——(2003) 'Introspection and Internalism', in Susana Nuccetelli (ed.), *New Essays on Semantic Externalism, and Self-Knowledge* (Boston), 257–76.
——(2004) 'Epistemic Probability', *Philosophical Issues*, 14: 149–64.
Goldman, Alvin (1967) 'A Causal Theory of Knowing', *Journal of Philosophy*, 64: 355–72.
——(1979) 'What is Justified Belief?', in George Pappas (ed.), *Justification and Knowledge* (Dordrecht), 1–23.
——(1986) *Epistemology and Cognition* (Cambridge).
——(1988) 'Strong and Weak Justification', in James Toberlin (ed.), *Philosophical Perspectives 2: Epistemology* (Atascadero, Calif.), 51–69.
——(1999) 'Internalism Exposed', *Journal of Philosophy*, 96: 271–93.
Huemer, Michael (2001) *Skepticism and the Veil of Perception* (Lanham, Md.).
——(2006) 'Phenomenal Conservatism and the Internal of Intuition', *American Philosophical Quarterly*, 43: 147–58.
——(forthcoming) 'Compassionate Phenomenal Conservatism', *Philosophy and Phenomenological Research*.
Hume, David (1888) *A Treatise of Human Nature*, ed. L. A. Selby-Bigge (London).
Ludlow, Peter, and Norah Martin (eds.) (1998) *Externalism and Self-Knowledge* (Stanford, Calif.).
McGrew, Timothy (1999) 'A Defense of Classical Foundationalism', in Louis Pojman (ed.), *The Theory of Knowledge*, 2nd edn. (New York), 224–35.
Markie, Peter (forthcoming) 'Easy Knowledge', *Philosophy and Phenomenological Research*.
Plantinga, Alvin (1993) *Warrant and Proper Function* (Oxford).
——(2000) *Warranted Christian Belief* (New York).
Quine, W. V. O. (1969) 'Epistemology Naturalized', in *Ontological Relativity and Other Essays* (New York).
Sandberg, Thomas (2004) 'Thomas Reid's Providentialist Epistemology', Ph.D. thesis, University of Iowa.
Van Cleve, James (2003) 'Is Knowledge Easy—or Impossible: Externalism as the Only Alternative to Skepticism', in Steven Luper (ed.), *The Skeptics* (Aldershot), 45–60.
Williamson, Timothy (2000) *Knowledge and its Limits* (Oxford).

4. The Evolution of Irrationality: Insights from Non-Human Primates

Laurie R. Santos

Making decisions is like speaking prose—
people do it all the time, knowingly or unknowingly.
(Kahneman and Tversky, 1984)

1. INTRODUCTION

From the moment we wake up in the morning, we are confronted by a staggering array of choices. (Should I hit the shower or the snooze button, the highway or the byway?) Though sometimes our preferences are well-known to us, often we must make decisions with limited information about how different outcomes will affect our overall happiness and utility. (Having never tried Korean *sannakji*[1] for example, I am unsure whether I would find the experience glorious or repulsive.) Moreover, we can rarely be certain of what the outcome of our choice will be. (Generally, taking the highway increases my overall well-being, but if I have an accident, the utility is decidedly otherwise.)

As decision-makers go, however, human adults are fairly lucky: we have large brains, language, psychic advisers, and Blackberries to help us navigate our myriad choices. Non-human animals face a corresponding array of complex options with a rather more limited

The author would like to thank Mark Maxwell and the editors for helpful comments on an earlier version of this paper. Address correspondence to Laurie R. Santos, Yale University, Department of Psychology, Box 208205, New Haven, CT 06510 or via email at laurie.santos@yale.edu.

[1] *Sannakji*, I'm told, is a Korean delicacy. To prepare it, the chef slices the tentacles off of a small live octopus, which arrive at the table still moving around using their suction cups. Though the squirming is considered a highlight of the experience, it does pose a health hazard; every once in a while a diner chokes as the suction cups stick to his mouth and throat.

set of resources. Like us, non-human animals have constrained time and energy, and must decide between different time- and energy-use alternatives. A female capuchin monkey waking up in the canopy, for example, must decide whether to get up or stay asleep, forage in well-trod paths or try new areas, lunch on boring leaves or search out rarer but more delicious insects, and so on. Such daily decisions may have far-reaching consequences, both in terms of immediate individual utility—how full, tired, comfortable, and happy she is that day—and for her survival and reproductive success.

