
encodes whether an outcome is a gain or a loss (see section 5), why self-reported happiness (and behavioral indicators like suicide) depend on changes in income and wealth, rather than levels (Andrew Oswald 1997), and why violations of expectations trigger powerful emotional responses (George Mandler 1982).

3.4 *Interactions between the Systems*

Behavior emerges from a continuous interplay between neural systems supporting activity within each of the four quadrants. Three aspects of this interaction bear special emphasis, which we labeled "collaboration," "competition," and "sense-making." Collaboration captures the insight that decision making, which is to say "rationality" in the broad, nontechnical sense of the word, is not a matter of shifting decision-making authority from quadrants II, III, and IV toward the deliberative, nonaffective quadrant I, but more a matter of maintaining proper collaboration in activity across all four quadrants. If quadrant I tries to do the job alone, it will often fail.[16] Competition reflects the fact that different processes—most notably affective and cognitive—often drive behavior in conflicting directions and compete for control of behavior. Sense-making refers to how we make sense of such collaboration and competition—how we make sense of our behavior. While behavior is, in fact, determined by the interaction of all four quadrants, conscious reflection on our behavior, and articulating reasons for it, is basically quadrant I trying to make sense of that interaction. And, not surprisingly,

---

[15] As Rolls (1999) writes, "We are sensitive to some extent not just to the absolute level of reinforcement being received, but also to the change in the rate or magnitude of reinforcers being received. This is well shown by the phenomena of positive and negative contrast effects with rewards. Positive contrast occurs when the magnitude of a reward is increased. An animal will work very much harder for a period (perhaps lasting for minutes or longer) in this situation, before gradually reverting to a rate close to that at which the animal was working for the small reinforcement. A comparable contrast effect is seen when the reward magnitude (or rate at which rewards are obtained) is reduced—there is a negative overshoot in the rate of working for a time."

[16] Roy Baumeister, Todd Heatherton, and Dianne Tice (1994) review studies which suggest that excessively high incentives can result in supra-optimal levels of motivation that have perverse effects on performance (known as the Yerkes–Dodson law). Beyond documenting such "choking under pressure," the review research showing that it often occurs because it causes people to utilize controlled processes for performing functions, such as swinging a golf club, that are best accomplished with automatic processes. See Dan Ariely et al. (2004) for the latest evidence.

quadrant I has a tendency to explain behavior in terms it can understand – in terms of quadrant I processes.

3.4.1 *Collaboration and Competition*
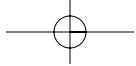
Although it is heuristically useful to distinguish between cognitive and affective processes, and between controlled and automatic processes, most judgments and behaviors result from interactions between them. Collaboration, delegation of activity, and proper balance across the quadrants are essential to normal decision making. Many decision-making disorders may originate in an improper division of labor between the quadrants. For example, psychiatry recognizes a decision-making continuum defined by the impulsive, "light" decision-making style at one end and the compulsive, "heavy" style at the other. The decisions of an impulsive individual are excessively influenced by external stimuli, pressures, and demands. Such a person may not be able to give a more satisfying explanation of an action except that "he felt like it" (David Shapiro 1965). By contrast, an obsessive–compulsive person will subject even the most trivial decisions to extensive deliberation and calculation. In situations in which it is entirely appropriate to make a quick decision based "on impulse"— e.g., when choosing which video to rent for the evening or what to order in a restaurant—the obsessive-compulsive will get stuck.

We are only now beginning to appreciate the importance of affect for normal decision making. The affective system provides inputs in the form of affective evaluations of behavioral options—what Damasio (1994) refers to as "somatic markers." Damasio and his colleagues show that individuals with minimal cognitive, but major affective deficits have difficulty making decisions, and often make poor decisions when they do (Antoine Bechara et al. 1994; Bechara, Damasio, Damasio, and Lee 1999; Damasio 1994). It is not enough to "know" what should be done; it is also necessary to "feel" it.

There is also interesting experimental evidence that deliberative thinking blocks access to one's emotional reactions to objects and so reduces the quality of decisions (e.g., Wilson and Schooler 1991). In one study (Wilson et al. 1993) college students selected their favorite poster from a set of posters. Those who were instructed to think of reasons why they liked or disliked the posters before making their selection ended up less happy on average with their choice of posters (and less likely to keep them up on their dorm room wall) than subjects who were not asked to provide reasons.

Needless to say, affect can also distort cognitive judgments. For example, emotions have powerful effects on memory—e.g., when people become sad, they tend to recall sad memories (which often increases their sadness). Emotions also affect perceptions of risks—anger makes people less threatened by risks, and sadness makes them more threatened (Jennifer Lerner and Dacher Keltner 2001). Emotions also create "motivated cognition"—people are good at persuading themselves that what they would *like* to happen is what *will* happen. Quack remedies for desperate sick people, and get-rich-quick scams are undoubtedly aided by the human propensity for wishful thinking. Wishful thinking may also explain high rates of new business failure (Camerer and Dan Lovallo 1999), trading in financial markets, undersaving, and low rates of investment in education (foregoing large economic returns). As LeDoux (1996) writes, "While conscious control over emotions is weak, emotions can flood consciousness. This is so because the wiring of the brain at this point in our evolutionary history is such that connections from the emotional systems to the cognitive systems are stronger than connections from the cognitive systems to the emotional systems."

When it comes to spending money or delaying gratification, taking or avoiding risks, and behaving kindly or nastily toward other people, people often find themselves

of "two minds"; our affective systems drive us in one direction and cognitive deliberations in another. We find ourselves almost compulsively eating our children's left-over Halloween candy, while obsessing about how to lose the extra ten pounds; gambling recklessly at the casino, even as a small voice in our head tell us to rein it in; trying to build up courage to step up to the podium; or tempted to give a little to the pathetic street-corner beggar though we know our crumpled dollar will go further if donated to United Way.

All of these deviations occur because our affective system responds to different cues, and differently to the same cues, as our cognitive system does. As Rolls (1999) writes,

> emotions often seem very intense in humans, indeed sometimes so intense that they produce behaviour which does not seem to be adaptive, such as fainting instead of producing an active escape response, or freezing instead of avoiding, or vacillating endlessly about emotional situations and decisions, or falling hopelessly in love even when it can be predicted to be without hope or to bring ruin. The puzzle is not only that the emotion is so intense, but also that even with our rational, reasoning, capacities, humans still find themselves in these situations, and may find it difficult to produce reasonable and effective behaviour for resolving the situation (p. 282).

Such divergences between emotional reactions and cognitive evaluations arise, Rolls (1999) argues, because

> in humans the reward and punishment systems may operate implicitly in comparable ways to those in other animals. But in addition to this, humans have the explicit system (closely related to consciousness) which enables us consciously to look and predict many steps ahead. (p. 282).
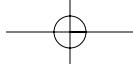
Thus, for example, the sight or smell of a cookie might initiate motivation to consume on the part of the affective system, but might also remind the cognitive system that one is on a diet.

Exactly how cognitive and affective systems interact in the control of behavior is not well understood. At a neurological level, it appears that an organ in the reptilian brain called the striatum (part of a larger system called the basal ganglia) plays a critical role. The striatum receives inputs from all parts of the cerebral cortex, including the motor cortex, as well as from affective systems such as the amygdala. Lesions of pathways that supply dopamine to the striatum leads, in animals, to a failure to orient to stimuli, a failure to initiate movements, and a failure to eat or drink (J. F. Marshall et al. 1974). In humans, depletion of dopamine in the striatum is found in Parkinson's disease, the most dramatic symptom of which is a lack of voluntary movement. The striatum seems to be involved in the selection of behaviors from competition between different cognitive and affective systems—in producing one coherent stream of behavioral output, which can be interrupted if a signal of higher priority is received, or a surprising stimulus appears (Carolyn Zink et al. 2003).

The extent of collaboration and competition between cognitive and affective systems, and the outcome of conflict when it occurs, depends critically on the intensity of affect (Loewenstein 1996; Loewenstein and Lerner 2003). At low levels of intensity, affect appears to play a largely "advisory" role. A number of theories posit that emotions carry information that people use as an input into the decisions they face (e.g., Damasio 1994; Ellen Peters and Paul Slovic 2000). The best-developed of these approaches is affect-as-information theory (Norbert Schwarz and Gerald Clore 1983; Schwarz 1990; Clore 1992).

At intermediate levels of intensity, people begin to become conscious of conflicts between cognitive and affective inputs. It is at such intermediate levels of intensity that one observes the types of efforts at self-control that have received so much attention in the literature (Jon Elster 1977; Walter Mischel, Ebbe Ebbesen, and Antonette Zeiss 1972; Thomas Schelling 1978, 1984).

Finally, at even greater levels of intensity, affect can be so powerful as to virtually preclude decision making. No one "decides" to fall asleep at the wheel, but many people do. Under the influence of intense affective motivation, people often report themselves as being "out of control" or "acting against their own self-interest" (Baumeister, Heatherton, and Tice 1994; Stephen Hoch and Loewenstein 1991; Loewenstein 1996). As Rita Carter writes, "where thought conflicts with emotion, the latter is designed by the neural circuitry in our brains to win" (1999).

3.4.2 *Spurious Sense-Making*

Sense-making is an important form of interaction between quadrant I and the other quadrants. The brain's powerful drive toward sense making leads us to strive to interpret our own behavior. Since quadrant I often does not have conscious access to activity in the other quadrants, it is perhaps not surprising that it tends to over-attribute behavior to itself—i.e., to deliberative decision processes. Even though much of the brain's activity is "cognitively inaccessible," we have the illusion that we are able to make sense of it, and we tend to make sense of it in terms of quadrant I processes.

Research with EEG recordings has shown (Benjamin Libet 1985) that the precise moment at which we become aware of an intention to perform an action trails the initial wave of brain activity associated with that action (the EEG "readiness potential") by about 300 msec. The overt behavioral response itself then follows the sensation of intention by another 200 msec. Hence, what is registered in consciousness is a regular pairing of the sensation of intention followed by the overt behavior. Because the neural activity antecedent to the intention is inaccessible to consciousness, we experience "free will" (i.e., we cannot identify anything that is causing the feeling of intention). Because the behavior reliably follows the intention, we feel that this "freely willed" intention is *causing* the action—but in fact,
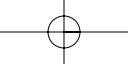
both the sensation of intention and the overt action are caused by prior neural events which are inaccessible to consciousness (see Daniel Wegner and Thalia Wheatley 1999).[17]

Quadrant I tends to explain behavior egocentrically—to attribute it to the types of deliberative processes that it is responsible for (see Richard Nisbett and Wilson 1977). A dramatic study demonstrating this phenomenon was conducted with a "split-brain" patient (who had an operation separating the connection between the two hemispheres of his brain). The patient's right hemisphere could interpret language but not speak, and the left hemisphere could speak (LeDoux 1996). The patient's right hemisphere was instructed to wave his hand (by showing the word "wave" on the left part of a visual screen, which only the right hemisphere processed). The left hemisphere saw the right hand waving but was unaware of the instructions that had been given to the right hemisphere (because the cross-hemisphere connections were severed). When the patient was asked why he waved, the left hemisphere (acting as spokesperson for the entire body) invariably came up with a plausible explanation, like "I saw somebody I knew and waved at them."

4. *General Implications of Neuroscience for Economics*

To add value to economics, neuroscience needs to suggest new insights and useful perspectives on old problems. This section discusses some broad implications for economics of the ideas and findings reviewed in the previous section. First, we show that neuroscience findings raise questions about the usefulness of some of the most common constructs that economists commonly use, such as risk aversion, time preference, and

---

[17] "… the brain contains a specific cognitive system that binds intentional actions to their effects to construct a coherent experience of our own agency" (Patrick Haggard, Sam Clark, and Jeri Kalogeras 2002).

altruism. Second, we show how the existence of specialized systems challenges standard assumptions about human information processing and suggests that intelligence and its opposite—bounded rationality—are likely to be highly domain-specific. Third, brain-scans conducted while people win or lose money suggest that money activates similar reward areas as do other "primary reinforcers" like food and drugs, which implies that money confers *direct* utility, rather than simply being valued only for what it can buy. Fourth, we show that research on the motivational and pleasure systems of the brain human challenges the assumed connection between motivation and pleasure. Finally, we describe some of the important implications of cognitive inaccessibility for economics.

4.1 *Economic Constructs*
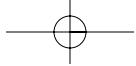
Knowing how the brain solves problems, and what specialized systems it has at its disposal to do so, challenges some of our fundamental assumptions about how people differ from one-another when it comes to economic behavior. Economists currently classify individuals on such dimensions as "time preference," "risk preference," and "altruism." These are seen as characteristics that are stable within an individual over time and consistent across activities; someone who is risk-seeking in one domain is expected to be risk-seeking in other domains as well. But empirical evidence shows that risk-taking, time discounting, and altruism are very weakly correlated or uncorrelated across situations. This inconsistency results in part from the fact that preferences are state-contingent (and that people may not recognize the state-contingency, which—if they did—would trigger overrides that impose more consistency than observed). But it also may point to fundamental problems with the constructs that we use to define how people differ from each other.

As an illustration, take the concept of time-preference. In empirical applications, economic analyses typically assume that the same degree of time preference is present for all intertemporal tradeoffs— saving for the future, flossing your teeth, dieting, and getting a tattoo. For example, whether a person smokes is sometimes taken as a crude proxy for low rates of time discounting, in studies of educational investment or savings. Thinking about the modularity of the brain suggests that while different intertemporal tradeoffs may have *some* element of planning in common (e.g., activity in prefrontal cortex), different types of intertemporal choices are likely to invoke qualitatively different mixtures of neural systems and hence to produce entirely different patterns of behavior. Then measured discount rates will not be perfectly correlated across domains, and might hardly be correlated at all.

Much as the study of memory has been refined by the identification of distinct memory systems, each with its own properties of learning, forgetting, and retrieval; the study of intertemporal choice might be enhanced by a similar decomposition, most likely informed by neuroscience research. For example, unpublished research by Loewenstein and Roberto Weber suggests that, in normal samples, future-oriented behaviors tend to cluster in tasks which tap different dimensions of self-control. For example, flossing your teeth is statistically associated with a number of other minor, repetitive, helpful behaviors such as feeding parking meters religiously and being on time for appointments. These behaviors seem to involve "conscientiousness," an important measure in personality theory.[18] John Ameriks, Andrew Caplin, and John Leahy (in press) measured the propensity to plan by asking TIAA–CREF enrollees to disagree or agree with statements like "I have spent a great deal of time developing a financial plan." These measures correlate with actual savings rates.

Dieting and use of addictive drugs, in contrast to punctuality and flossing, might be involve very different circuitry and hence reveal fundamentally different dis-

counting behavior. Both of these activities involve visceral motives that seem to differ reliably across persons and may be only weakly linked to conscientiousness. If you're the kind of person who loves to eat, or drink alcohol, then resisting indulging in those activities requires difficult exertion of self-control.

4.2 *Domain-Specific Expertise*

Economics implicitly assumes that people have general cognitive capabilities that can be applied to any type of problem and, hence, that people will perform equivalently on problems that have similar structure. The existence of systems that evolved to perform specific functions, in contrast, suggests that performance will depend critically on whether a problem that one confronts can be, and is in fact, processed by a specialized system that is well adapted to that form of processing. When a specialized system exists and is applied to a particular task, processing is rapid and the task feels relatively effortless. Automatic processes involved in vision, for example, are lightning fast and occur with no feeling of mental effort, so people are not aware of the power and sophistication of the processes that allow it to happen. Even the most powerful computers don't hold a candle to humans when it comes to visual perception or voice recognition.
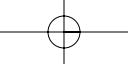
When we lack such tailored systems, however, we are likely to seem extraordinarily flat-footed because we will be forced to "muscle it out" with quadrant 1 processing, much as autistic individuals seem to solve theory of mind problems by building up a statistical understanding of appropriate social behavior. As a general rule, we should expect people to be geniuses when presented with problems that can be, and are

[18] Based on the observation that drugs affecting serotonin receptors are used to treat compulsion disorders, compulsivity appears to have some connection to the neurotransmitter serotonin, In addition, compulsion disorders are thought to be related to hyperactivity in the caudate nucleus, a "reminder" region of the limbic system.

processed by specialized systems, but relatively obtuse when they are forced to rely on controlled processes.

A neat illustration of this is provided by the "Wason four-card problem" in logic. Subjects are shown four cards, each with a letter on one side and a number on the other. The exposed card faces read "X," "Y," "1," and "2." Subjects are asked which cards would need to be turned over to test the rule: "If there is an X on one side there is a 2 on the other." Few subjects give the right answers, which are X and 1 (if there's an X on the opposite side of the "1" the rule is broken). However, most subjects give the right answer when the logically equivalent problem is put in a cheating-detection frame. For example, if there are four children from two different towns and two school districts and the rule is "If a child lives in Concord he or she must go to Concord High," most subjects realize that the home address of the student who does not go to Concord High must be checked to see if she is cheating and not going to Concord High when she should (Cosmides 1989).