How do humans and other animals actually navigate the decisions that we face each day? In this chapter, I will challenge what has typically been considered the standard descriptive account of the mechanisms underlying human and animal decision-making, the notion of *rational utility-maximization*—the idea that organisms make decisions rationally, choosing alternatives that maximize their expected payoffs. After presenting a brief overview of this standard theory (section 2) I will review the results of classic studies on decision-making in humans which suggest that even experienced decision-makers violate rationality in a number of systematic and important ways (section 3). (Readers familiar with this literature may wish to skim these portions of the paper.) I will then present new evidence from my lab on decision-making in non-human primates indicating that humans are not alone in their irrational decision-making tendencies (section 4). I will then use this evidence to support an alternative claim—that humans and other animals make decisions using evolutionarily shared (possibly innately specified) cognitive shortcuts, ones that do not adhere strictly to the rules of rationality. I will then very briefly discuss some implications of this notion for cognitive evolution generally and for the idea of rationality in humans and animals (section 5).

2. THE CLASSICAL APPROACH TO RATIONAL CHOICE: EXPECTED UTILITY MAXIMIZATION

The classical view of human decision-making—which I'll refer to throughout this article as rational expected-utility-maximization—starts with a simple assumption about decision-making organisms:

they are rational. Rationality in this case means that organisms will behave in ways that they believe will maximize their own utility. For this notion to make sense, we must presuppose a few things about rational agents. First, self-interested rational agents must have more and less preferred consequences—*preferences*—which are *reasonable*, in the sense that they are consistent over time and transitive across different options (if an agent prefers A to B and B to C then he must also prefer A to C). Second, rational agents must be endowed with certain *reasoning capacities*. They must, at least at some level, be able to make connections between the actions they take and the consequences of those actions, and when actions do not consistently lead to the same consequences, rational organisms must employ the basic tenets of probability to determine the likelihood that a particular action will yield a given consequence. In this way, organisms must factor likelihood information into their calculation of which behaviors can be expected to bring them the best returns. Thus rational agents are assumed to compute *expected utilities*—the value of the consequences of each action adjusted by the likelihood that this consequence will actually occur—for each possible action and then choose the action that, on average, leads to the maximum expected utility. Finally, rational organisms must *consistently behave rationally*—that is, they must always act in ways that are consistent with their own self-interest and preferences, and must always choose options that maximize their own average expected utility, no matter what their current wealth level or situation.

Although the basic idea of rational expected-utility-maximization was first described hundreds of years ago, its popularity reached a pinnacle in the mid-twentieth century when two very separate fields, behavioral ecology and economics, attempted to formulate normative models of optimal behavior. Behavioral ecologists in the 1950s were centrally concerned with the adaptive nature of animal behavior, the extent to which an individual animal's behavior was optimized by natural selection for maximizing that individual's survival and reproductive success. With this in mind, behavioral ecologists became interested in the behavioral trade-offs that animals make on a daily basis, particularly within the domain of foraging (for elegant reviews of this literature, see Glimcher, 2003; Krebs and Davies, 1993). This interest led to the development of

optimal foraging theory, a normative model of optimal choice behavior and a set of mathematical predictions governing how the ideal rationally self-interested individual animal should forage given different environmental payoffs. Optimal foraging theory was thus concerned both with determining how organisms maximize expected daily payoffs, like overall daily energy attainment, and with how they maximize the ultimate evolutionary currency, survival, and reproductive success.

At around the same time that behavioral ecologists were formulating normative models of animal decision-making, economists were developing normative models for human decision-making. Conceptualizing economic situations as multi-player games, these models allowed individual decision-making to be described in terms of choices among different "gambles," each with its own associated payoff and probability of occurring. Under this framework, optimal human decision-making could be understood as a process of computing and comparing different average expected payoffs with the goal of maximizing utility. Such models were taken to apply to the behavior both of individuals (persons) and groups (corporations or markets).

Though rational expected-utility-maximization models were originally formulated as *normative* models, both economists and behavioral ecologists have often adopted the same models as a *descriptive* framework. The idea that rational expected-utility-maximization models accurately describe the behavior of individual animals and investors has held intuitive appeal for economists and biologists for a number of reasons. First, these models fit well with the widespread belief that, in general, people and animals behave in ways that cause their desires to be satisfied. Second, rational expected-utility-maximization models are attractive to economists and biologists because they have a formal appeal; such models are easy to quantify mathematically, and therefore allow for the types of predictive modeling to which economists and mathematical biologists are accustomed. Third, rational expected-utility-maximization models provide a natural explanation for the observation that, in general, large groups of decision-makers—markets in the case of economics and species in the case of ecology—do seem relatively optimized to solve particular problems and achieve specific goals. Finally, rational utility-maximization models mesh well with the first-principle

assumptions of each of these two fields. Behavioral ecologists work under the assumption that the behavior of modern organisms has been shaped over time by the process of natural selection: behavioral strategies observed today are the result of generations of competition for scarce resources. Because optimal decision-making behavior should increase an organism's chance of survival in harsh competitive times, optimal behaviors are more likely to persist across generations of evolutionary selection. In the same way, economists assume that market competition should serve as a strong selection force against suboptimal decision-making strategies. In this way, generations of market forces should select for normatively optimal decision-making behavior. As such, observed market strategies should on average yield relatively optimal payoffs, just as observed animal behavior should yield relatively optimal energy returns that can be translated into relatively optimal reproductive fitness returns.

3. THE MODERN SYNTHESIS: CHOICES, VALUES, AND FRAMES

The normative appeal of rational expected-utility-maximization models led many to the view that such models provide adequate descriptive accounts of behavior, both that of individual human investors and other non-human species. But an enormous (and still growing) body of empirical work suggests that human agents diverge from what rational expected-utility-maximization models would predict, both in the laboratory and in the real world.

Rational expected-utility-maximization models were first questioned by the behavioral economists Daniel Kahneman and Amos Tversky. One of Kahneman and Tversky's earliest and most important observations was that human decision-makers seem to violate a primary assumption of rational expected-utility-maximization models—they don't *always* choose the option with the highest expected utility. In addition, human decision-makers (including experienced ones like economists and investors) generally do not describe the outcomes of their choices in terms of overall utility. In ordinary conversation, people tend to refer to the outcome of their choices as a gain or loss relative to some starting point (e.g.

"I lost $20 because of that parking ticket!") rather than in terms of their overall utility or wealth level (e.g. "My entire net worth is now only $227,364 because of that parking ticket!") Kahneman and Tversky wondered if this relativist rather than absolutist perspective actually affected people's choices. Would people behave differently when faced with outcomes that felt like relative gains than they would for ones that felt like relative losses? They presented participants with the one of the two following scenarios (see Kahneman and Tversky, 1979). The actual percentage of participants that chose each scenario is given in brackets after each scenario.

Scenario 1. You have been given $1000. You are now asked to choose between: (A) a 50% chance of receiving another $1000, and 50% chance of receiving nothing [16%], or (B) receiving $500 with certainty [84%].

Scenario 2. You have been given $2000. You are now asked to choose between: (C) a 50% chance of losing $1000, and a 50% chance of losing nothing [69%], or (D) losing $500 with certainty [31%].

From the perspective of overall utility maximization, each scenario has exactly the same two choices: options A and C each give a 50% chance of a final result of $1000 and a 50% chance of a final result of $2000, and options B and D each guarantee $1500. Rational expected-utility-maximization models would thus predict that human subjects should show the same preference in each of the two scenarios. In contrast to this prediction, participants show quite different preferences across the two scenarios. In the first situation, where both options are framed as gains, participants reliably preferred the safe option B over the risky option A; in the second situation, where options are framed in terms of losses, participants reliably preferred the risky option C over the safe option D.

Kahneman and Tversky used evidence from this and numerous similar cases to argue that human decision-makers do not evaluate choices in terms of overall utility, as the classic rational descriptive account predicts. Instead, they seem to consider different options in regards to a particular (usually arbitrary) *reference point* (e.g. one's current position in a particular experimental gamble, etc.). Kahneman and Tversky further observed that subjects seemed to treat changes from a reference point differently depending on whether

those changes were positive (gains) or negative (losses): people tended to be risk averse when dealing with perceived gains—they chose sure, smaller gains over larger, riskier gains—but risk-seeking when dealing with perceived losses—they preferred a risky chance not to have any loss over a sure small loss. This phenomenon of changing risk-preferences—often termed the *reflection effect*—is observed even in decisions that don't involve monetary gains. Consider another problem presented by Tversky and Kahneman (1981), a scenario commonly referred to as the "Asian disease problem":