Of particular interest to economics, many neuroscientists believe there is a specialized "mentalizing" (or "theory of mind") module, which controls a person's inferences about what other people believe, feel, or might do. The hint at the existence of such a dedicated module came from tests conducted by developmental psychologists in which two children are shown an object in the process of being concealed (see Uta Frith 2001a). One child then leaves, and the other child observes as the object is moved to a new location. The child who remains in the room is then asked to predict where the child who left will look for the object when she returns. Normal children are typically able to solve this problem around age three or four. Autistic children as a rule master this distinction much later (8–12 years), and with great difficulty, although some (especially those with "Asperger syndrome") have normal or superior intelligence. At the same time, the autistic child will have no

difficulty with general inferences having a similar logical form (e.g., if a photo is taken of the object's location, and the object is then moved, they will correctly infer that the photo will register the object in the old place, before it was moved).

Autistic adult individuals may compensate in many ways and eventually pass such basic tests of mentalizing. However, they have difficulty appreciating more subtle social meanings (e.g., irony), and will sometimes wonder at the "uncanny ability" of non–autistic persons to "read minds" (Frith 2001b). People with Asperger's syndrome, who are intelligent but have trouble understanding emotions in others, show lower activation in the medial prefrontal regions as compared with normals when presented with problems that involve mentalizing, but also show greater activation in more ventral (lower) region of the prefrontal cortex, which is normally responsible for general reasoning (Francesca Happe et al. 1996). A natural interpretation is that Asperger individuals eventually deduce the answer through a complex process of reasoning instead of grasping it directly by a specialized module, as if they create a neural "workaround" or "long-cut." In the medical literature, one can also find patients with brain lesions who exhibit difficulties with mentalizing tasks but not with other types of cognition (Andrea Rowe et al. 2001; James Blair and Lisa Cipolotti 2000). This, too, is consistent with the hypothesis of a separable mentalizing module.

The possibility of a mentalizing module has gained credibility and substance through converging neuroscientific evidence. fMRI studies have shown that when normal adults are given pairs of closely matching judgment problems, differing only in whether they do or do not require mentalizing, the mentalizing problems lead to greater activation in the left medial prefrontal cortex (Fletcher et al. 1995; Rebecca Saxe and Nancy Kanwisher 2003). As ultimate proof, one would like to identify neuronal populations that are specifically turned on by mentalizing activity. Neuroscience is not there yet, but recent single-cell recordings in monkeys have identified an intriguing class of "mirror neurons" in the prefrontal cortex, which fire either when an experimenter performs a physical action (e.g., grasping a peanut) or when the monkey performs ("mirrors") the same action. Having such neurons makes learning by imitation easy and supports mind reading by, for example, internally simulating the facial expressions of others.

Mentalizing is relevant for economics because many judgments require agents to make guesses about how other people feel or what they will do. The concept of equilibrium requires that agents correctly anticipate what others will do; presumably this arises because of accurate mentalizing, or through some kind of specific learning about behavior which may not transfer well to new domains or when variables change. Furthermore, the kind of learning about players' "types" from observation in Bayesian–Nash equilibrium is modeled as simply Bayesian updating of the likelihoods of chance events in the face of new information. Since mentalizing is a special ability, and logical-deductive reasoning can only partially compensate for its absence, treating inference about behavior of other agents and "nature" is a simplification which may be wrong.[19]

---

[19] Our point here is simple: Circuitry that controls updating of frequencies of events in the world—whether it snows in Chicago in January—may be dissociated from circuitry that controls updating of personal "types" from behavior. The circuitry which controls attributions of payoff types to people, from their behavior, requires a concept of how types link to behavior (i.e., what *are* types?), as well as updating of types (i.e., reputations) from behavior. An autist with theory of mind deficits, for example, might keep very accurate counts of relative frequencies (in fact, many autists are obsessed with counting objects in special categories) but be unable to tell whether a person behaved badly in a game because of bad intentions or situational influences (such as a bureaucrat who is pressured to stick to rules). So there is no reason to think that updating about event frequencies is necessarily the same as updating about personal types.

The existence of domain-specific expertise suggests that people will appear to be geniuses at some tasks but will seem remarkably flat-footed when dealing with other tasks that may be only superficially different. Domain-specific processing has important implications for economics, and specifically for the organization of labor. Bundling tasks together into jobs, for example, requires an understanding of which kind of skills are general (useful for many tasks) and which are neurally separate.

4.3 *Utility for Money*

As discussed earlier, neuroscience can point out commonalities between categories that had been viewed as distinct. An example of this with important implications for economics is the utility for money. The canonical economic model assumes that the utility for money is indirect—i.e., that money is a mere counter, only valued for the goods and services it can procure. Thus, standard economics would view, say, the pleasure from food or cocaine and the "pleasure" from obtaining money as two totally different phenomena. Neural evidence suggests, however, that the same dopaminergic reward circuitry of the brain in the midbrain (mesolimbic system) is activated for a wide variety of different reinforcers (Montague and Berns 2002), including attractive faces (Itzhak Aharon et al. 2001), funny cartoons (Dean Mobbs et al. 2003), cultural objects like sports cars (Susanne Erk et al. 2002), drugs (Schultz 2002), and money (e.g., Hans Breiter et al. 2001; Brian Knutson and Richard Peterson, in press; Delgado et al. 2000). This suggests that money provides *direct* reinforcement.

Figure 4 is a rough guess about the neural circuitry of reward (from Schultz 2002). It is useful as a pictorial reminder that, while neuroscience (and our review) often emphasize specific regions which are central in different kinds of processing, the focus in thinking about economic decisions should be on circuitry or *systems* of regions and how they interact. The diagram also shows how frontal regions (at the top of Figure 4) both receive input from "lower" systems (dopaminergic neurons and the amygdala), and also feed back processed information to the striatum.

The idea that many rewards are processed similarly in the brain has important implications for economics, which assumes that the marginal utility of money depends on what money buys. Of course, it is possible that money rewards activate the same circuitry as sports cars, cocaine, and jokes because the circuitry is being used to evaluate money in terms of the goods that it buys (which requires a cortical process that "simulates" the value of those goods internally to the brain). But it is also possible that money becomes what psychologists call a "primary reinforcer," which means that people value money without carefully computing what they plan to buy with it. Neuroeconomics is not nearly advanced enough to separate these two roles for midbrain responses to money, but suggests the *possibility* that the brain value of money is only loosely linked to consumption utility, which in turn suggests a wide range of potential experiments to explore implications.

An example of how consumption-driven and primary reinforcement from money can matter is asset pricing. Since Robert Lucas (1978), many models of stock prices have assumed that investors care about the utility of consumption they can finance with stock market returns, rather than with returns per se. Simple models of this sort make many counterfactual predictions—most famously that the "equity premium," or marginal return to stocks over bonds, would be much lower than it has actually been if investors only disliked risk because of its impact on consumption, as conventional models assume. Shlomo Benartzi and Richard Thaler (1995) and Nicholas Barberis, Tano Santos, and Ming Huang (2001) have had more success explaining returns patterns using a model in which investors care *directly* about stock returns. This alternative assumption fits a brain that
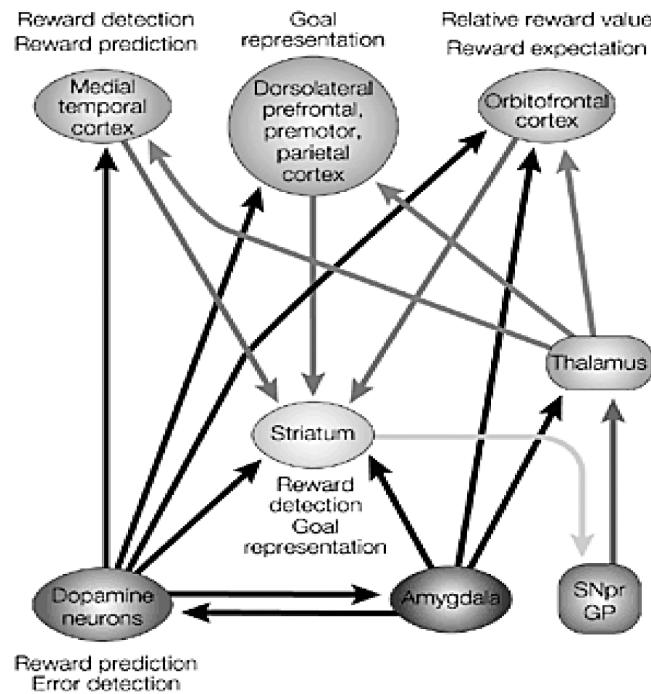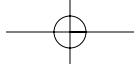
Figure 4. Hypothesized neural circuitry of reward processing (Schultz 2000).

gets a kick out of earning high returns for their own sake.

If gaining money provides direct pleasure, then the experience of parting with it is probably painful. While there is no direct evidence that paying is painful, the assumption that paying hurts can explain many market phenomena which are otherwise puzzling (Drazen Prelec and Loewenstein 1998). An example is the effect of payment-neutral pricing schemes on choices. Companies often go to great lengths to disguise payments, or reduce their pain. Consumers appear to oversubscribe to flat-rate payment plans for utilities and telephone service and health clubs (Kenneth Train 1991; Train, Daniel McFadden, and Moshe Ben-Akiva 1987; Della Vigna and Ulrike Malmendier 2003, Anja Lambrecht and Bernd Skiera 2004). A flat-rate plan eliminates marginal costs and allows consumers to enjoy the service without thinking about the marginal cost. Similarly, travel plans are often sold as packages, making it impossible to compute the cost of the individual components (hotel, food, transportation). Often components of the package are presented as "free" (like Microsoft's internet browser) even though the claim is meaningless from an economic standpoint, given that the package is presented on a take-it-or-leave-it basis. One can interpret the appeal of ad-hoc currencies, such as frequent-flyer miles, chips in casinos, or the beads used for incidental expenses at all-inclusive resorts like Club Med, as an attempt to reduce the pain-of-payment. The ad-hoc currency, whether miles or beads, feels like "play money," and spending it does not seem to exact the same psychic cost.

In experiments, we have observed a preference for prepayment for certain items, even where prepayment is financially irrational because it incurs an opportunity cost of lost interest payments (Prelec and Loewenstein 1998). When questioned, respondents claim

to prefer prepayment for, e.g., a vacation trip because then they could relax and enjoy the vacation more knowing it was paid up. It is an interesting question whether part of the motive for owning rather than renting products is precisely to create prepayment, and so enjoy consumption without being reminded of the cost. The pain-of-paying may also explain why we are willing to pay less for a product if paying in cash than by credit card. Although there are financial reasons to prefer paying by credit-card, the size of credit card relative to cash premium that people reveal, in an experiment, is much too high (up to 100 percent) to be rationalized by liquidity preference and other economic considerations (Prelec and Duncan Simester 2001).

4.4 *Wanting and Liking*

Economists usually view behavior as a search for pleasure (or, equivalently, escape from pain). The subfield of welfare economics, and the entire ability of economists to make normative statements, is premised on the idea that giving people what they want makes them better off. But, there is considerable evidence from neuroscience and other areas of psychology that the motivation to take an action is not always closely tied to hedonic consequences.

Berridge (1996) argues that decision making involves the interaction of two separate, though overlapping systems, one responsible for pleasure and pain (the "liking" system), and the other for motivation (the "wanting" system). This challenges the fundamental supposition in economics that one only strives to obtain what one likes. Berridge finds that certain lesions and pharmacological interventions can selectively enhance a rat's willingness to work for a food, without changing the pleasure of eating the food, as measured, admittedly somewhat questionably, by the animal's facial expression (Animal facial expressions, like those of humans, provide at least a clue about whether something tastes good, bad or indifferent. In economic language, the
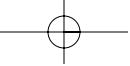
experiments create a situation where the utility of food and disutility of work remain the same, but the amount of work-for-reward goes up. This implies that it is possible to be motivated to take actions that bring no pleasure.

Berridge believes that the later stages of many drug addictions presents prototypical examples of situations of what he terms "wanting" without "liking"; drug addicts often report an absence of pleasure from taking the drugs they are addicted to, coupled with an irresistible motivation to do so. Other examples of situations in which there often seems to be a disconnect between one's motivation to obtain something and the pleasure one is likely to derive from it are sex and curiosity (Loewenstein 1994). Thus, for example, you can be powerfully motivated to seek out information, even when you are quite certain that it will make you miserable, and can feel quite unmotivated to engage in activities that, at a purely cognitive level, you are quite sure you would find deeply pleasurable.

Economics proceeds on the assumption that satisfying people's wants is a good thing. This assumption depends on knowing that people will like what they want. If likes and wants diverge, this would pose a fundamental challenge to standard welfare economics. Presumably welfare should be based on "liking." But if we cannot infer what people like from what they want and choose, then an alternative method for measuring liking is needed, while avoiding an oppressive paternalism.

4.5 *Cognitive Inaccessibility*

The fact that people lack introspective access to the sources of their own judgments of behavior, and tend to overattribute both to controlled processes, has many important implications for economics. When it comes to discriminatory biases, for example, because people lack introspective access to the processes that produce such biases, they are unable to correct for them
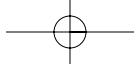
even when they are motivated to make impartial judgments and decisions. Indeed, they are very likely to deny that they are biased and, hence, likely to not even perceive the need for such correction. Such unconscious discrimination may explain why otherwise identical job application resumes for candidates with statistically "white" rather than "African–American" names have a 50 percent higher chance of generating callbacks from job application letters, as shown recently by Marianne Bertrand and Mullainathan (2004). It can also help explain why physicians are so convinced that gifts from pharmaceutical companies do not bias their prescription practices even though research (and the continuing practice of gift giving by the companies) suggests that it does (Jason Dana and Loewenstein 2003).

A second class of implications is related to the phenomena of apparent self-deception and self-manipulation, where, for instance, economic agents (investors, consumers, entrepreneurs) are overly optimistic about their chances for success. These phenomena have been richly catalogued by social psychologists, beginning with the research on motivated cognition and cognitive dissonance in the 1950s. Neuroscience suggests that they are all related to chronic cognitive inaccessibility of automatic brain processes. Attention, for example, is largely controlled by automatic processes, and attention in turn determines what information we absorb. If attention is chronically drawn to information that is favorable to us, we will emerge with an overoptimistic sense of our abilities and prospects. It is a case of quadrants III and IV collaborating, without the "adult supervision" provided by quadrant I.

A third class of implications arises from the cognitive inaccessibility of our own motives for action. The fact that we are not consciously aware of the moment of decision (as shown by Libet's research, mentioned in the previous section) strongly suggests that we may also not understand the reasons why we choose one way or another. Paradoxically, this can be very beneficial. In many situations, an action may be diagnostic of a good outcome, without having any significant ability to cause that outcome. For instance, participating in some socially desirable activity (e.g., voting or not littering) may be quite diagnostic of the desired collective outcome (your candidate winning, clean streets) without being able to cause that outcome, because the impact of your action is negligible. This gives rise to the notorious "voter's paradox" in rational choice theory, which seems to imply that no one should ever vote. Fortunately, the distinction between diagnosticity and causality, which is absolute in the rational-choice model, is quite fuzzy psychologically. There is experimental evidence that people will cheat on their own medical tests so as to "manufacture" a good diagnosis (George Quattrone and Amos Tversky 1984), and their own personality inventories, to yield a personality assessment diagnostic of success. This can only be possible if the true motive for action — the desire to get "good news" — is hidden from the agent at the moment of choice; if it were not hidden, then awareness that the action was taken precisely to get the good news would instantly void the diagnostic value of the action (see Ronit Bodner and Prelec 2003 for an economic self-signaling model that allows for noncausal motives on action). Cognitive inaccessibility prevents this logical short-circuit and dramatically expands the range of motives that can influence behavior. Because the hedonic system (quadrant IV) is not constrained by logical considerations, a person who takes a small step toward a larger virtuous objective, such as joining a gym to get fit, or buying a copy of Stephen Hawking's *A Brief History of Time* to feel scientifically literate, may experience a feeling of pleasure from the small step "as if" they have in fact secured their objective (which they most likely will not). At the same time, cognitive inaccessibility means that quadrant I doesn't have to acknowledge that the feeling of pleasure was in fact the cause of his action.

5. *Specific Economic Applications*

We now go into greater detail about ramifications of neuroscience findings for four specific topics in economics: intertemporal choice, decision making under risk and uncertainty, game theory, and labor-market discrimination.

5.1 *Intertemporal Choice and Self-Control*

The standard perspective in economics views intertemporal choice as a trade-off of utility at different points in time. Individual differences in the way that people make this tradeoff are captured by the notion of a discount rate—a rate at which people discount future utilities as a function of when they occur. The notion of discounting, however, gained currency not because of any supportive evidence, but based only on its convenient similarity to financial net present value calculations (Loewenstein 1992). Indeed, in the article that first proposed DU in detail, Samuelson (1937) explicitly questioned its descriptive validity, saying "It is completely arbitrary to assume that the individual behaves so as to maximize an integral of the form envisaged in [the DU model]."

In fact, more recent empirical research on time discounting challenges the idea that people discount all future utilities at a constant rate (see Frederick, Loewenstein, and O'Donoghue 2002). The notion of time discounting, it seems, neither describes the behavior of individuals nor helps us to classify individuals in a useful fashion. How can an understanding of the brain help us to establish a better understanding of intertemporal choice behavior? Our two central distinctions, between affect and cognition, and between automatic and controlled processes, both have important ramifications.
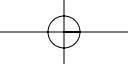
5.1.1 *Affect and Cognition in Intertemporal Choice*

As discussed above, the affective system is designed to ensure that certain survival and reproduction functions are met and it achieves this function in part by motivating individuals to take certain actions. In most animals, emotions and drives motivate behaviors that have short-term goals, such as eating, drinking, and copulating. As a result, the affective systems that we share with wide range of other animals are inherently myopic. Though *some* animals display far-sighted behaviors, such as storing food for winter, these are specialized, preprogrammed, behaviors that are distinctly different from the type of spontaneous delay of gratification observed in humans. Humans appear to be unique among animals in terms of caring about, making immediate sacrifices for, and flexibly responding to, desired future consequences.