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved [72%]
If Program B is adopted, there is a 1/3 probability that 600 people will be saved, and 2/3 probability that nobody will be saved [28%]
If Program C is adopted 400 people will die [22%]
If Program D is adopted there is a 1/3 probability that nobody will die, and 2/3 probability that 600 people will die [78%]

As in the previous set of scenarios, programs A and C are equivalent (200 people will live for sure, and 400 will die for sure), and programs B and D are equivalent (there is a 1/3 chance of 600 people will live and zero will die, and a 2/3 chance that no one will live and 600 die). Nevertheless, as in the case described above, participants presented with the first two options preferred the certain gain (A) to the risky gain (B), while those presented with the second two options preferred the risky loss (D) to the certain loss (C). Participants' choices thus seemed to be based solely on how the problem was written or *framed*: when the choice was described in terms of people dying (i.e. lives *lost*) people chose to avoid a sure loss; when the mathematically identical choice was described in terms of survival rates (i.e. lives gained, so to speak), participants switched their preference and sought out safe options. As Kahneman and Tversky (1981) observed in this and other problems (see Kahneman and Tversky, 2000, for a review), the utility that decision-makers feel they lose with losses tends to be greater than the utility they feel they obtain with identically sized gains. This feature leads to *loss aversion*—people tend to avoid losses more than they tend to seek out equally sized gains. (Human loss

aversion can also be observed in the fact that most average-salaried academics would pass up the following gamble: a 50% chance to win $1001 and a 50% chance to lose $1000.)

The phenomenon of reference dependence and loss aversion led Kahneman and Tversky (1979, 1981) to develop a new descriptive account of human decision-making under uncertainty. This new account, *prospect theory*, begins by treating decisions as measuring value (described in relative terms like gains or losses) rather than overall utility. Under prospect theory, values are measured as losses or gains relative to a specified (yet often arbitrary) reference point. Because of loss aversion, there is a kink in the value curve at the reference point, such that a given absolute-sized loss (e.g. a $5 loss) will decrease value more than an identically sized gain (e.g. a $5 gain) will increase value. One of the major implications of prospect theory is that people naturally frame their decisions as gains or losses relative to a particular reference point. This feature of prospect theory leads to odd and often irrational *framing effects*, in which decision-makers' responses may vary with how the choice is presented, worded, or described. Because of loss aversion, framing effects may cause decision-makers to make decisions that lower their overall expected utility.

We can see these effects in a number of real-world economic situations. Take, for example, the tendency for people to overvalue objects that they own relative to objects that they don't yet own, a phenomenon that Thaler (1980) christened "*the endowment effect.*" In one classic study, Kahneman and colleagues (1990) gave one of two identically priced items—either a coffee mug or a box of pens—to each member of a participants group. They then examined subjects' willingness to trade the item they were given for the alternative item. Interestingly, the researchers observed very few examples of trading. Instead, coffee mug owners tended to demand a larger price to sell or trade their mug than non-owners were willing to pay to buy or trade for it. Kahneman and colleagues hypothesized that the act of becoming an owner of an object changes one's view of that object because it changes one's reference point—object owners will experience a loss when trading an object (and thus, value it more strongly) while object buyers will experience the new object as a gain (and thus, value it less strongly than they would if they were experiencing it as a loss). Because of loss aversion, people's willingness-to-sell an item is often less than their willingness-to-buy an identical item.

4. NEW EMPIRICAL WORK ON IRRATIONALITY: FRAMING EFFECTS IN ANIMALS

The ground-breaking work of Kahneman and Tversky led to a new view of human decision-making: we, humans, are not the normatively obedient decision-makers we had once assumed. Instead, humans behave in ways that violate a number of the major tenets of economic rationality. We have preferences that are inconsistent across time and presentation. We evaluate our payoffs in relative rather than absolute terms, and therefore occasionally fail to choose options with the highest absolute payoffs. Finally, and perhaps most problematic, our view of a decision can be changed arbitrarily by framing—simply wording a problem differently or changing how it's presented can affect both how we value our alternative outcomes and how we behave.