This capacity to take long-term consequences of our behavior into account seems to be the product of our prefrontal cortex, which, tellingly, is the part of the brain that is uniquely human (see, e.g., Manuck et al. 2003). Patients with damage to prefrontal regions—most famously Phineas Gage, whose injury led to a reformulation of our understanding of the function of the prefrontal cortex—tend to behave myopically, paying little heed to the delayed consequences of their behavior. As Hersh Shefrin and Thaler (1988) suggested many years ago, intertemporal choice can be viewed as a splice of two processes—an impulsive, affective, process and a more far-sighted process guided by the prefrontal cortex.

Recent brain imaging research provides support for such an account. Samuel McClure et al. (2004) scanned subjects using fMRI while they made a series of preference judgments between monetary reward options that varied by amount and delay to delivery. For some pairs of choices, the earlier reward would be received immediately; for others, both rewards were delayed (though one by more than the other). Consistent with the idea that intertemporal choice is driven by two systems, one more cognitive and one more deliberative, they found that parts of the limbic—i.e., affective—system associated with the midbrain dopamine system were

preferentially activated by options involving immediately available rewards. In contrast, regions of the lateral prefrontal cortex and posterior parietal cortex—typically viewed as more cognitive regions—were engaged uniformly by intertemporal choices irrespective of delay. Furthermore, the relative activity of the two systems actually predicted the choices that subjects made: Greater relative activity in affective systems was associated with choosing earlier rewards more often.

The notion of quasi-hyperbolic time discounting, of course, provides a mathematical representation of precisely such a splicing of two processes, and has been shown to usefully describe behavior in a wide range of domains. However, an understanding of the neural underpinnings of these dual processes allows for more nuanced predictions and formulations. The notion of hyperbolic time-discounting predicts that people will always behave impulsively when faced with the right combination of incentives (typically those involving some immediate cost and benefit), but this does not seem to be the case. Understanding that hyperbolic time discounting stems, in part, from competition between the affective and cognitive systems, leads to the prediction that factors that strengthen or weaken one or the other of these influences will cause people to behave more or less impulsively.

5.1.2 *Determinants of the Relative Strength of Affect and Cognition*
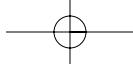
There are a variety of factors that affect the relative strength of affective and cognitive influences on intertemporal choice which can help to explain what could be called "intraindividual" variability in impatience.

First, any factor that increases the demands on the prefrontal cortex—on the controlled, cognitive, system, should decrease the influence of this system and, hence, decrease individuals' control over their own behavior. This possibility was demonstrated by Baba Shiv and Alexander Fedorikhin (1999). To manipulate "cognitive load," half of their subjects

(low-load) were given a two-digit number that they were instructed to memorize, and the other half (high load) were given a seven-digit number to memorize. Subjects were then instructed to walk to another room in the building. On the way they passed by a table at which they were presented with a choice between a caloric slice of cake or a bowl of fruit salad. More than half (59 percent) chose the cake in the high load (seven-digit) condition, but only 37 percent chose the cake in the low load (two-digit number) condition. This finding is consistent with the idea that the effort required to memorize seven-digit numbers drew deliberative resources away from self-control, leading those subjects who had to remember more to eat more cake.

Second, prior exercise of self-control seems to diminish the capacity and, hence, propensity to exert self-control in the present. Recall that the prefrontal cortex is the part of the brain that is associated with a subjective feeling of effort. It is tempting to attribute this to the fact that self-control involves the same part of the brain—the executive prefrontal cortex—that is itself associated with feelings of mental effort. Perhaps this is why exercising willpower feels so difficult, and why exercising self-control in one domain can undermine its exercise in another, as demonstrated by a series of clever experiments conducted by Baumeister and colleagues (see, e.g., Baumeister and Kathleen Vohs 2003). In a typical study, subjects on diets who resisted temptation (by foregoing the chance to grab snacks from a nearby basket) later ate more ice cream in an ice-cream taste test and also quit earlier when confronted with an intellectual problem they couldn't solve. They acted as if their ability to resist temptation was temporarily "used up" by resisting the snacks (or, alternatively, that they had "earned" a reward of ice cream by skipping the tempting snacks). Other factors that seem to undermine this self-control resource are alcohol, stress, and sleep deprivation.

Turning to the other side of the equation, activation of affective states should, by the

same token, tend to accentuate temporal myopia. Indeed, there is considerable evidence of such effects (Janet Metcalfe and Mischel 1999). Research has shown, for example, that addicts display higher discount rates, not only for drugs but also for money, when they are currently craving drugs than when they are not (Giordano et al. 2002). Other research has shown that sexual arousal produces greater time-discounting of money rewards (Ariely and Loewenstein 2003).

The main exceptions to the rule that affective states tend to engender short-sighted behavior involve interactions between the cognitive and affective systems. In fact, decisions to delay gratification often involve a mixture of affect and cognition. They require a cognitive awareness of the delayed benefits in delaying gratification—e.g., that desisting from eating cake today will mean a more pleasing body type in the future. But, as many researchers have observed, cognitive awareness alone is insufficient to motivate delay of gratification; emotions play a critical role in forward-looking decision making. As Barlow (1988) notes, "The capacity to experience anxiety and the capacity to plan, are two sides of the same coin."
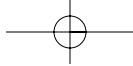
Thomas Cottle and Stephen Klineberg (1974), similarly argue that people care about the delayed consequences of their decisions only to the degree that contemplating such consequences evokes immediate affect. In support of this view, they cite the effects of frontal lobotomies which harm areas of the brain that underlie the capacity for images of absent events to generate experiences of pleasure or discomfort. The neurosurgeons who performed these operations wrote of their frontal-lobotomy patients that: "the capacity for imagination is still present, and certainly not sufficiently reduced to render the patients helpless, and affective responses are often quite lively, [but there is] a separation of one from the other" (Freeman and Watts 1942).

The work of Damasio and colleagues discussed earlier (Bechara, Damasio, Daniel Tranel, and Damasio 1997; Damasio 1994) lends further credence to this perspective, as does research on psychopaths, who are characterized by both emotional deficits when it comes to imagining the future and by insensitivity to the future consequences (as well as consequences to others) of their behavior (Hervey Cleckley 1941; Hare 1965, 1966; David Lykken 1957). Whether deliberately, as when one conjures up an image of a "fat self" exposed on the beach, or without conscious intention, self-control often involves an interaction of affective and cognitive mechanisms.

How might one model intertemporal choice differently as a result of the insights from neuroscience? First, the neuroscience research points to ways to "unpack" the concept of time preference. Clearly, ability to think about future consequences is important, which is probably why time preference is correlated with measured intelligence (Mischel and Robert Metzner 1962). Second, because people are likely to make myopic choices when under the influence of powerful drives or emotions (Loewenstein 1996), this suggests that a key to understanding impulsivity in individuals might be to understand what types of situations get them "hot." Third, we might be tempted to look for individual differences in what could be called "willpower"—i.e., the availability of the scarce internal resource that allows people to inhibit viscerally driven behaviors (see Loewenstein and O'Donoghue 2004 for a recent two-system model of intertemporal choice and other economic behaviors).

A model of intertemporal choice that took account of interactions between affect and cognition can help to explain not only impulsivity, but also why many people have self-control problems of the opposite type of those typically examined in the literature— e.g., tightwads who can't get themselves to spend enough; workaholics who can't take a break, and people who, far from losing control in the bedroom, find themselves frustratingly unable to do so. All of these patterns

of behavior can easily be explained by the possibly uniquely human propensity to experience emotions, such as fear, as a result of thinking about the future. Indeed, it is likely that one of the main tools that the prefrontal cortex uses to impose self-control when affective forces would otherwise favor short-sighted self-destructive behavior is to create "deliberative affect" via directed imagery and thought (Roger Giner-Sorolla 2001).

Such a framework might also help to explain why people appear so inconsistent when their behavior is viewed through the lens of discounted utility. The ability to think about future consequences may not be strongly correlated with the degree to which different experiences produce visceral reactions, and these in turn might not be correlated with an individual's level of willpower. Indeed, Loewenstein et al. (2001) found close to zero correlations between numerous behaviors that all had an important intertemporal component, but much higher correlations between behaviors that seemed to draw on the same dimension of intertemporal choice—e.g., which required suppression of specific emotions such as anger.
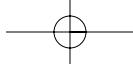
5.1.3 *Automatic Processes in Intertemporal Choice*

To the extent that intertemporal choice is, in fact, driven by cognitive considerations, much of this cognition does not take the usually assumed form of a weighing of costs and benefits discounted according to when they occur in time. Rather, consistent with the earlier claim that people often make decisions via a two-part process that takes the form of first asking "what situation am I in?" then continuing "how does one behave in such a situation," much intertemporal choice is driven by automatic processes involving pattern-matching, recognition, and categorization.

A pattern of choice that seems to be driven by such a process is the preference for sequences of outcomes that improve over time. In one demonstration of this effect, Loewenstein and Prelec (1993) asked some

subjects to choose whether to eat a fancy French Restaurant dinner in one month or two, and others to choose between the sequence [fancy French this month, eat at home next month] or the same choice in reverse order. A majority of subjects preferred the isolated French dinner earlier but deferred the French dinner when it was embedded in a sequence with eating at home. Frederick and Loewenstein (2000) conducted a study in which subjects were presented with a series of intertemporal choices that were framed in ways intended to evoke different considerations. For example, in one version they asked respondents to allocate pleasurable outcomes, such as massages, over time. They expected, and found, that the allocation format evoked a *choice heuristic* that caused people to spread consumption relatively evenly over time, implying a preference for flat sequences. In another version, they asked respondents to state a maximum buying price for the Greek-then-French and French-then-Greek sequences, rather than choose between them, anticipating that the mention of money would evoke considerations of the time value of money and, hence, cause subjects to place higher value on the option that provided greater value earlier—i.e., the declining sequences—which is what they found.[20]

---

[20] Another phenomenon that may be driven by such automatic processes is the "diversification bias" (Itamar Simonson 1989). When people choose several alternatives from a set, they choose more variety when they choose them all simultaneously than when they choose them sequentially. This phenomenon has been demonstrated with snack food, audio pieces, gambles, and lottery tickets. Daniel Read and Loewenstein (1995) tested various explanations for overdiversification (e.g., they diversify in simultaneous choice to gather information), but concluded that it results from a rule of thumb that that they apply whenever choices are expressed in a fashion that highlights diversification (see also Thomas Langer and Craig Fox 2004). This viewpoint is supported by a study (Read, Loewenstein, and S. Kalyanaraman 1999) in which subjects make successive choices between groups of objects which are easily categorized or not. When categorization was easy (e.g., virtues and vices), subjects diversified more in simultaneous than in sequential choice. When there were multiple competing categorizations, the overdiversification bias disappeared.

In sum, neuroscience points to some deficiencies in the way that economists currently model intertemporal choice and also suggests directions for future modeling. A somewhat stylized interterpretation of the results just reviewed would be that some intertemporal decisions are, in fact, well represented by the discounted utility model—specifically those involving detailed deliberation but minimal affect. However, a wide range of other intertemporal choices are influenced by affectively "hot" processes such as drives and emotions, or result from processes that automatically evoke a response which depends on the situation. Models which focus on how these discrepant processes interact are promising (e.g., Bernheim and Rangel 2004; Loewenstein and O'Donoghue 2004).

5.2 *Decision-Making under Risk and Uncertainty*

Both collaboration and competition between affect and cognition, and between controlled and automatic processes, can also be seen in the domain of decision making under risk and uncertainty.

5.2.1 *Affect versus Cognition*
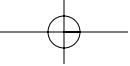
The expected utility model views decision making under uncertainty as a tradeoff of utility under different states of nature—i.e., different possible scenarios. But, much as they do toward delayed outcomes, people react to risks at two different levels. On the one hand, as posited by traditional economic theories and consistent with quadrant I of table 1, people do attempt to evaluate the objective level of risk that different hazards could pose. On the other hand, and consistent with quadrant IV, people also react to risks at an emotional level, and these emotional reactions can powerfully influence their behavior (Loewenstein, Weber, Christopher Hsee, and Ned Welch 2001).

The existence of separate affective and cognitive systems that respond differently to risks is most salient when the two systems clash. People are often "of two minds" when

it comes to risks; we drive (or *wish* we were driving as we sit white-knuckled in our airplane seat) when we know at a cognitive level that it is safer to fly. We fear terrorism, when red meat poses a much greater risk of mortality. And, when it comes to asking someone out on a date, getting up to speak at the podium, or taking an important exam, our deliberative self uses diverse tactics to get us to take risks, or to perform in the face of risks, that our visceral self would much prefer to avoid. Perhaps the most dramatic illustrations of the separation of visceral reactions and cognitive evaluations, however, comes from the phobias that so many people suffer from; the very hallmark of a phobia is to be unable to face a risk that one recognizes, objectively, to be harmless. Moreover, fear unleashes preprogrammed sequences of behavior that aren't always beneficial. Thus, when fear becomes too intense it can produce counterproductive responses such as freezing, panicking, or "dry-mouth" when speaking in public. The fact that people pay for therapy to deal with their fears, and take drugs (including alcohol) to overcome them, can be viewed as further "evidence" that people, or more accurately, people's deliberative selves, are not at peace with their visceral reactions to risks.

5.2.2 *Affective Reactions to Uncertainty*

A lot is known about the neural processes underlying affective responses to risks. Much risk averse behavior is driven by immediate fear responses to risks, and fear, in turn, seems to be largely traceable to the amygdala. The amygdala constantly scans incoming stimuli for indications of potential threat and responds to inputs both from automatic and controlled processes in the brain. Patrik Vuilleumier et al. (2001) observed equivalent amygdala activation in response to fearful faces that were visually attended to or in the peripheral region which falls outside of conscious perception (cf. de Beatrice de Gelder, Jean Vroomen, Gilles Pourtois, and Lawrence Weiskrantz 1999;

Nouchine Hadjikhani and de Gelder 2003; LeDoux 1996; John Morris, C. Buchel, and Raymond Dolan 2001; Whalen et al. 1998). But the amygdala also receives cortical inputs, which can moderate or even override its automatic quadrant IV response.

In a paradigmatic experiment that illustrates cortical overriding of amygdala activation (LeDoux 1996), an animal such as a rat is "fear-conditioned"—by repeatedly administering a signal such as a tone followed by administration of a painful electric shock. Once the tone becomes associated in the animal's mind with the shock, the animal responds to the tone by jumping or showing other over signs of fear. In the next phase of the experiment, the tone is played repeatedly without administering the shock, until the fear response becomes gradually "extinguished." At this point, one might think that the animal has "unlearned" the connection between the tone and the shock, but the reality is more complicated and interesting. If the neural connections between the cortex and the amygdala are then severed, the original fear response to the tone reappears, which shows that fear conditioning is not erased in "extinction" but is suppressed by the cortex and remains latent in the amygdala. This suggests that fear learning may be permanent, which could be an evolutionarily useful adaptation because it permits a rapid relearning if the original cause of the fear reappears. Indeed, other studies (reported in the same book) show that fear-conditioning can be reinstated by the administration of a single shock.

Decision making under risk and uncertainty, like intertemporal choice, nicely illustrates both collaboration and competition between systems. When it comes to collaboration, risk taking (or avoiding) behavior involves an exquisite interplay of cognitive and affective processes. In a well-known study that illustrates such collaboration (Bechara et al. 1997), patients suffering prefrontal damage (which, as discussed above, produces a disconnect between cognitive and affective systems) and normal subjects chose a sequence of cards from four decks whose payoffs the subjects only learned from experience (a "multiarmed bandit" problem). Two decks had more cards with extreme wins and losses (and negative expected value); two decks had less extreme outcomes but positive expected value. Both groups exhibited similar skin conductance (sweating—an indication of fear) after large-loss cards were encountered, but, compared to normals, prefrontal subjects rapidly returned to the high-paying risky decks after suffering a loss and, as a result, went "bankrupt" more often. Although the immediate emotional reaction of the prefrontal patients to losses was the same as the reaction of normals (measured by skin conductance), the damaged patients apparently do not store the pain of remembered losses as well as normals, so their skin conductance rose much less than normals when they resampled the high risk decks. Subsequent research found a similar difference between normal subjects who were either high or low in terms of emotional reactivity to negative events. Those who were more reactive were more prone to sample from the lower-paying, safer decks of cards (Peters and Slovic 2000).

Damasio et al.'s research shows that insufficient fear can produce nonmaximizing behavior when risky options have negative value. But, it is well established that fear can also discourage people from taking advantageous gambles (see, e.g., Uri Gneezy and Jan Potters 1997). Indeed, Shiv et al. (2005) found that frontal patients actually make more money on a task in which negative emotions cause normal subjects to be extremely risk averse: a series of take-it-or-leave-it choices to play a gamble with a 50 percent chance of losing $1.00 or gaining $1.50. Normal subjects and frontal subjects were about equally likely to play the gamble on the first round, but normals stopped playing when they experienced losses, while frontal patients kept playing. Clearly, having frontal damage undermines

the overall quality of decision making; but there are situations in which frontal damage can result in superior decisions.