The finding that human decision-making violates these normative standards leads to a number of questions concerning where these decision-making biases come from. How is it that adult human decision-makers (like those tested in studies) came to be loss averse and reference dependent? Do these decision-making biases come about because of specific economic experiences or cultural training? Or are these biases instead a more basic feature of the way humans make decisions? Could they even be present as part of our evolutionary heritage?

My colleagues and I have taken a somewhat radical view of the origins of our own species' irrational decision-making biases (see Chen *et al.*, 2006; Santos and Lakshminarayanan, in press): we hypothesize that at least some aspects of our irrational decision-making are innately specified. More specifically, we have argued that phenomena like loss aversion and reference dependence are not solely the result of experience and cultural learning. Instead, we hypothesize that these decision-making biases result from more evolutionarily ancient tendencies, ones that are likely to be present in the decision-making of other non-human species as well.

To test our somewhat radical idea, my colleagues and I have begun to examine whether the irrational biases that govern human choice are demonstrated in the choice behavior of other animals as well (see Marsh and Kacelnik, 2002, for a similar approach). In particular, we examined whether a closely related non-human primate species, the capuchin monkey (*Cebus apella*) exhibits reference dependence and loss aversion when making decisions conceptually similar to those of the Kahneman and Tversky studies. We chose to begin our comparative investigation with a primate comparative model for a number of reasons. First, non-human primate (hereafter, "primate") models provide one of the best windows into the evolutionary origins of human cognition. Much of human cognition was likely to have been shaped for problems that our ancestors faced over their evolutionary history. The ancestors of closely related primates presumably faced selection pressures similar to those that plagued our human ancestors; as such, modern primates can often provide a useful model for determining the strategies that our ancestors may have used to solve ancestral problems. In addition, primate models can be particularly useful tools for examining the role of experience in shaping how a particular cognitive mechanism operates. Adult primates, unlike adult humans, have little experience in real-world economic markets, and thus they provide an excellent test case for what economic decision-making looks like in the absence of market training and experience. Third, because primates do not naturally use money, any decision-making mechanisms we observe in primates could not have evolved for economic exchange *per se*. Instead, if loss aversion and reference dependence were observed in primates, they would have to be mechanisms that evolved not for market use, but instead as general problem-solving and decision-making strategies.

The overarching goal of our work, then, was to examine how capuchin monkeys would perform on the gambles presented to human participants in classic behavioral economic studies (see reviews in Kahneman and Tversky, 2000). The first challenge of this approach was to develop a way to measure and quantify monkey preferences. Unlike human participants, our capuchin subjects were unable to fill out questionnaires, list what they prefer, or verbally report how much they'd be willing to spend to take part in a particular gamble. We therefore needed to develop a method for

testing primate preferences that could approximate the tasks typically used to test human subjects in choice experiments. In addition, we also hoped to find a capuchin choice method that required as little training as possible. Spontaneous studies, which require very little reinforced training, represent a particularly useful way of assessing how an animal would make decisions in its natural environment. We therefore wanted to develop a task that required little training with the specific gambles in which we were interested in testing.

In the end, we decided that if we really wanted to test monkeys' preferences in a way that was directly comparable to that used by economists to test humans, then the easiest course of action was to teach our monkeys how to use a fungible currency like money and let them use this currency to establish preferences as humans did. Although teaching monkeys to use money might seem a daunting task, a number of labs had previously taught primates to trade token-like "money" in exchange for food (see Brosnan and de Waal, 2003, 2004; Liv *et al.*, 1999; Westergaard *et al.*, 1998, 2004). These labs had used token-trading methods as an easy way of getting monkeys to make choices, but no one to date had used monkey tokens to set up a real economy, one geared to illustrate monkeys' economic preference patterns and compare such patterns with those of human participants.

We began by presenting monkeys with novel tokens—small metal discs. Once monkeys became familiar with the tokens, we then began reinforcing them for an exchange, basically handing them food whenever they gave one of their tokens to a human experimenter. Monkeys quickly got the hang of the set-up, spontaneously taking tokens when they were available and waiting patiently with the tokens for a human willing to trade (see Figure 4.1 for a

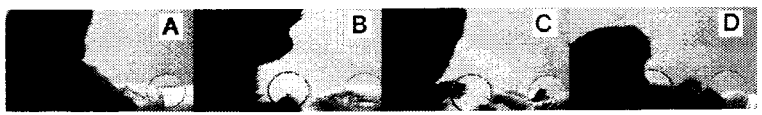

Figure 4.1. Depiction of the Trading Situation. The capuchin trader delivers a token to the salesman's hand (A). He then takes it away (B), and delivers a piece of food in return (C), which the monkey then takes and eats (D).

depiction of a single exchange). Within a few weeks of this simple training, our monkey economy was born. Each morning before testing, we handed monkeys a wallet of tokens and let them enter into a "market," in which two different experimental salesmen—two different research assistants dressed in dissimilar outfits—offered different kinds of food at different prices. Monkeys were then given a set amount of time in the market in which they could spend their wallet of tokens however they chose, buying goods from one or both of the salesmen. With this set-up, we could observe monkeys' preferences just as economists observe human preferences—by looking at what options the monkeys are interested in buying and the proportion of their budget they're willing to spend to obtain these different options.

Our first task was to demonstrate that our new monkey market worked in an economic sense, that is, that monkeys acting in our new market would reveal preferences that generally remained stable over time and responded rationally to price changes (Chen *et al.*, 2006). To assess this, we presented the monkeys with a market that involved a choice between two equally preferred goods, which for most monkeys involved a slice of Jello and a small apple chunk. Monkeys liked these two types of food about equally, so they tended to purchase them at equal rates—spending half of their budget on Jello and half on apples. We then presented monkey buyers with a situation in which the price of one of the two equally preferred goods fell by half—monkeys who previously paid one token for a small chunk of apple now got two chunks for the same price. If monkeys, like humans, adjust their preferences based on the price of the objects, then they should buy more of the cheaper good. Monkeys did just this, switching to buying more apples and avoiding goods that cost more. Like human economies, our newly introduced monkey market seemed to pay attention to price shifts, with their previously established preferences remaining stable across changes.

Having established that monkeys performed rationally in the types of situations where human agents performed rationally, we then turned to our real question of interest: whether monkeys would demonstrate the *irrationalities* that humans classically demonstrate—reference dependence and loss aversion. Put more specifically, could we get monkey buyers to change their preferences

based simply on how a salesman framed his offer? To explore these questions, we presented monkeys with a market in which the two experimenters each offered different numbers of apples. Importantly, however, the experimenters in this market didn't always hand over the number of food pieces they originally displayed—some experimenters gave more pieces of food than they originally offered, whereas others gave less. Our hypothesis was that monkeys might decide between the two offers not just based on how much food they get, but how much food they get relative to what they were originally offered. In other words, we predicted that monkeys might use the original offer as a reference point, and frame their final outcome not in terms of its overall value, but instead, as a loss or gain relative to that initial reference point, with less food being considered a relative loss and more food being considered a relative gain.

In the first study (Chen *et al.*, 2006), monkeys had a choice between two experimenters. The first experimenter offered two pieces of food but, half the time, only handed over one of the promised pieces of food. As such, this experimenter represented a risky chance of losing one piece of food. The second experimenter always offered one piece of food but, half the time, added a second unexpected piece of food. In contrast to the first experimenter, the second experimenter represented a risky chance of gaining one piece of food. As such, the only thing that differed across the two experimenters was how their actual offering was framed relative to what they had originally promised. Overall, both of the two experimenters delivered the same amount of food; on average, each experimenter gave 1.5 pieces of apple. But our monkey participants did not treat the two offers similarly—all of our monkey buyers significantly preferred the experimenter who appeared to give them gains, the salesman who promised only one piece and sometimes gave more. Like humans, our monkey buyers seemed to like receiving things that appeared to be gains more than they liked experiencing perceived losses, even though this didn't increase the overall amount of food they received on average.

But were our monkey buyers seeking out perceived gains or were they, like human actors, actively avoiding losses? To test this, we presented monkeys with a choice between an experimenter who always delivered what he initially offered and an experimenter who always delivered less than he initially offered. Monkeys chose

between one experimenter who promised and delivered one chunk of apple and an experimenter who promised two chunks and always delivered only one. Again, monkeys avoided the experimenter giving losses, and this time they showed an even stronger preference than in the previous study. Monkeys thus seemed to avoid a 100 per cent chance of loss even more than they avoided a 50 per cent chance of loss.