At a more macro level, emotional reactions to risk can help to explain risk-seeking as well as risk-aversion (Caplin and Leahy 2001). Thus, when gambling is pleasurable, a model that incorporates affect naturally predicts that people will be risk-seeking and that self-control will be required to rein in risk-taking. Indeed, about 1 percent of the people who gamble are diagnosed as "pathological"—they report losing control, "chasing losses," and harming their personal and work relationships by gambling (National Academy of Sciences 1999). The standard economic explanations for gambling—convex utility for money or a special taste for the act of gambling—don't help explain why some gamblers binge and don't usefully inform policies to regulate availability of gambling. Neuroscience may help. Pathological gamblers tend to be overwhelmingly male and tend to also drink, smoke, and use drugs much more frequently than average. Genetic evidence shows that a certain gene allele (D2A1), which causes gamblers to seek larger and larger thrills to get modest jolts of pleasure, is more likely to be present in pathological gamblers than in normal people (David Comings 1998). One study shows tentatively that treatment with naltrexone, a drug that blocks the operation of opiate receptors in the brain, reduces the urge to gamble (e.g., Paula Moreyra et al. 2000). The same drug has been used to successfully treat "compulsive shopping" (Susan McElroy et al. 1991).
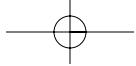
Neural evidence also substantiates the distinction between risk (known probability) and "Knightian" uncertainty, or ambiguity. Subjects facing ambiguous gambles—knowing they lack information they would like to have about the odds—often report a feeling of discomfort or mild fear. Brain imaging shows that different degrees of risk and uncertainty activate different areas of the brain (McCabe et al. 2001; Aldo Rustichini

et al. 2002) which corroborates the subjects' self-reports. Using fMRI, Ming Hsu et al. (2005) found frontal insula and amygdala activation when subjects faced ambiguous choices, compared to risky ones. They also found that patients with orbitofrontal cortical (OFC) lesions are ambiguity-neutral, compared to brain-damaged controls. Since the OFC receives input from the limbic system (including the insula and amygdala), the fMRI and lesion evidence together imply that in normal subjects, ambiguous gambles often create discomfort or fear which is transmitted to the OFC. Ironically, patients with OFC brain damage therefore behave more "rationally" (treating ambiguous and risky gambles similarly) than normals, a reminder that logical principles of rationality and biological adaptations can be different (cf. Shiv et al. 2005).

5.2.3 *Automatic Versus Controlled Processes*

The divergence between different systems' evaluations of risk can also be seen when it comes to judgments of probability. Numerous studies by psychologists have observed systematic divergences between explicit judgments of probability in different settings (presumably the product of controlled processing) and implicit judgments or judgments derived from choice (which are more closely associated with automatic processing and/or emotion). For example, Kirkpatrick and Epstein (1992) found that people prefer to draw a bean from a bowl containing ten winning beans and ninety losing beans than from a bowl containing one winning bean and nine losing beans (see also Veronika Denes-Raj, Epstein, and Jonathon Cole 1995; Paul Windschitl and Gary Wells 1998). Subjects say that they *know* the probabilities of winning are the same, but they still have an automatic quadrant III preference for the bowl with more winning beans.

An important feature of good probability judgment is logical coherence: probabilities

of mutually exclusive and exhaustive events should add to one, and conditional probabilities should be linked to joint and marginal probability according to Bayes rule $(P(A|B)=P(A \text{ and } B)/P(B))$. Logical coherence is violated in at least two neurally interesting ways. One is "conjunction fallacy"—the tendency to judge events with two components A and B as more likely than A or B alone. While most subjects (even statistically sophisticated ones) make conjunction errors on some problems, when those errors are pointed out, quadrant I wakes up, and the subjects sheepishly recognize the error and correct it (Kahneman and Frederick 2002). For example, the famous "Linda" problem describes an earnest young politically minded student. In one condition, subjects are asked to rank statements about Linda—Is she a bank teller? A feminist bank teller? A large majority of subjects, even highly educated ones, say Linda is more likely to be a feminist bank teller than she is to be a bank teller, a violation of the conjunction principle. But when subjects are asked, "Out of one hundred people like Linda, how many are bank tellers? Feminist bank tellers?" conjunction errors disappear (Tversky and Kahneman 1983).

Another violation is that subjects often report probabilities which are logically incoherent. fMRI evidence suggests an explanation for why probability judgments are incoherent, but can be corrected upon reflection: when guessing probabilities, the left hemisphere of the brain is more active; but when answering logic questions, the right hemisphere is more active (Lawrence Parsons and Daniel Osherson 2001). Since enforcing logical coherence requires the right hemisphere to "check the work" of the left hemisphere, there is room for slippage.

Once again, it can be seen that neuroscience, and specifically, a consideration of affective and automatic processes that have been largely neglected by economists, could potentially inform an important line of research and theory. We are skeptical of whether any theory that fails to incorporate the affective dimensions of risk will be capable of shedding much light on such important phenomena as stock market booms and busts, the ubiquity of gambling (e.g., slot machines revenues dwarf revenues from movies), and the vicissitudes of public responses to threats as diverse as terrorism and global warming.

5.3 *Game Theory*

Neuroscientific data are well-suited to exploring the central assumptions on which game theory predictions rest. These assumptions are that players: (1) have accurate beliefs about what others will do (i.e., players are in equilibrium); (2) have no emotions or concern about how much others earn (a useful auxiliary assumption); (3) plan ahead; and (4) learn from experience.

5.3.1 *Theory of Mind and Autism*

In strategic interactions (games), knowing how another person thinks, and how another person thinks *you* think, etc., is critical to predicting the other person's behavior (and for inferring the other player's intentions, which underlie emotional judgments of fairness and obliged reciprocity in more modern theories (e.g., Rabin 1993). From a neural view, iterated strategic thinking consumes scarce working memory and also requires a player to put herself in another player's "mind." There may be no generic human capacity to iterate this kind of thinking beyond a couple of steps.[21] Studies that examine either subject's choices, or that monitor what type of information subjects look up or pay attention to in experimental games, suggest only one–two steps of strategic thinking are typical in most populations (e.g., Eric Johnson et al. 2002; Miguel Costa-Gomes et al. 2001; Camerer, Teck Ho, and Kuan Chong 2004), though up to

---

[21] Of course, equilibration might occur through a process other than introspection—e.g., adaptive learning, imitation, communication, or evolution. But these processes must have a neural basis too.

three–four steps are observed in analytically skilled and specially trained populations.

As discussed in section 4.2, many neuroscientists believe there is a specialized "mind-reading" (or "theory of mind") area, perhaps in prefrontal area Brodmann 10, which generates reasoning about what others believe and might do (e.g., Simon Baron-Cohen 2000).[22] Autism is thought to be a deficit in this area (and related circuitry). People with autism often have trouble figuring out what other people think and believe, and are consequently puzzled by behavior that most people would consider normal.

One tool used in behavioral game theory is the "ultimatum game." In this game, a "proposer" offers a division of a sum of money, generically $10.00, to another "responder" who can accept or reject it, ending the game. If the responder has no emotional reaction to the fact that the proposer is earning more than she is ("envy" or "disgust"), then the responder should accept the smallest offer. If the proposer also has no emotional reaction to earning more ("guilt") and anticipates correctly what the responder will do, then the proposer should offer the lowest amount. This pattern is rarely observed: Instead, in most populations the proposer offers 40–50 percent and about half the responders reject offers less than 20 percent.

When players *do* follow the dictates of game theory, the result can be a low payoff and confusion. Consider this quote from an upset subject, an Israeli college student, whose low offer in a $10.00 ultimatum game was rejected (from Shmuel Zamir 2000):

> I did not earn any money because all the other players are *stupid*! How can you reject a positive amount of money and prefer to get zero? They just did not understand the game! You should have stopped the experiment and explained it to them . . .
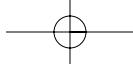
[22] Grether et al. (2004) also find BA 10 activity, along with interesting activity in anterior cingulate and the basal forebrain, as subjects bid in second-price auctions.

Ironically, while the subject's reasoning matches exactly how conventional game theory approaches the game, it also sounds *autistic*, because this subject is surprised and perplexed by how normal people behave.

This anecdote is complemented by Elizabeth Hill and David Sally's (2003) extensive comparison of normal and autistic children and adults playing ultimatum games. Nearly half of the autistic children offer zero or one unit (out of ten) in the ultimatum game, and relatively few offer half. Many autistic adults also offered nothing, but a large number of autistic adults offered half—as if they had developed a reasoning or experiential workaround which tells them what other people think is fair in games that involve sharing, even though they cannot guess what others will do using normal circuitry.

McCabe et al. (2001) used fMRI to measure brain activity when subjects played games involving trust, cooperation, and punishment. They found that players who cooperated more often with others showed increased activation in Brodmann area 10 (thought to be one part of the mind-reading circuitry) and in the thalamus (part of the emotional "limbic" system). Players who cooperated less often showed no systematic activation.

Bhatt and Camerer (in press) used fMRI to compare brain activity when subjects make choices and express beliefs in matrix games. They found that when players were in equilibrium (choices were best-responses and beliefs were accurate), the same circuitry was being used when making a choice and expressing a belief. This means equilibrium is a "state of mind," which can be identified by a tight overlap in activity in the two tasks, as well as a mathematical restriction on best response and belief accuracy. They also found evidence suggesting that, when subjects guessed what beliefs other subjects had about their own behavior, they tended to anchor on their own choices. This means

beliefs about beliefs are not just iterations of some belief processing mechanism; instead, self-referential beliefs use components of circuitry for forming beliefs and for making choices.

5.3.2 *Emotions and Visceral Effects*

One of the most striking neuroscientific findings about game theory comes from Sanfey et al.'s (2003) fMRI study of ultimatum bargaining. By comparing the brains of subjects responding to unfair ($1.00–$2.00 out of $10.00) and fair ($4.00–5.00) offers, they found that very unfair offers differentially activated three regions: Dorsolateral prefrontal cortex (DLPFC), anterior cingulate (ACC), and insula cortex (see Figure 4). DLPFC is an area involved in planning. The insula cortex is known to be activated during the experience of negative emotions like pain and disgust. ACC is an "executive function" area which often receives inputs from many areas and resolves conflicts among them.[23] Therefore, it appears that, after an unfair offer, the brain (ACC) struggles to resolve the conflict between wanting to accept the money because of its planned reward value (DLPFC) and disliking the "disgust" of being treated unfairly (insula).

In fact, whether players reject unfair offers or not can be predicted rather reliably (a correlation of 0.45) by the level of their insula activity. It is irresistible to speculate that the insula is a neural locus of the distaste for inequality or unfair treatment posited by models of social utility, which have been successfully used to explain many varying patterns in experiments — robust ultimatum rejections, public goods contributions, and trust and gift-exchange (e.g., Bazerman, Loewenstein, and Leigh Thompson 1989;

Ernst Fehr and Simon Gachter 2000; Camerer 2003). The fact that unfair offers activate insula means that a verbal statement like "I am so disgusted about being treated that way" is literal, not metaphorical—they really *do* feel disgusted.[24]

Zak et al. (2003) explored the role of *hormones* in trust games. In a canonical trust game, one player can invest up to $10.00, which is tripled. A second "trustee" player can keep or repay as much of the tripled investment as they want. Zak et al. measured eight hormones at different points in the trust game. The hormone with the largest effect was oxytocin—a hormone that rises during social bonding (such as breast-feeding and casual touching). They found that oxytocin rose in the trustee if the first player "trusts" her by investing a lot. (They also found that ovulating women were particularly untrustworthy—they did not repay as much of the investment.)

Roxanna Gonzalez and Loewenstein (2004) examined the impact of circadian rhythms in a repeated trust (centipede) game. They sorted people into "morning" and "night" people (which can be done rather reliably) and had them play a centipede game when they were on- or off-peak (e.g., the morning people were off-peak when playing in the evening). Based on prior research showing that sleep-cycle affects emotional regulation—i.e., people's ability to suppress or avoid acting on unwanted feelings—they predicted and found much lower levels of cooperative behavior when people played at off-peak times.

Tania Singer et al. (2004) report an important link between reward and behavior in games. They played repeated prisoners' dilemma games in which one player, in the fMRI scanner, faced a series of opponents.

---

[23] The ACC also contains a large concentration of "spindle cells"—large neurons shaped like spindles, which are almost unique to human brains (Allman et al. 2002). Loosely speaking, these cells are probably important for many of the activities which distinguish humans from our primate cousins, particularly language, and complex decision making.

[24] It also suggests an intriguing follow-up experiment that no previous theory would have predicted—patients with damage to their insula regions should feel no disgust and accept low offers, unless those patients have developed a "workaround" or alternative method to "feel" unfairness in nonvisceral terms.
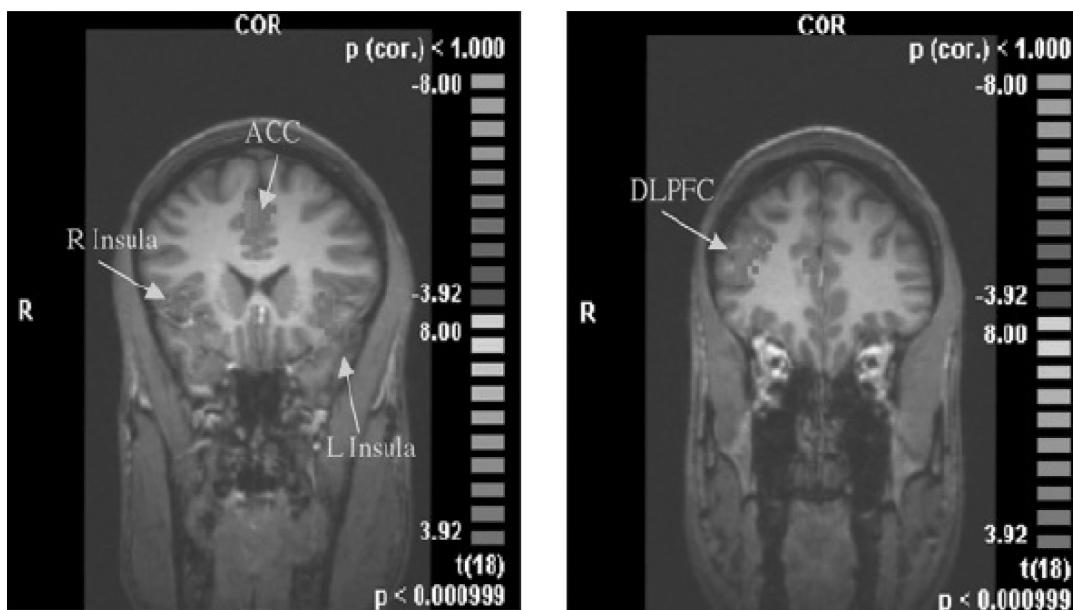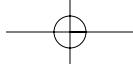
Figure 5. Coronal slices showing regions which are differentially active after an unfair offer ($1–2 out of $10), relative to activity after a fair off ($4–5). Regions are anterior cingulate (ACC), right and left insula, and dorsolateral prefrontal cortex (DLPFC). See Sanfey et al. (2003).

The scanned subject was told that some opponents cooperated intentionally (i.e., they could choose freely) while others cooperated, but unintentionally. Afterwards, subjects were shown different faces of those they had played against. The faces of the intentional cooperators activated insula, amygdala, and ventral striatal areas (among others). Since the striatum is an all-purpose reward area, activation in that region means that simply seeing the face of a person who intentionally cooperated with you is rewarding. In game theory terms, a person's "reputation" in a repeated game is a perception by other players of their "type" or likely behavior based on past play. Singer et al.'s results mean that a good reputation may be neurally encoded in a way similar to beautiful or other rewarding stimuli.

The facts that insula activity, oxytocin levels, and sleep cycles affect behavior in games, and that cooperators' faces "feel beautiful," does not "disprove" game theory, per se, because preferences for different outcomes, and reasoning ability, might legitimately vary with these biological factors. The theory is easily patched by inserting variables, like "an envy/disgust coefficient," which depends on some biological state. But game theory also assumes that players will recognize the state-dependence in others and adjust their guesses about how other people will play. We have no idea if they can, and it is likely that people are generally limited at simulating emotional states of others (see Leaf van Boven, Loewenstein, and David Dunning 2003).

Cognitive inaccessibility also implies that people may not fully understand the influence of exogenous changes in visceral states on their own behavior. For example, if being trusted produces oxytocin, then when oxytocin surges for exogenous reasons—from a relaxing massage, or when synthetic oxytocin is administered—the brain might misread this surge in oxytocin as a sign of being trusted and react accordingly (e.g., by acting in a more reciprocal trustworthy way).