In a later study (Lakshminarayanan *et al.*, in preparation), we went on to explore whether monkeys' risk-taking behavior is also susceptible to framing effects. To do so, we developed a monkey version of Kahneman and Tversky's Asian disease problem. We presented monkeys with a choice between two salesmen that varied in their riskyness: the first salesman was a safe trader—he always did the same thing on each trial. The second salesman, in contrast, was a risky trader—he varied what he did from trial to trial. We then altered how the risky and safe salesmen framed what they offered. In the first condition, both salesmen framed their offer as a gain. The safe trader promised one piece of apple and, on every trial, added a second piece to his actual offer. The risky trader on the other hand always began by offering one piece of apple, but then varied his final offer: on some trials he gave a large gain of two apples (for a resulting offer of three) while on other trials he gave no bonus. With this set-up both the safe and the risky trader resulted in an average absolute offer of two pieces of apple. Nevertheless, our monkey subjects did not treat them equally. Our monkey buyers preferred to trade with the safe experimenter over the risky experimenter. The second condition, in contrast, presented monkeys a choice between risky and safe options that were instead framed as losses. Monkeys were allowed to choose between a safe salesman and a risky salesman, each of whom began by offering three chunks of apple. Both salesmen then delivered a loss relative to this initial reference point: the safe salesman always delivered a small loss of one apple piece (resulting in an offer of two pieces), whereas the risky salesmen sometimes delivered no loss (resulting in an offer of three pieces) and sometimes delivered a big loss (resulting in an offer of only one apple piece). Like human subjects, monkeys changed their risk seeking when problems were framed as losses rather than gains. When the options were framed as losses, monkeys

behaved more risk-seeking—they reliably preferred to trade with the risky salesman over the safe salesman. Even when evaluating when to take risks, monkeys seem to respond to how a particular problem is framed and whether options appear to be gains or losses relative to an arbitrary reference point.

In our most recent study (Lakshminarayanan, Chen, and Santos, in preparation), we examined whether our monkeys' loss aversion would also lead them to experience an endowment effect, an asymmetry between their willingness-to-buy and willingness-to-sell the same good. To test this, we identified another set of two different foods that the monkeys preferred about equally, in this case, cereal and pear pieces, and therefore purchased at equal rates. We then changed the trading task just a bit. Rather than provide the monkeys with a wallet of tokens, we instead gave them a wallet of one of these two goods. Half our monkeys, for example, became an owner of several wheat cereal pieces, which they could then trade for equally valued pear slices. Because these pear and cereal pieces are equally preferred, one might predict that monkeys should be willing to trade about half of their cereal pieces for pear slices. However, if monkeys, like humans, experience an endowment effect then they should be reluctant to trade the good that they own for an equally preferred good that they do not own. This is just what we observed. Like humans, monkeys who are made owners of one kind of food are reluctant to trade this food for another equally preferred option. Even across control conditions in which monkeys are compensated for any perceived transaction costs, our capuchin participants failed to trade the owned good for an equally preferred or slightly more valuable good. As in humans, it seems that when a monkey becomes an owner of an object, his views about the value of that object change. Monkey owners seem to avoid losing objects that they already own even if it means forgoing the gain of an equally valued alternative object. As with loss averse humans, monkeys' willingness-to-sell an item is less than their willingness-to-buy an identical item.

Taken together, the studies described above suggest that monkeys share a number of the irrational tendencies that humans demonstrate. First, they seem to evaluate their choices not in terms of their overall utility, but instead in terms of arbitrary reference points—how much food they currently have or how much food

they are expecting. In addition, monkeys seem to pay more attention to losses than they do to gains, demonstrating an asymmetry in the amount of work they will expend to avoid losses as opposed to seeking out equally sized gains. Such loss aversion also plays out in the way monkeys deal with risk—they seem to choose risky situations more when they think they are gambling to lose than when they think they are gambling to gain. Finally, monkeys' loss aversion also leads to problems with trading that humans experience; like human traders, monkeys demonstrate an endowment effect when trading equally priced goods.