5.3.3 *Backward Induction*

A central principle in game theory is "backward induction" in extensive-form ("tree") games that are played over time. Backward induction means figuring out what to do today by and reasoning how others will behave at all possible future points and working backward. Behavioral evidence, and direct evidence from measuring where players look on a computer screen, shows that people have trouble doing more than a couple of steps of backward induction (e.g., Johnson et al. 2002). However, Johnson et al. also found that, when players were briefly instructed about how to do backward induction, they could learn to do it rapidly and with little effort (total response times were similar to in pre-instruction trials). This is a reminder, in the game theory context, of the important distinction between controlled and automatic behavior stressed in section 3 above. Initially, subjects look automatically at early periods and hardly notice future periods they realize (correctly, in a statistical sense) are unlikely to occur. However, with instruction and practice, backward induction quickly becames automated, yielding fast responses and few errors.

Economists are naturally inclined to model cognition in terms of costs and benefits. If forced into this framework, backward induction would probably be characterized as cognitively costly. But the fact that it is easily learned and automated suggests the costs of backward induction has a special structure— at first, it is unnatural (not spontaneously intuited by subjects), like a piece of software not yet installed; but once installed, it is cheap to run.

5.3.4 *Learning*

The idea that a game-theoretic equilibrium resulted from learning, imitation, or evolution, rather than simple introspection, has led to a large literature on what results in the long-run from different types of learning models (e.g., Drew Fudenberg and David Levine 1998; George Mailath 1998). Since many types of learning rules have been proposed, fitting rules to data from experiment has proved useful in showing when intuitively appealing rules might fit badly, and suggesting improvements to those rules (e.g., Camerer 2003). Camerer and Ho (1999) showed that both simple reinforcement of chosen strategies, and learning by updating beliefs about other players, are really two polar opposite types of generalized reinforcement learning in which strategies have numerical propensities or attractions which are adjusted over time by experience.

In neural terms, the Camerer–Ho theory can be interpreted as a splice of two processes—a rapid, emotional process in which a chosen strategy is quickly reinforced by the gain or loss that resulted and a slower deliberative process that requires players to create counterfactuals about how much they would have earned from other strategies that were not chosen. Conventional reinforcement learning neglects the second process. "Fictitious play" belief learning assumes the second process completely overrides the first. A parameter in the theory which represents the relative strength of the second process, compared to the first, is usually estimated to be between zero and one. This implies that reinforcement is stronger, but both processes are at work.

Michael Platt and Paul Glimcher (1999) found corroborating evidence for reinforcement learning of the first, rapid-process sort in single-neuron recording in monkey parietal cortex. They measured neuron firing rates in advance of choices in a game between a monkey and a computerized opponent. They found that firing rates are closely related to the average reinforcement received for that choice in the last ten trials. Dominic Barraclough, Michelle Conroy, and Daeyeol Lee (2004) found similar evidence of "learning neurons" in dorsolateral prefrontal cortex of rhesus monkeys; their parameter estimates of learning models also supports the two-process Camerer–Ho
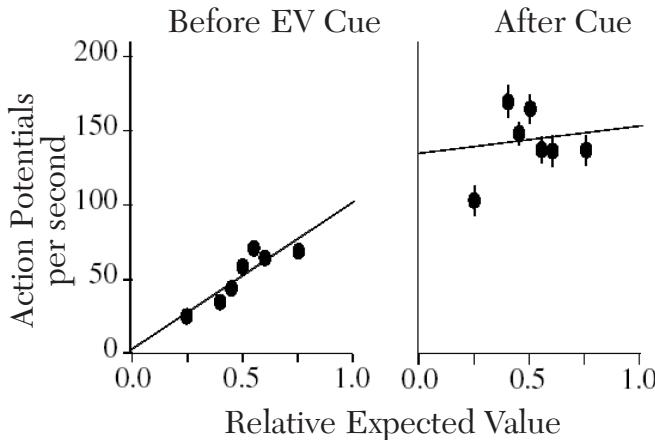
Figure 6. (from Glimcher et al., in press). The left graph ("before EV cue") shows the LIP neuron firing rate (y axis) plotted against the relative expected value of one of two possible reward-producing locations. The right graph ("after cue") shows the firing rate after the monkey learns which movement is rewarding.

theory.[25] Gathering neural evidence from humans (to see whether a second, deliberative process is being used as the parametric estimates suggest) is feasible and could prove insightful.

Figure 6 illustrates the behavior of Glimcher's parietal neurons in a task with two targets of different expected value. The left graph shows the actual firing rates (on the y axis) plotted against the expected reward value of the target before the "best" target is revealed. The correspondence is remarkably close—*these neurons are encoding expected value*. The righthand graph shows that, after the winning target is revealed (so that the expected value is no longer relevant), the firing rates increase to close to their maximum capacity.
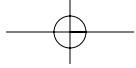
In sum, neuroscience provides some new ways to think about central elements of game

theory. Strategic thinking is likely to require specialized circuitry and iterated strategic reasoning is likely to be bounded (as indicated by ample behavioral data). Early studies suggest hormones and biological factors (such as circadian rhythms) play a role in social preferences like trust. Thinking about the brain—and some experimental facts— also suggest that logical principles like backward induction may be neurally unnatural and that learning processes are likely to be a splice of cognitive and affective processes. At the same time, single-neuron monkey measurement suggests that expected-value computation and reinforcement learning may control behavior in highly adapted tasks. The emerging picture is one in which the simplest elements of game theory, like keeping track of what has worked in a mixed-equilibrium game with limited scope for outguessing one's opponent, may be alive and well in the brain, while higher-order cognition and affective influence on social preferences depart from the standard ideas we teach our students.

5.4 *Labor-Market Discrimination*

Our last specific application is labor-market discrimination. Economic models assume

---

[25] They estimate two reinforcements: When the monkeys choose and win (reinforcement by Δ_1), and when they choose and lose (Δ_2). In their two-strategy games, the model is equivalent to one in which monkeys are not reinforced for losing, but the unchosen strategy is reinforced by Δ_2. The fact that Δ_2 is usually less than Δ_1 in magnitude (see also Lee et al., in press) is equivalent to $\delta < 1$ in the Camerer–Ho theory.
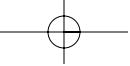
that labor-market discrimination against minorities is either a taste (a distaste for working with minorities, or a distaste passed on from customers), or a belief that minority workers are less productive (for example, a belief that minority status is a proxy for unobservable differences in skill, also known as "statistical discrimination").

Neuroscience suggests a different answer. Automaticity contributes to discrimination because neural networks rapidly spread activation through associated concepts and stereotypes. Affect contributes to discrimination because automatic affective reactions have such a powerful effect on cognitive judgments. Discrimination in this view involves rapid, automatic, associations between social categories, stereotypes, and affect.

Such an account receives support from remarkable experiments that demonstrated subtle "implicit associations" between demographic categories and good or bad adjectives (try it out on yourself at http://buster.cs.yale.edu/implicit/). Subjects taking this computer-administered implicit association test (IAT) are shown a mixed up series of stereotypically black or white names (Tyrone or Chip) and positive or negative adjectives (mother or devil). They are asked to tap one key when they see one type of name or adjective, and a different key if they see the other type of name or adjective, at which point the computer moves on to the next name or adjective. The dependent variable is how long it takes the subject to work their way through the complete list of names and adjectives. White subjects work their way through the list much more quickly when one key is linked to the pair (black or negative) and the other to (white or positive) than they are when one key is linked to [black or positive] and the other to [white or negative]. What's going on? The brain encodes associations in neural networks, which spread activation to related concepts. For white students, black names are instantly associated with negative concepts, whether they realize it or not, because the association is automatic (rapid and unconscious).

As the name suggest, the implicit association test taps into "implicit," as opposed to "explicit," attitudes. One can think of implicit attitudes, roughly, as those associated with automatic processing, whereas explicit attitudes are associated with conscious, controlled, processing. New methods developed by psychologists, such as the IAT, have begun to reveal that implicit and explicit attitudes can sometimes diverge from one-another, with implicit attitudes, in some cases, exerting a more reliable influence on behavior. Thus, in one recent study, Allen McConnell and Jill Leibold (2001) administered the IAT to subjects, had them complete measures of explicit attitudes toward blacks and whites, and also had them interact with two experimenters, one black and the other white. Coders blind to either the implicit or explicit measures then coded the subjects' interactions with the experimenters, including such objective measures as how closely the subject moved their chair to the experimenter, and the experimenter also rated their perception of the prejudice of the subject. Although the experimenters ratings of the subjects' degree of prejudice correlated with both explicit ($r = 0.33$, $p < 0.05$) and implicit ($r = 0.39$, $p < 0.05$) measures, the other behavioral measures of bias that correlated with either the IAT or self-reports of prejudice (these were: coders' overall ratings, speaking time, smiling, speech errors, speech hesitation, and extent of extemporaneous social comments) all correlated with the IAT but not with the explicit measures of prejudice.

Implicit attitudes have also been linked to neural processing. In one study (Elizabeth Phelps et al. 2000), researchers administered the IAT to Caucasian subjects and also asked them explicit questions about their attitudes toward African–Americans. Then they scanned the subjects' brains with fMRI while exposing them to photographs of unfamiliar black and white males, and focusing specifically on a subcortical structure in the brain called the amygdala, which numerous

studies have linked to processing of fear. They found that the relative strength of amygdala activation to black as compared with white faces was correlated with the IAT measure of implicit attitudes, but not with direct, conscious, expression of race attitudes. (A study by Hart et al. (2000) suggests the effect is roughly symmetric for black subjects seeing white faces.) Furthermore, the same pattern was not observed when the black and white faces were well known, positively regarded, celebrities (such as Michael Jordan and Denzel Washington). People hold implicit attitudes involving not only race, but also a wide range of other individual characteristics, such as height, weight, attractiveness, religion, and national origin, and, like race, many of these attitudes undoubtedly have real economic consequences. Thus, although most people reject height and attractiveness as indicators of marginal productivity, taller and more attractive people are more likely to earn higher wages (e.g., Nicola Persico, Andrew Postlewaite, and Dan Silverman 2002) and other rewards (e.g., the U.S. Presidency).

Does the implicit association view imply that labor-market discrimination is due to a taste, a statistical shortcut, or something else? One piece of evidence suggests the statistical interpretation is on the right track: the amygdala that is active when people see other-race faces seems to be sensitive to *familiarity* of faces, not race per se (Dubois et al. 1999). This is consistent with an interpretation of statistical discrimination in which employers do not *fear* minority workers, they are simply less sure of their abilities. But other evidence suggests discrimination is "something else." Amygdala activity can be dampened if subjects are shown pictures of black and white faces and asked to judge how much the pictured people like vegetables (Wheeler and Fiske 2005). This suggests that automatic reactions to race can be erased (or substituted for) depending on the question being asked when the faces are perceived. These reactions respond to variables other than prices and choice sets so it is a stretch to think of them of as conventional tastes or beliefs.
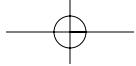
6. *Conclusions*

Economics parted company from psychology in the early twentieth century. Economists became skeptical that basic psychological forces could be measured without inferring them from behavior (the same position "behaviorist" psychologists of the 1920s reached), which led to adoption of the useful tautology between unobserved utilities and observed (revealed) preferences. But remarkable advances in neuroscience now make direct measurement of thoughts and feelings possible for the first time, opening the "black box" which is the building block of any economic interaction and system—the human mind.

Most economists are curious about neuroscience, but instinctively skeptical that it can tell us how to do better economics. The tradition of ignoring psychological regularity in making assumptions in economic theory is so deeply ingrained—and has proved relatively successful—that knowing more about the brain seems unnecessary. Economic theory *will* chug along successfully for the next few years paying no attention at all to neuroscience (just as it paid little attention to psychology until recently). But it is hard to believe that *some* neuroscientific regularities will not help explain *some* extant anomalies, particularly those that have been debated for decades.

Indeed, in many areas of economics there are basic constructs or variables at the heart of current debates which can be usefully thought of as neural processes, and studied using fMRI and other tools. For example, finance is a field awash in literally millions of observations of daily price movements. Despite having widespread access to terrific data, after decades of careful research there is no agreed-upon theory of why stock prices fluctuate, why people trade, and why there

are so many actively managed mutual funds despite poor fund performance. Perhaps knowing more about basic neural mechanisms that underlie conformity, attention paid to large price changes, wishful thinking, sense-making of random series, and perceptions of expertise can help explain these puzzles (e.g., Lo and Repin 2002). Furthermore, some scholars have argued that large fluctuations in stock prices are due to reasonable time-variation of risk premia. But there is no theoretical basis in finance for why attitudes toward risk would vary over time. Maybe neuroscience can supply one.

In labor markets, a major puzzle is why wages are so downward sticky. Firms say they are afraid to cut wages because they want to maintain worker morale (e.g., Truman Bewley 2002); and experiments show that when worker productivity is valuable, paying a high wage induces effort, even when workers are free to shirk (Fehr and Gachter 2000). Presumably morale is some combination of workers' emotional feelings toward their employer and may be very sensitive to recent experience, to what other workers think, to whether wage cuts are "procedurally just," and so forth. There is no reason these processes could not be described as neural processes and studied that way.

There are many anomalies in intertemporal choice. In the United States, credit card debt is substantial ($5,000.00 per household) and a million personal bankruptcies have been declared in each of several years (Laibson, Andrea Repetto, and Jeremy Tobacman 1998). Healthier food is cheaper and more widely available than ever before, but spending on dieting and obesity are *both* on the rise. Surely understanding how brain mechanisms process reward, and curb or produce compulsion, and their evolutionary origins (e.g., Trent Smith 2003) might help explain these facts and shape sensible policy and regulation.
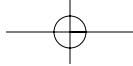
Prevailing models of advertising assume that ads convey information or signal a product's quality or, for "network" or "status" goods, a product's likely popularity. Many of these models seem like strained attempts to explain effects of advertising without incorporating the obvious intuition that advertising taps neural circuitry of reward and desire.

Finally, economic models do not provide a satisfying theory of how individuals differ. As laymen, we characterize other people as impulsive or deliberate, stable or neurotic, decisive or indecisive, mature or immature, foolish or wise, depressed or optimistic, scatterbrained or compulsively organized. The consumers who spend countless dollars on self-help, "organize your life" manuals, and who sustain the huge, and infinitely varied psychological counseling industry, are typically unhappy where they stand on some of these dimensions and are looking for ways to change. Comparative economic development, entrepreneurial initiative and innovation, business cycle sensitivity, and other important macroeconomic behaviors are probably sensitive to the distribution of these and other psychological "assets." Yet there is no complete way to discuss them with the language of beliefs and desires, which is the only language operating in quadrant I.

6.1 *Can Neuroscience Save Rational Choice Economics?*

Many neuroscientists are now using the most basic elements of rational choice theory to explain what they see. Ironically, they are taking up rational choice theory at the same time as more and more economists are moving away from rational choice toward a behavioral view anchored in limits on rationality, willpower, and greed (which we expect to be informed by neural detail). For example, neuroscientific studies of simple reward circuitry in rats and primates vindicate some of the simplest ideas of economics—namely, a "common brain currency" which permits substitution (Shizgal 1999), existence of neurons which encode expected reward (see Glimcher 2002 and Figure 3), and revealed

preference (macaques viewing socially valued images, see Deaner, Khera, and Platt 2005). Other groups are using Bayesian models. One study concluded that "the central nervous system therefore employs [Bayesian] probabilistic models during sensorimotor learning" (Konrad Körding and Daniel Wolpert 2004).

Our view is that establishing a neural basis for some rational choice principles will not necessarily vindicate the approach as widely applied to *humans*. The reason is that the studies which have most clearly established expected-value and Bayesian encoding use simple tasks which monkeys and humans are well-evolved to perform (e.g., reaching to earn a juice reward). It is quite possible that simple rational mechanisms are neurally instantiated to do these tasks well. But the most important kinds of economic behavior involve manipulating abstract symbols, thinking about groups of people and complex institutions, trading off very different types of rewarding objects across time, and weighing them by probabilities which are not always learnable by experience. Ironically, rational choice models might therefore be most useful in thinking about the simplest kinds of decisions humans and other species make— involving perceptual tradeoffs, motor movements, foraging for food, and so forth—and prove least useful in thinking about abstract, complex, long-term tradeoffs which are the traditional province of economic theory.

6.2 *Incremental Versus Radical Neuroeconomics*

How should neuroeconomics contribute to advances in economic analysis? In the short-run, an "incremental" approach in which psychological evidence suggests functional forms will help enhance the realism of existing models. For example, the two-parameter "ß–δ" hyperbolic discounting approach (where ß expresses preference for immediacy, and is equal to one in the standard model) is an example that has proven productive theoretically (e.g., Laibson,

Repetto and Tobacman 1998; O'Donoghue and Rabin 1999). Laibson's (2001) model of homeostatic response to environmental cues in addiction is an incremental model well grounded in recent neuroscience.

However, we believe that in the long run a more "radical" departure from current theory will become necessary, in the sense that the basic building blocks will not just consist of preferences, constrained optimization and (market or game-theoretic) equilibrium. After all, the point of constrained optimization is to model behavior precisely and predict how behavior changes in response to changes in budget constraints and prices. There is no reason other models starting from a very different basis could not be constructed, while also predicting behavioral responses to constraints and prices, *and* predicting responses to other variables as well. Furthermore, thinking about the brain does not so much "falsify" rational choice theories as suggest entirely new distinctions and questions.

For example, is learning that a particular person cooperated with you in a game encoded as a cognitive reputational statistic about that person, like a numerical test score, or a "warm glow" which produces a surge of dopamine when you see the person's face (Singer et al. 2004; cf. James Rilling et al. 2002). Standard game theory has no answer for this question. But the distinction matters. If reputations are encoded dopaminergically, then they may spill over across groups of people who look the same, for example, or who are perceived as being part of a group who behave similarly (Bill McEvily et al. 2003; Paul Healy 2004). Or the mechanism can work in reverse—since attractive faces are known to produce dopaminergic surges (Aharon et al. 2001), the cortex may mistake these positive signals of cooperativeness (which also produces pleasurable sensations) and automatically judge attractive faces as likely to be cooperative. This simple neural question about reward spillover might form the foundation of group affiliation and social

capital and have something to do with why some countries are rich and others poor.

Our main theme in this paper is that radical models should respect the fact that brain mechanisms combine controlled and automatic processes, operating using cognition and affect. The Platonic metaphor of reason as a charioteer, driving twin horses of passion and appetite, is on the right track—except reason has its hands full with headstrong passions and appetites.[26] Of course, the challenge in radical-style theorizing is to develop models of how multiple mechanisms interact which are precise. Can this be done? The answer is Yes. Bernheim and Rangel (2004), Loewenstein and O'Donoghue (2004), and Jess Benhabib and Alberto Bisin (2002) have all proposed recent models with interacting mechanisms much like those in our Table 1.

Furthermore, while interactions of multiple brain mechanisms might appear to be too radical a change from equilibrium with utility maximization, we think many familiar tools can be used to do radical neuroeconomics. Interactions of cognition and affect might resemble systems like supply and demand, or feedback loops which exhibit multiple equilibria. The interaction of controlled and automatic processes might be like an inventory policy or agency model in which a controller only steps in when an extreme state of the system (or unusual event) requires controlled processes to override automatic ones. The influence of affect on choices is a very general type of state-dependence (where the "state" is affective, and is influenced by external cues and also by internal deliberation and restraint). Instead of solving for equilibria in these interacting-mechanism models, solve for steady states or cyclic fluctuations. Instead of summarizing responses to changes by comparative statics, study impulse-response functions.

Although we focused solely on applications of neuroscience to economics, intellectual

trade could also flow in the opposite direction. Neuroscience is shot through with familiar economic language — delegation, division of labor, constraint, coordination, executive function — but these concepts are not formalized in neuroscience as they are in economics. There is no overall theory of how the brain allocates resources that are essentially fixed (e.g., blood flow and attention). An "economic model of the brain" could help here. Simple economic concepts, like mechanisms for rationing under scarcity, and general versus partial equilibrium responses to shocks, could help neuroscientists understand how the entire brain interacts. At a technical level, neuroscientists are using tools imported from econometrics, such as Granger causality (e.g., Wolfram Hesse et al. 2003), to draw better inferences from neural time series. Finally, as the center of gravity in neuroscience research shifts from elementary cognitive processes to the study of so-called higher functions—reasoning, social inference, and decision making—neuroscientists will increasingly reference, and draw inspiration from, the conceptual apparatus of economics, a unique distillate of our century-long reflection on individual and strategic behavior.

REFERENCES

Adolphs, Ralph, Daniel Tranel, Hanna Damasio, and Antonio R. Damasio. 1995. "Fear and the Human Amygdala." *Journal of Neuroscience*, 15(9): 5879–91.
Aharon, Itzhak, Nancy Etcoff, Dan Ariely, Chris F. Chabris, Ethan O'Connor, and Hans C. Breiter. 2001. "Beautiful Faces Have Variable Reward Value: fMRI and Behavioral Evidence." *Neuron*, 32(3): 537–51.
Allman, John, Atiya Hakeem, and Karli Watson. 2002. "Two Phylogenetic Specializations in the Human Brain." *Neuroscientist*, 8(4): 335–46.
Ameriks, John, Andrew Caplin, and John Leahy. 2003. "Wealth Accumulation and the Propensity to Plan." *The Quarterly Journal of Economics*, 118(3): 1007–47.
Anderson, Adam K., Kalina Christoff, David Panitz, Eve De Rosa, and John D. Gabrieli. 2003. "Neural Correlates of the Automatic Processing of Threat Facial Signals." *Journal of Neuroscience*, 23(13): 5627–33.
Ariely, Dan and George Loewenstein. 2003. "The Heat of the Moment: The Effect of Sexual Arousal on Decision Making," CMU Working Paper.
Ariely, Dan, Uri Gneezy, George Loewenstein, and

---

[26] Thanks to Nava Ashraf for clarifying this point.

Nina Mazar. 2004. "Large Stakes and Big Mistakes," CMU Working Paper.

Armony, Jorge L., David Servan-Schreiber, Jonathan D. Cohen, and Joseph E. LeDoux. 1995. "An Anatomically Constrained Neural Network Model of Fear Conditioning." *Behavioral Neuroscience*, 109(2): 246–57.

Armony, Jorge L., David Servan-Schreiber, Jonathan D. Cohen, and Joseph E. LeDoux. 1997. "Computational Modeling of Emotion: Explorations through the Anatomy and Physiology of Fear Conditioning." *Trends in Cognitive Sciences*, 1(1): 28–34.

Barberis, Nicholas, Ming Huang, and Tano Santos. 2001. "Prospect Theory and Asset Prices." *The Quarterly Journal of Economics*, 116(1): 1–53.

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

Bargh, John A., Shelly Chaiken, Paula Raymond, and Charles Hymes. 1996. "The Automatic Evaluation Effect: Unconditional Automatic Attitude Activation with a Pronunciation Task." *Journal of Experimental Social Psychology*, 32(1): 104–28.

Bargh, John A. and Tanya L. Chartrand. 1999. "The Unbearable Automaticity of Being." *American Psychologist*, 54(7): 462–79.

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

Baron-Cohen, Simon. 2000. "Theory of Mind and Autism: A Fifteen Year Review," in *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*. Simon Baron-Cohen, Helen Tager-Flusberg and Donald J. Cohen, eds. Oxford: Oxford University Press, 1–20.

Barraclough, Dominic J., Michelle L. Conroy, and Daeyeol Lee. 2004. "Prefrontal Cortex and Decision Making in a Mixed-Strategy Game." *Nature Neuroscience*, 7(4): 404–10.

Baumeister, Roy F., Todd F. Heatherton, and Dianne M. Tice. 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 *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. George F. Loewenstein, Daniel Read and Roy F. Baumeister, eds. New York: Russell Sage Foundation, 13–86.

Bazerman, Max, George Loewenstein, and Leigh Thompson. 1989. "Social Utility and Decision Making in Interpersonal Contexts." *Journal of Personality and Social Psychology*, 57(3): 426–41.

Bechara, Antoine, Antonio R. Damasio, Hanna Damasio, and Steven W. Anderson. 1994. "Insensitivity to Future Consequences Following Damage to Human Prefrontal Cortex." *Cognition*, 50(1-3): 7–15.

Bechara, Antoine, Hanna Damasio, Antonio R. Damasio, and G. P. Lee. 1999. "Different Contributions of the Human Amygdala and Ventromedial Prefrontal Cortex to Decision-Making." *Journal of Neuroscience*, 19(13): 5473–81.

Bechara, Antoine, Hanna Damasio, Daniel Tranel, and Antonio R. Damasio. 1997. "Deciding Advantageously Before Knowing the Advantageous Strategy." *Science*, 275(5304): 1293–95.

Becker, Gary and Luis Rayo. 2004. "Evolutionary Efficiency and Mean Reversion in Happiness." University of Chicago Graduate School of Business.

Benartzi, Shlomo and Richard H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *The Quarterly Journal of Economics*, 110(1): 73–92.

Benhabib, Jess and Alberto Bisin. 2002. "Self-Control and Consumption-Savings Decisions: Cognitive Perspectives." New York University.

Berkowitz, Leonard. 1999. "Anger," in *Handbook of Cognition and Emotion*. Tim Dalgleish and Mick J. Power, eds. New York: Wiley, 411–28.

Bernheim, B. Douglas and Antonio Rangel. 2004. "Addiction and Cue-Conditioned Cognitive Processes." *American Economic Review,* 94(5): 1558–90.

Berridge, Kent C. 1996. "Food Reward: Brain Substrates of Wanting and Liking." *Neuroscience and Biobehavioral Reviews*, 20(1): 1–25.

Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," NBER Working Paper 9873.

Bewley, Truman F. 2002. "Fairness, Reciprocity, and Wage Rigidity," Cowles Foundation Working Paper. http://cowles.econ.yale.edu/P/cd/d13b/d1383.pdf.

Bhatt, Meghana and Colin F. Camerer. In Press. "Self-Referential Strategic Thinking and Equilibrium as States of Mind in Games: Evidence from fMRI." *Games and Economic Behavior.*

Blair, R. James and Lisa Cipolotti. 2000. "Impaired Social Response Reversal: A Case of 'Acquired Sociopathy.'" *Brain*, 123(6): 1122–41.

Bodner, Ronit and Drazen Prelec. 2003. "The Diagnostic Value of Actions in a Self-Signaling Model," in *The Psychology of Economic Decisions Vol. 1*. Isabelle Brocas and Juan D. Carillo, eds. Oxford: Oxford University Press.

Breiter, Hans C., I. Aharon, Daniel Kahneman, A. Dale, and Peter Shizgal. 2001. "Functional Imaging of Neural Responses to Expectancy and Experience of Monetary Gains and Losses." *Neuron*, 30(2): 619–39.

Buck, Ross. 1999. "The Biological Affects: A Typology." *Psychological Review,* 106(2): 301–36.

Cabanac, Michael. 1979. "Sensory Pleasure." *Quarterly Review of Biology*, 54(1): 1–29.

Caggiula, Anthony R. and Bartley G. Hoebel. 1966. "'Copulation–Reward Site' in the Posterior Hypothalamus." *Science*, 153(3741): 1284–85.

Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Roundtable Series in Behavioral Economics. Princeton: Princeton University Press; New York: Russell Sage Foundation.

Camerer, Colin F. and George Loewenstein. 2004. "Behavioral Economics: Past, Present, and Future," in *Advances in Behavioral Economics*. Colin F. Camerer, George Loewenstein and Matthew Rabin, eds. Princeton: Princeton University Press.

Camerer, Colin F. and Teck Hua Ho. 1999.

"Experience-Weighted Attraction Learning in Normal Form Games." *Econometrica*, 67(4): 827–74.

Camerer, Colin F., Teck Hua Ho, and Kuan Chong. 2004. "A Cognitive Hierarchy Model of One-Shot Games." *Quarterly Journal of Economics*, 119(3): 861–98.

Camerer, Colin F., George Loewenstein, and Martin Weber. 1989. "The Curse of Knowledge in Economic Settings: An Experimental Analysis." *Journal of Political Economy*, 97(5): 1232–54.

Camerer, Colin F. and Dan Lovallo. 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review*, 89(1): 306–18.

Campfield, L. Arthur and Francoise J. Smith. 1990. "Systemic Factors in the Control of Food Intake," in *Handbook of Behavioural Neurobiology, Vol. 10: Neurobiology of Food and Fluid Intake.* E. M. Stricker, ed. New York: Plenum, 183–206.

Caplin, Andrew and John Leahy. 2001. "Psychological Expected Utility Theory and Anticipatory Feelings." *The Quarterly Journal of Economics*, 116(1): 55–79.

Carter, Rita. 1999. *Mapping the Mind*. Berkeley: University of California Press.

Chorvat, Terrence, Kevin McCabe and Vernon L. Smith. 2004. "Law and Neuroeconomics." George Mason Working Paper. http://papers.ssrn.com/sol13/papers.cfm?abstract_id=501063.

Cleckley, Hervey. 1941. *The Mask of Sanity.* St. Louis: C. V. Mosby.

Clore, Gerald L. 1992. "Cognitive Phenomenology: Feelings and the Construction of Judgment," in *The Construction of Social Judgments*. Leonard L. Martin and Abraham Tesser, eds. Hillsdale: Erlbaum, 133–63.

Comings, David E. 1998. "The Molecular Genetics of Pathological Gambling." *CNS Spectrums*, 3(6): 20–37.

Cosmides, Leda. 1989. "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task." *Cognition*, 31(3): 187–276.

Cosmides, Leda and John Tooby. 2004. "Evolutionary Psychology and the Emotions," in *Handbook of Emotions*. Michael Lewis and Jeannette M. Haviland-Jones, eds. New York: Guilford.

Costa-Gomes, Miguel, Vincent P. Crawford, and Bruno Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica*, 69(5): 1193–235.

Costa, Dora L. and Matthew E. Kahn. "Shame and Ostracism: Union Army Deserters Leave Home." MIT Working Paper. http://web.mit.edu/costa/www/shame11b.html.

Cottle, Thomas. 1974. *The Present of Things Future: Explorations of Time in Human Experience*. New York: Free Press.

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

Dana, Jason and George Lowenstein. 2003. "A Psychological Perspective on the Influence of Gifts to Physicians from Industrry." *Journal of the American Medical Association*, 290(2): 252–55.

de Becker, Gavin. 1997. *The Gift of Fear: Survival Signals That Protect Us from Violence*. Boston: Little, Brown and Co.

Deaner, Robert O., Amit V. Khera, and Michael Platt. 2005. "Monkeys Pay Per View: Adaptive Valuation of Social Images by Rhesus Macaques," *Current Biology.* http://download.current-biology.com/pdfs/0960-9822/PIIS960982205001041.pdf

de Gelder, Beatrice, Jean Vroomen, Gilles Pourtois, and Lawrence Weiskrantz. 1999. "Non-Conscious Recognition of Affect in the Absence of Striate Cortex." *Neuroreport*, 10(18): 3759–63.

De Houwer, Jan, Dirk Hermans, and Paul Eelen. 1998. "Affective and Identity Priming with Episodically Associated Stimuli." *Cognition and Emotion,* 12(2): 145–69.

Delgado, Mauricio R., Leigh E. Nystrom, C. Fissell, D. C. Noll, and Julie A. Fiez. 2000. "Tracking the Hemodynamic Responses to Reward and Punishment in the Striatum." *Journal of Neurophysiology*, 84(6): 3072–77.

DeLillo, Don. *White Noise.* East Rutherford, NJ: Penguin Books.

della Vigna, Stefano and Ulrike Malmendier. 2003. "Overestimating Self-Control: Evidence from the Health Club Industry," Berkeley Working Paper. http://emlab.berkeley.edu/users/sdellavi/wp/self_control_nov03.pdf.

Denes-Raj, Veronika, Seymour Epstein, and Jonathan Cole. 1995. "The Generality of the Ratio–Bias Phenomenon." *Personality and Social Psychology Bulletin*, 21(10): 1083–92.

Dubois, Samuel, Bruno Rossion, Christine Schiltz, Jean-Michel Bodart, Christian Michel, Raymond Bruyer, and Marc Crommelinck. 1999. "Effect of Familiarity on the Processing of Human Faces." *Neuroimage*, 9: 278–89.

Eisenberger, Naomi I., Matthew D. Lieberman, and Kipling D. Williams. 2003. "Does Rejection Hurt? An fMRI Study of Social Exclusion." *Science*, 302(5643): 290–92.

Elbert, Thomas, Christo Pantev, Christian Wienbruch, Brigitte Rockstroh, and Edward Taub. 1995. "Increased Cortical Representation of the Fingers of the Left Hand in String Players." *Science*, 270(5234): 305–07.

Elster, Jon. 1977. *Ulysses and the Sirens*. Cambridge: Cambridge University Press.

Erk, Suzanne, Manfred Spitzer, Arthur P. Wunderlich, Lars Galley, and Henrik Walter. 2002. "Cultural Objects Modulate Reward Circuitry." *Neuroreport*, 13(18): 2499–503.

Farrer, Chlöé and Chris D. Frith. 2001. "Experiencing Oneself versus Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency." *Neuroimage*, 15(3): 596–603.

Fazio, Russell, David M. Sanbonmatsu, Martha C. Powell, and Frank Kardes. 1986. "On the Automatic Activation of Attitudes." *Journal of Personality and Social Psychology*, 50(2): 229–38.

Fehr, Ernst and Simon Gachter. 2000. "Cooperation and Punishment in Public Goods Experiments."

American Economic Review, 90(4): 980–94.

Fletcher, Paul C., Francesca Happe, Uta Frith, S. C. Baker, Ray J. Dolan, Richard S. Frackowiak, and Chris D. Frith. 1995. "Other Minds in the Brain: A Functional Imaging Study of "Theory of Mind" in Story Comprehension." *Cognition*, 57(2): 109–28.

Frederick, Shane and George Loewenstein. 2000. "The Psychology of Sequence Preferences." CMU Working Paper.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, 40(2): 351–401.

Freeman, Walter J. and James W. Watts. 1942. *Psychosurgery in the Treatment of Mental Disorders and Intractable Pain.* Springfield: Thomas.

Fried, Itzhak. 1998. "Technical Comment: The Hippocampus and Human Navigation." *Science*, 282(5397): 2151.

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

Frith, Uta. 2001a. "Mind Blindness and the Brain in Autism." *Neuron*, 32(6): 969–79.

Frith, Uta. 2001b. "What Framework Should We Use for Understanding Developmental Disorders?" *Developmental Neuropsychology*, 20(2): 555–63.

Fudenberg, Drew and David K. Levine. 1998. *The Theory of Learning in Games*. Cambridge and London: MIT Press.

Gazzaniga, Michael S. and Joseph E. LeDoux. 1978. *The Integrated Mind*. New York: Plenum.

Gibbs, James, S. P. Madison, and Edmund T. Rolls. 1981. "Satiety Role of the Small Intestine Examined in Sham-Feeding Rhesus Monkeys." *Journal of Comparative Physiology and Psychology*, 95(6): 1003–15.

Gilbert, Daniel T. 2002. "Inferential Correction," in *Heuristics and Biases: The Psychology of Intuitive Judgment*. Thomas Gilovich, D. Griffin and Daniel Kahneman, eds. Cambridge: Cambridge University Press, 167–84.

Gilbert, Daniel T. and Michael J. Gill. 2000. "The Momentary Realist." *Psychological Science*, 11(5): 394–98.

Giner-Sorolla, Roger. 2001. "Guilty Pleasures and Grim Necessities: Affective Attitudes in Dilemmas of Self-Control." *Journal of Personality and Social Psychology*, 80(2): 206–21.

Giordano, Louis A., Warren K. Bickel, George Loewenstein, Eric A. Jacobs, Lisa Marsch, and Gary J. Badger. 2002. "Opioid Deprivation Affects How Opioid-Dependent Outpatients Discount the Value of Delayed Heroin and Money." *Psychopharmacology*, 163(2): 174–82.

Glimcher, Paul W. 2003. *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. Cambridge and London: MIT Press.

Glimcher, Paul W., Michael Dorris, and Hannah Bayer. In Press. "Physiological Utility Theory and the Neuroeconomics of Choice." *Games and Economic Behavior*.

Gneezy, Uri and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *The Quarterly Journal of Economics*, 112(2): 631–45.

Gneezy, Uri and Aldo Rustichini. 2004. "Incentives, Punishment and Behavior," in *Advances in Behavioral Economics*. Colin F. Camerer, George F. Loewenstein and Matthew Rabin, eds. New York: Russell Sage Foundation, 572–89.

Gobet, Ferhand and Herbert A. Simon. 1996. "Recall of Random and Distorted Chess Positions: Implications for the Theory of Expertise." *Memory and Cognition*, 24(4): 493–503.

Goldman, Alvin. 2003. "Emotion, Mindreading, Simulation, and Modularity." *Other Minds: An Interdisciplinary Conference*. Institute of Cognitive and Decision Sciences: University of Oregon, September 27–28, 2003.

Gollwitzer, Peter M., Kentaro Fujita, and Gabriele Oettingen. 2004. "Planning and Implementation of Goals," in *Handbook of Self-Regulation: Research, Theory, and Applications*. Roy F. Baumeister and Kathleen D. Vohs, eds. New York: Guilford Press, 211–28.

Gonzalez, Roxanna and George Loewenstein. 2004. "Effects of Circadian Rhythm on Cooperation in an Experimental Game." Working Paper. http://ssrn.com/abstract=485442.

Green, Leonard J. and Howard Rachlin. 1991. "Economic Substitutability of Electrical Brain Stimulation, Food, and Water." *Journal of the Experimental Analysis of Behavior*, 55(2): 133–43.

Greenberg, D., G. P. Smith, and J. Gibbs. 1990. "Intraduodenal Infusions of Fats Elicit Satiety in Sham-Feeding Rats." *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, 259(1 Pt. 2): R110–18.

Greenwald, Anthony G. 1992. "New Look 3: Unconscious Cognition Reclaimed." *American Psychologist*, 47(6): 766–79.

Greenwald, Anthony G., M. R. Klinger, and T. J. Liu. 1989. "Unconscious Processing of Dichoptically Masked Words." *Memory and Cognition*, 17(1): 35–47.

Grether, David, Charles Plott, Daniel Rowe, Martin Sereno, and John Allman. 2004. "Mental Processes and Strategic Equilibration: An fMRI Study of Selling Strategy in Second Price Auctions." Caltech Working Paper No. 1189. http://www.hss.caltech.edu/SSpapers/wp1189.pdf.

Hadjikhani, Nouchine and Beatrice de Gelder. 2003. "Seeing Fearful Body Expressions Activates the Fusiform Cortex and Amygdala." *Current Biology*, 13(24): 2201–05.

Haggard, Patrick, Sam Clark, and Jeri Kalogeras. 2002. "Voluntary Action and Conscious Awareness." *Nature Neuroscience*, 5(4): 382–85.

Hahnloser, Richard H., Rahul Sarpeshkar, Misha A. Mahowald, Rodney J. Douglas, and H. Sebastian Seung. 2000. "Digital Selection and Analogue Amplification Coexist in a Cortex-Inspired Silicon Circuit." *Nature*, 405(6789): 947–51.

Haier, Richard J., Benjamin V. Siegel, Jr., Andrew MacLachlan, Eric Soderling, Stephen Lottenberg, and Monte S. Buchsbaum. 1992. "Regional Glucose Metabolic Changes After Learning a Complex

Visuospatial/Motor Task: a PET Study." *Brain Research*, 570(1–2): 134–43.

Happe, Francesca, Stefan Ehlers, Paul Fletcher, Uta Frith, Maria Johansson, Christopher Gillberg, Ray Dolan, Richard Frackowiak, and Chris Frith. 1996. "'Theory of Mind' in the Brain: Evidence from a PET Scan Study of Asperger Syndrome." *Neuroreport*, 8(1): 197–201.

Hare, Robert D. 1965. "Psychopathy, Fear Arousal And Anticipated Pain." *Psychological Reports*, 16: 499–502.

Hare, Robert D. 1966. "Temporal Gradient of Fear Arousal in Psychopaths." *Journal of Abnormal Psychology*, 70(6): 442–45.

Hart, Allen J., Paul Whalen, Lisa Shin, Sean McInerney, Hakan Fischer, and Scott Rauch. 2000. "Differential Response in the Human Amygdala to Racial Outgroup versus Ingroup Face Stimuli." *Neuroreport*, 11(11): 2351–55.

Hastie, Reid. 1984. "Causes and Effects of Causal Attribution." *Journal of Personality and Social Psychology*, 46(1): 44–56.

Healy, Paul J. 2004. "Group Reputations and Stereotypes as a Contract Enforcement Device." Caltech Working Paper. http://kakutani.caltech.edu/~pj/papers/Healy-StereotypeContracts.pdf.

Hebb, Donald. 1949. *The Organization of Behavior: A Neuropsychological Theory.* New York: Wiley.

Hershey, John C., Howard C. Kunreuther, and Paul J. H. Schoemaker. 1982. "Sources of Bias in Assessment Procedures for Utility Functions." *Management Science*, 28(8): 936–54.

Hesse, Wolfram, Eva Möller, Matthias Arnold, and Bärbel Schack. 2003. "The Use of Time-Variant EEG Granger Causality for Inspecting Directed Interdependencies of Neural Assemblies." *Journal of Neuroscience Methods*, 124(1): 27–44.

Hill, Elisabeth and David Sally. 2003. "Dilemmas and Bargains: Autism, Theory-of-Mind, Cooperation and Fairness." University College London Working Paper. http://ssrn.com/abstract=407040.

Hirst, William. 1993. "On the Nature of Systems," in *Conceptions of the Human Mind: Essays in Honor of George A. Miller*. Gilbert Harman and George A. Miller eds. Hillsdale: Erlbaum, 1–11.

Hoch, Stephen J. and George F. Loewenstein. 1991. "Time-Inconsistent Preferences and Consumer Self-Control." *Journal of Consumer Research*, 17(4): 492–507.

Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon L. Smith. 1994. "Preferences, Property Rights, and Anonymity in Bargaining Games." *Games and Economic Behavior*, 7(3): 346–80.

Houston, David and Russell H. Fazio. 1989. "Biased Processing as a Function of Attitude Accessibility: Making Objective Judgments Subjectively." *Social Cognition,* 7(1): 51-66.

Hsu, Ming, Meghana Bhatt, Ralph Adolphs, Daniel Tranel, and Colin Camerer. 2005. "Ambiguity-Aversion in the Brain: fMRI and Lesion-Patient Evidence." Caltech Working Paper.

Hursh, Steven R. and B. H. Natelson. 1981. "Electrical Brain Stimulation and Food Reinforcement Dissociated by Demand Elasticity." *Physiological Behavior,* 26(3): 509–15.

Jevons, William S. 1871. *The Theory of Political Economy.* London: Macmillan and Co.

Johnson, Eric J., Colin F. Camerer, Sen Sankar, and Talia Tymon. 2002. "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining." *Journal of Economic Theory*, 104(1): 16–47.

Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics." *American Economic Review*, 93(5): 1449–75.

Kahneman, Daniel and Shane Frederick. 2002. "Representativeness Revisited: Attribution Substitution in Intuitive Judgment," in *Heuristics of Intuitive Judgment: Extensions and Applications*. Thomas Gilovich, Dale Griffin and Daniel Kahneman, eds. New York: Cambridge University Press, 49–81.

Kirkpatrick, Lee A. and Seymour Epstein. 1992. "Cognitive–Experiential Self-Theory and Subjective Probability: Further Evidence for Two Conceptual Systems." *Journal of Personality and Social Psychology*, 63(4): 534–44.

Knutson, Brian and Richard Peterson. In Press. "Neurally Reconstructing Expected Utility." *Games and Economic Behavior*.

Körding, Konrad P. and Daniel M. Wolpert. 2004. "Bayesian Integration in Sensorimotor Learning." *Nature*, 427(6971): 244–47.

Kosslyn, Stephen M. 1994. *Image and Brain*. Cambridge: MIT Press.

Kruglanski, Arie W., Erik P. Thompson, E. Tory Higgins, M. Nadir Atash, Antonio Pierro, James Y. Shah, and Scott Spiegel. 2000. "To 'Do the Right Thing' or to 'Just Do It': Locomotion and Assessment as Distinct Self-Regulatory Imperatives." *Journal of Personality and Social Psychology*, 79(5): 793–815.

Laibson, David. 2001. "A Cue-Theory of Consumption." *The Quarterly Journal of Economics*, 116(1): 81–119.

Laibson, David I., Andrea Repetto, and Jeremy Tobacman. 1998. "Self-Control and Saving for Retirement." *Brookings Papers on Economic Activity.* 1: 91–196.

Lambrecht, Anja and Bernd Skiera. 2004. "Paying Too Much and Being Happy About It: Causes and Consequences of Tariff Choice–Bias." Johann Wolfgang Goethe-University Frankfurt am Main Working Paper.

Langer, Thomas and Craig R. Fox. "Partition Dependence and Naïve Diversification in Investment Decision Making." University of Mannheim Working Paper.

Le Bihan, Denis, Jean-François Mangin, Cyril Poupon, Chris A. Clark, Sabina Pappata, Nicolas Molko, and Hughes Chabriat. 2001. "Diffusion Tensor Imaging: Concepts and Applications." *Journal of Magnetic Resonance Imaging*, 13(4): 534–46.

Leboeuf, Robyn Aimee. 2002. "Alternating Selves and Conflicting Choices: Identity Salience and Preference Inconsistency." *Dissertation Abstracts International*, 63(2–B): 1088.

LeDoux, Joseph E. 1996. *The Emotional Brain: The*

Mysterious Underpinnings of Emotional Life. New York: Simon and Schuster.

Lee, Daeyeol, Michelle L. Conroy, Benjamin P. McGreevy and Dominic J. Barraclough. 2004. "Reinforcement Learning and Decision Making in Monkeys During a Competitive Game." *Cognitive Brain Research,* 22(1): 45–58.

Lerner, Jennifer S. and Dacher Keltner. 2001. "Fear, Anger, and Risk." *Journal of Personality and Social Psychology*, 81(1): 146–59.

Libet, Benjamin. 1985. "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action." *Behavior and Brain Sciences*, 8(4): 529–66.

Lieberman, Matthew D. In Press. "Reflective and Reflexive Judgment Processes: A Social Cognitive Neuroscience Approach," in *Responding to the Social World: Implicit and Explicit Processes in Social Judgments and Decisions*. Joseph P. Forgas, Kipling D. Williams and William von Hippel, eds. Philadelphia: Psychology Press.

Lieberman, Matthew D., Ruth Gaunt, Daniel T. Gilbert, and Yaacov Trope. 2002. "Reflection and Reflexion: A Social Cognitive Neuroscience Approach to Attributional Interference," in *Advances in Experimental Social Psychology*. Mark P. Zanna, ed. New York: Academic Press, 199–249.

Lo, Andrew W. and Dmitry V. Repin. 2002. "The Psychophysiology of Real-Time Financial Risk Processing." *Journal of Cognitive Neuroscience*, 14(3): 323–39.

Loewenstein, George F. 1992. "The Fall and Rise of Psychological Explanations in the Economics of Intertemporal Choice," in *Choice Over Time*. George Loewenstein and Jon Elster, eds. New York: Russell Sage Foundation, 3–34.

Loewenstein, George F. 1994. "The Psychology of Curiosity: A Review and Reinterpretation." *Psychological Bulletin*, 116(1): 75–98.

Loewenstein, George F. 1996. "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, 65(3): 272–92.

Loewenstein, George F. 1999. "Because It Is There: The Challenge of Mountaineering…for Utility Theory." *Kyklos*, 52(3): 315–43.

Loewenstein, George F. and Jennifer S. Lerner. 2003. "The Role of Affect in Decision Making," in *Handbook of Affective Sciences*. Richard J. Davidson, Klaus R. Scherer and H. Hill Goldsmith, eds. Oxford: Oxford University Press, 619–42.

Loewenstein, George F. and Ted O'Donoghue. 2004. "Animal Spirits: Affective and Deliberative Influences on Economic Behavior," CMU Working Paper.

Loewenstein, George F. and Drazen Prelec. 1993. "Preferences for Sequences of Outcomes." *Psychological Review*, 100(1): 91–108.

Loewenstein, George F., Leigh Thompson, and Max H. Bazerman. 1989. "Social Utility and Decision Making in Interpersonal Contexts." *Journal of Personality and Social Psychology*, 57(3): 426–41.

Loewenstein, George F., Elke U. Weber, Christopher K. Hsee, and Ned Welch. 2001. "Risk as Feelings." *Psychological Bulletin*, 127(2): 267–86.

Loewenstein, George F., Robert Weber, Janine Flory,

Stephen Manuck and Matthew Muldoon. 2001. "Dimensions of Time Discounting." *Conference on Survey Research on Household Expectations and Preferences*. Ann Arbor, MI: November 2–3, 2001.

Logothetis, Nikos K., Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. 2001. "Neurophysiological Investigation of the Basis of the fMRI Signal." *Nature*, 412(6843): 150–57.

Lucas, Robert E., Jr. 1978. "Asset Prices in an Exchange Economy." *Econometrica*, 46(6): 1429–45.

Lykken, David T. 1957. "A Study of Anxiety in the Sociopathic Personality." *Journal of Abnormal Psychology*, 55(1): 6–10.

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

Maguire, Eleanor A., David G. Gadian, Ingrid S. Johnsrude, Catriona D. Good, John Ashburner, Richard S. Frackowiak, and Christopher D. Frith. 2000. "Navigation-Related Structural Change in the Hippocampi of Taxi Drivers." *Proceedings of the National Academy of Sciences of the United States of America*, 97(8): 4398–403.

Mailath, George. 1998. "Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory." *Journal of Economic Literature,* 36(3): 1347–74.

Mailath, George J., Michihiro Kandori, and Rafael Rob. 1993. "Learning, Mutation, and Long Run Equilibria in Games." *Econometrica*, 61(1): 29–56.

Mandler, George. 1982. "The Structure of Value: Accounting for Taste," in *Affect and Cognition: The Seventeenth Annual Carnegie Symposium on Cognition*. M. S. Clark and S. T. Fiske, eds. Hillsdale: Erlbaum, 3–36.

Manuck, Stephen B., Janine Flory, Matthew Muldoon, and Robert E. Ferrell. 2003. "A Neurobiology of Intertemporal Choice," in *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*. George F. Loewenstein, Daniel Read and Roy F. Baumeister, eds. New York: Russell Sage Foundation, 139–72.

Marshall, J. F., J. S. Richardson, and P. Teitelbaum. 1974. "Nigrostriatal Bundle Damage and the Lateral Hypothalamic Syndrome." *Journal of Comparative and Physiological Psychology*, 87(5): 808–30.

McCabe, Kevin, Daniel Houser, Lee Ryan, Vernon L. Smith, and Theodore Trouard. 2001. "A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange." *Proceedings of the National Academy of Sciences of the United States of America*, 98(20): 11832–35.

McClure, Samuel M., David I. Laibson, George F. Loewenstein and Jonathan D. Cohen. 2004. "Separate Neural Systems Value Immediate and Delayed Monetary Rewards." *Science*, 306(5695): 503–07.

McConnell, Allen and Jill Leibold. 2001. "Relations Among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes." *Journal of Experimental Social Psychology*, 37(5): 435–42.

McElroy, Susan L., Andrew Satlin, Harrison Pope, Paul E. Keck, and James I. Hudson. 1991.

"Treatment of Compulsive Shopping with Antidepressants: A Report of Three Cases." *Annals of Clinical Psychiatry*, 3(3): 199–204.

McEvily, Bill, Roberto Weber, Cristina Bicchieri and Violet Ho. 2003. "Can Groups Be Trusted? An Experimental Study of Collective Trust," CMU Working Paper. http://www.andrew.cmu.user/rwe-ber/CollectiveTrust10.pdf.

McNeil, Barbara J., Stephen G. Pauker, Harold C. Sox, Jr., and Amos Tversky. 1982. "On the Elicitation of Preferences for Alternative Therapies." *New England Journal of Medicine*, 306(21): 1259–62.

Medin, Douglas L. and Max H. Bazerman. 1999. "Broadening Behavioral Decision Research: Multiple Levels of Cognitive Processing." *Psychonomic Bulletin and Review*, 6(4): 533–47.

Mendelson, Joseph. 1967. "Lateral Hypothalamic Stimulation in Satiated Rats: The Rewarding Effects of Self-Induced Drinking." *Science*, 157(3792): 1077–79.

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

Miller, George A., Eugene Galanter, and Karl H. Pribram. 1960. *Plans and the Structure of Behavior*. New York: Holt.

Mischel, Walter, Ebbe B. Ebbesen, and Antonette R. Zeiss. 1972. "Cognitive and Attentional Mechanisms in Delay of Gratification." *Journal of Personality and Social Psychology*, 21(2): 204–18.

Mischel, Walter and Robert Metzner. 1962. "Preference for Delayed Reward as a Function of Age, Intelligence, and Length of Delay Interval." *Journal of Abnormal Social Psychology*, 64(6): 425–31.

Mobbs, Dean, Michael D. Greicius, Eiman Abdel-Azim, Vinod Menon, and Allan L. Reiss. 2003. "Humor Modulates the Mesolimbic Reward Centers." *Neuron*, 40(5): 1041–48.

Montague, P. Read and Gregory S. Berns. 2002. "Neural Economics and the Biological Substrates of Valuation." *Neuron*, 36(2): 265–84.

Montague, P. Read, Gregory S. Berns, Jonathan D. Cohen, Samuel M. McClure, Giuseppe Pagnoni, Mukesh Dhamala, Michael C. Wiest, Igor Karpov, Richard D. King, Nathan Apple, and Ronald E. Fisher. 2002. "Hyperscanning: Simultaneous fMRI During Linked Social Interactions." *Neuroimage*, 16(4): 1159–64.

Moreyra, Paula, Angela A. Ibáñez, Jerónimo Saiz-Ruiz, Kore Nissenson, and Carlos Blanco. 2000. "Review of the Phenomenology, Etiology and Treatment of Pathological Gambling." *German Journal of Psychiatry*, 3(2): 37–52.

Morris, John S., C. Buchel, and Raymond J. Dolan. 2001. "Parallel Neural Responses in Amygdala Subregions and Sensory Cortex During Implicit Fear Conditioning." *Neuroimage*, 13(6 Pt. 1): 1044–52.

Mullainathan, Sendhil. 2002. "Thinking Through Categories," MIT Working Paper.

National Academy of Sciences. 1999. *Pathological Gambling*. Washington, D.C.: National Academy Press.

Nichols, M. James and William T. Newsome. 2002. "Middle Temporal Visual Area Microstimulation

Influences Veridical Judgments of Motion Direction." *Journal of Neuroscience*, 22(21): 9530–40.

Nisbett, Richard E. and Timothy D. Wilson. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Bulletin*, 84(3): 231–59.

O'Donoghue, Ted and Matthew Rabin. 1997. "Addiction and Self-Control," in *Addiction: Entries and Exits*. Jon Elster, ed. New York: Russell Sage Foundation, 1999, 169–206.

O'Donoghue, Ted and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review*, 89(1): 103–24.

Ochsner, Kevin N., Silvia A. Bunge, James J. Gross, and James D.E. Gabrieli. 2002. "Rethinking Feelings: An fMRI Study of the Cognitive Regulation of Emotion." *Journal of Cognitive Neuroscience*, 14(8): 1215–29.

Ochsner, Kevin N. and Matthew D. Lieberman. 2001. "The Emergence of Social Cognitive Neuroscience." *American Psychologist*, 56(9): 717–34.

Olds, James and Peter Milner. 1954. "Positive Reinforcement Produced by Electrical Stimulation of Septal Area and Other Regions of Rat Brain." *Journal of Comparative and Physiological Psychology*, 47(6): 419–27.

Oswald, Andrew J. 1997. "Happiness and Economic Performance." *Economic Journal*, 107(445): 1815–31.

Panksepp, Jaak. 1998. *Affective Neuroscience*. Oxford: Oxford University Press.

Parsons, Lawrence M. and Daniel Osherson. 2001. "New Evidence for Distinct Right and Left Brain Systems for Deductive versus Probabilistic Reasoning." *Cerebral Cortex*, 11(10): 954–65.

Persico, Nicola, Andrew Postlewaite, and Dan Silverman. "The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height," University of Pennsylvania Working Paper.

Persinger, Michael A. and Faye Healey. 2002. "Experimental Facilitation of the Sensed Presence: Possible Intercalation Between the Hemispheres Induced by Complex Magnetic Fields." *Journal of Nervous and Mental Disease*, 190(8): 533–41.

Peters, Ellen and Paul Slovic. 2000. "The Springs of Action: Affective and Analytical Information Processing in Choice." *Personality and Social Psychology Bulletin*, 26(12): 1465–75.

Phelps, Elizabeth, Kevin J. O'Connor, William A. Cunningham, E. Sumie Funayama, J. Christopher Gatenby, John C. Gore, and Mahzarin R. Banaji. 2000. "Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activity." *Journal of Cognitive Neuroscience*, 12(5): 729–38.

Pillutla, Madan M. and Xiao-Ping Chen. 1999. "Social Norms and Cooperation in Social Dilemmas: The Effects of Context and Feedback." *Organizational Behavior and Human Decision Processes*, 78(2): 81–103.

Platt, Michael L. and Paul W. Glimcher. 1999. "Neural Correlates of Decision Variables in Parietal Cortex." *Nature*, 400(6741): 233–38.

Prelec, Drazen and George F. Loewenstein. 1998. "The Red and the Black: Mental Accounting of Savings and Debt." *Marketing Science*, 17(1): 4–28.

Prelec, Drazen and Duncan Simester. 2001. "Always Leave Home Without It." *Marketing Letters*, 12(1): 5–12.

Quattrone, George A. and Amos Tversky. 1984. "Causal Versus Diagnostic Contingencies: On Self-Deception and on the Voter's Illusion." *Journal of Personality and Social Psychology*, 46(2): 237–48.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review*, 83(5): 1281–302.

Read, Daniel and George Loewenstein. 1995. "The Diversification Bias: Explaining the Difference Between Prospective and Real-Time Taste for Variety." *Journal of Experimental Psychology: Applied*, 1(1): 34–49.

Read, Daniel, George Loewenstein, and Shobana Kalyanaraman. 1999. "Mixing Virtue and Vice: The Combined Effects of Hyperbolic Discounting and Diversification." *Journal of Behavioral Decision Making*, 12(4): 257–73.

Rilling, James K., D. A. Gutman, T. R. Zeh, Giuseppe Pagnoni, Greg S. Berns, and Clint D. Kilts. 2002. "A Neural Basis for Social Cooperation." *Neuron*, 35(2): 395–405.

Rolls, Edmund T. 1999. *The Brain and Emotion*. New York: Oxford University Press.

Romer, Paul M. 2000. "Thinking and Feeling." *American Economic Review*, 90(2): 439–43.

Ross, Lee, Mark R. Lepper, and Michael Hubbard. 1975. "Perseverance in Self-Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm." *Journal of Personality and Social Psychology*, 32(5): 880–92.

Rowe, Andrea D., Peter R. Bullock, Charles E. Polkey, and Robin G. Morris. 2001. "'Theory of Mind' Impairments and their Relationship to Executive Functioning Following Frontal Lobe Excisions." *Brain*, 124(3): 600–16.

Rumelhart, David E. and James L. McClelland. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. Cambridge: MIT Press.

Rustichini, Aldo, John Dickhaut, Paolo Ghirardato, Kip Smith and Jose V. Pardo. 2002. "A Brain Imaging Study of Procedural Choice," University of Minnesota Working Paper.

Samuelson, Paul. 1937. "A Note on Measurement of Utility." *Review of Economic Studies*, 4: 155–61.

Sanfey, Alan G., James K. Rilling, Jessica A. Aaronson, Leigh E. Nystrom, and Jonathan D. Cohen. 2003. "The Neural Basis of Economic Decision-Making in the Ultimatum Game." *Science*, 300(5626): 1755–58.

Saxe, Rebecca and Nancy Kanwisher. 2003. "People Thinking about Thinking People: The Role of the Temporo–Parietal Junction in 'Theory of Mind.'" *Neuroimage*, 19(4): 1835–42.

Schelling, Thomas C. 1978. "Egonomics, or the Art of Self-Management." *American Economic Review*, 68(2): 290–94.

Schelling, Thomas C. 1984. "The Intimate Contest for Self-Command," in *Choice and Consequence*. Cambridge: Harvard University Press, 57–81.

Schneider, Walter and Richard M. Shiffrin. 1977.

"Controlled and Automatic Human Information Processing: I. Detection, Search and Attention." *Psychological Review*, 84(1): 1–66.

Schultz, Wolfram. 2002. "Getting Formal with Dopamine and Reward." *Neuron*, 36(2): 241–63.

Schultz, Wolfram and Anthony Dickinson. 2000. "Neuronal Coding of Prediction Errors." *Annual Review of Neuroscience*, 23: 473–500.

Schwarz, Norbert. 1990. "Feelings as Information: Informational and Motivational Functions of Affective States," in *Handbook of Motivation and Cognition: Foundations of Social Behavior*. E. T. Higgins and R. M. Sorrentino, eds. New York: Guilford Press, 527–61.

Schwarz, Norbert and Gerald L. Clore. 1983. "Mood, Misattribution and Judgments of Well-Being: Informative and Directive Functions of Affective States." *Journal of Personality and Social Psychology*, 45(3): 513–23.

Shallice, Timothy and Paul W. Burgess. 1996. "The Domain of Supervisory Processes and Temporal Organization of Behaviour." *Philosophical Transactions of the Royal Society B: Biological Sciences*, 351(1346): 1405–12.

Shapiro, David. 1965. *Neurotic Styles*. New York: Basic Books.

Shefrin, Hersh and Richard Thaler. 1988. "The Behavioral Life Cycle Hypothesis." *Economic Inquiry,* 26(4): 609–43.

Sherry, David F. and Daniel L. Schacter. 1987. "The Evolution of Multiple Memory Systems." *Psychological Review*, 84(2): 127–90.

Shiffrin, Richard M. and Walter Schneider. 1977. "Controlled and Automatic Human Information Processing: II. Perceptual Learning, Automatic Attending and a General Theory." *Psychological Review*, 84(2): 127–90.

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(3): 278–92.

Shiv, Baba, George Loewenstein, Antoine Bechara, Antonio Damasio and Hanna Damasio. 2005. "Investment Behavior and the Dark Side of Emotion." *Psychological Science*. Forthcoming.

Shizgal, Peter. 1999. "On the Neural Computation of Utility: Implications from Studies of Brain Reward," in *Foundations of Hedonic Psychology: Scientific Perspectives on Enjoyment and Suffering*. Daniel Kahneman, Edward Diener and Norbert Schwarz eds. New York: Russell Sage Foundation, 502–26.

Simon, Herbert A. 1967. "Motivational and Emotional Controls of Cognition." *Psychological Review*, 74(1): 29–39.

Simonson, Itamar. 1990. "The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior." *Journal of Marketing Research*, 27(2): 150–62.

Singer, Tania, Stefan J. Kiebel, Joel S. Winston, Raymond J. Dolan, and Chris D. Frith. 2004. "Brain Responses to the Acquired Moral Status of Faces." *Neuron*, 41(4): 653–62.

Sloman, Steven A. 1996. "The Empirical Case for Two

Systems of Reasoning." *Psychological Bulletin*, 119(1): 3–22.

Smith, Trenton. "The McDonald's Equilibrium: Advertising, Empty Calories, and the Endogenous Determination of Dietary Preferences," University of Bonn Working Paper. http://www.international.ucla.edu/cms/files/smithnov13.pdf.

Strack, Fritz, Leonard L. Martin, and Sabine Stepper. 1988. "Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis." *Journal of Personality and Social Psychology*, 54(5): 768–77.

Titmuss, Richard M. 1970. *The Gift Relationship*. London: Allen and Unwin.

Train, Kenneth E. 1991. *Optimal Regulation: The Economic Theory of Natural Monopoly*. Cambridge: MIT Press.

Train, Kenneth E., Daniel L. McFadden, and Moshe Ben-Akiva. 1987. "The Demand for Local Telephone Service: A Fully Discrete Model of Residential Calling Patterns and Service Choices." *Rand Journal of Economics*, 18(1): 109–23.

Tversky, Amos and Daniel Kahneman. 1983. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review*, 90(4): 293–315.

Van Boven, Leaf, George Loewenstein, and David Dunning. 2003. "Mispredicting the Endowment Effect: Underestimation of Owners' Selling Prices by Buyer's Agents." *Journal of Economic Behavior and Organization*, 51(3): 351–65.

Viner, Jacob. 1925. "The Utility Concept in Value Theory and its Critics." *Journal of Political Economy*, 33(4): 369–87.

von Melchner, Laurie, Sarah L. Pallas, and Mriganka Sur. 2000. "Visual Behaviour Mediated by Retinal Projections Directed to the Auditory Pathway." *Nature*, 404(6780): 871–76.

Vuilleumier, Patrik, Jorge L. Armony, Jon Driver, and Ray J. Dolan. 2001. "Effects of Attention and Emotion on Face Processing in the Human Brain: An Event-Related fMRI Study." *Neuron*, 30(3): 829–41.

Wegner, Daniel M. and Thalia Wheatley. 1999. "Apparent Mental Causation. Sources of the Experience of Will." *American Psychologist*, 54(7): 480–92.

Wells, Gary L. and Richard E. Petty. 1980. "The Effects of Overt Head Movements on Persuasion: Compatibility and Incompatibility of Resources." *Basic and Applied Social Psychology*. 1(3): 219–30.

Whalen, Paul J., S. L. Rauch, Nancy Etcoff, S. C. McInerney, M. Lee, and M. A. Jenike. 1998. "Masked Presentations of Emotional Facial Expressions Modulate Amygdala Activity without Explicit Knowledge." *Journal of Neuroscience*, 18(1): 411–18.

Whaling, Carol S., M. M. Solis, Allison J. Doupe, Jill A. Soha, and Peter Marler. 1997. "Acoustic and Neural Bases for Innate Recognition of Song." *Proceedings of the National Academy of Sciences of the United States of America*, 94(23): 12694–98.

Wheeler, Mary E. and Susan T. Fiske. 2005. "Controlling Racial Prejudice: Social–Cognitive Goals Affect Amygdala and Stereotype Activation." *Psychological Science*, 16(1): 56–63.

Wilson, Timothy D., Samuel Lindsey, and Tonya Y. Schooler. 2000. "A Model of Dual Attitudes." *Psychological Review*, 107(1): 101–26.

Wilson, Timothy D., Douglas J. Lisle, Jonathan W. Schooler, Sara D. Hodges, Kristen J. Klaaren, and Suzanne J. LaFleur. 1993. "Introspecting about Reasons Can Reduce Post-Choice Satisfaction." *Personality and Social Psychology Bulletin*, 19(3): 331–39.

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(2): 181–92.

Windschitl, Paul D. and Gary L. Wells. 1998. "The Alternative–Outcomes Effect." *Journal of Personality and Social Psychology*, 75(6): 1411–23.

Winkielman, Piotr and Kent C. Berridge. 2004. "Unconscious Emotion." *Current Directions in Psychological Science*, 13(3): 120–23.

Wise, Roy A. 1996. "Addictive Drugs and Brain Stimulation Reward." *Annual Review of Neuroscience*, 19: 319–40.

Wolford, George, Michael B. Miller, and Michael Gazzaniga. 2000. "The Left Hemisphere's Role in Hypothesis Formation." *Journal of Neuroscience*, 20(6): 1–4.

Zajonc, Robert B. 1980. "Feeling and Thinking: Preferences Need No Inferences." *American Psychologist*, 35(2): 151–75.

Zajonc, Robert B. 1984. "On the Primacy of Affect." *American Psychologist*, 39(2): 117–23.

Zajonc, Robert B. 1998. "Emotions," in *Handbook of Social Psychology*. Daniel T. Gilbert, Susan T. Fiske and Gardner Lindzey, eds. New York: Oxford University Press, 591–632.

Zajonc, Robert B. and Daniel N. McIntosh. 1992. "Emotions Research: Some Promising Questions and Some Questionable Promises." *Psychological Science*, 3(1): 70–74.

Zak, Paul. In Press. "Neuroeconomics." *Transactions of the Royal Philosophical Society B*.

Zak, Paul, Robert Kurzban and William Matzner. 2003. "Oxytocin is Associated with Interpersonal Trust in Humans." Under Review.

Zamir, Shmuel. 2000. "Rationality and Emotions in Ultimatum Bargaining," Discussion Paper #222. http://www.ma.huji.ac.il/~zamir/dp222.pdf.

Zink, Caroline F., Giuseppe Pagnoni, M. E. Martin, M. Dhamala, and Greg S. Berns. 2003. "Human Striatal Response to Salient Nonrewarding Stimuli." *Journal of Neuroscience*, 23(22): 8092–97.