5. INNATE IRRATIONALITY? CHALLENGES FOR PSYCHOLOGISTS (AND PHILOSOPHERS)

I began this article with the claim that humans and other animals do not behave according to the tenets of the most widely accepted normative account of decision-making, the model of rational expected-utility-maximization. Having reviewed evidence that human and primate participants behave in ways that systematically violate the predictions of rational expected-utility-maximization models, I have argued that a better descriptive model of human and non-human decision-making involves evaluating payoffs in terms of their relative value rather than their absolute utility. This new model incorporates two cognitive biases—reference dependence and loss aversion—that produce framing effects in a number of situations. In addition, our new evidence seems to suggest that such effects are not the result of economic experience: capuchin monkeys (and possibly other animals, see Marsh and Kacelnik, 2002) seem to employ the same biased decision-making strategies as do human participants.

The findings reviewed above suggest that humans and animals share a set of decision-making strategies that deviate from classic normative utility maximization models. My colleagues and I have interpreted these collective findings not only as evidence that capuchins (and probably other animals) share the underlying decision-making heuristics that give rise to the irrationality of human choice, but also that these biases result from homologous cognitive mechanisms, ones that stem from a shared innate evolutionary history.

Let me close this chapter by very briefly highlighting three sorts of challenges that this comparative line of research raises for a psychological (and philosophical) understanding of the evolution of decision-making and rationality more generally. The first challenge concerns the claim that the biases we have observed in capuchin monkeys are part of an innately specified system and the question of how these purportedly innate biases came about over evolutionary time. Typically, when psychologists talk about an innately specified cognitive strategy, they at least implicitly assume that the strategy they're talking about is one that works pretty well. Take, for example, the nativist claim that human infants possess innate cognitive strategies for tracking objects as they move through time and space (see e.g. Spelke *et al.*, 1992). Although many have taken issue with the evidence for such an innate object knowledge system (e.g. Smith, 1999), no one would argue that such an innate system would constitute a bad evolutionary idea. In fact, possessing an innately specified set of cognitive strategies to accurately make sense of the physical world can only be thought of as an evolutionary blessing. Intuitively, at least, this *does not* seem to be the case for an innate system for making biased decisions (see also Santos and Lakshminarayan, in press). Rather than an evolutionary blessing, a set of cognitive strategies that systematically leads an individual away from normatively better choices would seem to be a pretty major evolutionary inconvenience, one that might systematically reduce an organism's chances of maximal reproductive success. How, then, could an innate system of biased decision-making strategies have come about over evolutionary time? To the extent that the strategies I've described cause both humans and non-humans to behave in ways that systematically deviate from what normative models dictate, how could they have been selected via natural selection? The question of whether and how these biased decision-making strategies are able to confer a selective advantage is a major challenge facing this research program, one with which philosophers interested in the evolution of rationality may have to contend.

The next set of challenges facing the comparative research program I've described—and one that also plagues work in human behavioral economics—arises from considering the contrast between poor individual performance on the tasks described above and the proficient aggregated performance demonstrated by groups

of those same individuals in a variety of macroeconomic situations. Whereas individual human decision-makers regularly make decisions that violate the maxims of rationality and utility-maximization, markets—which are in some ways just aggregates of individual decision-makers—*do not* seem to behave in these ways. Indeed, neo-classical economists critical of experimental work in behavioral economics have long pointed out that markets, on average, seem to approximate what rational expected-utility-maximization models predict. A parallel question arises for aggregates of non-human decision-makers. Whereas the studies described above suggest that at least one non-human species (the capuchin monkey) demonstrates irrational decision-making behavior involving reference dependence and loss aversion, most species do impressive jobs of maximizing their expected energy and mating payoffs over time. How could seemingly irrational strategies like framing and reference dependence at the proximate or individual level translate into seemingly rational strategies of payoff-maximization at the ultimate or group level? The idea that strategies at the individual level do not necessarily translate into strategies at the aggregated level suggests that researchers must be careful when discussing the idea of rationality. In particular, those interested in the evolution of rationality might need to be more careful in specifying the proper unit of analysis for the investigation—either at the level of individual strategies or at the level of an aggregated group-level strategy.

A third challenge facing researchers interested in rationality concerns the question of what these new (possibly adapative) descriptive models of decision-making mean for normative models of behavior and choice. Gigerenzer and colleagues (e.g. Gigerenzer and Selton, 2001; Gigerenzer and Todd, 1999) have forcefully argued that "irrational" strategies may outperform rational expected-utility-maximization models when time and energy are of the essence. Because rational models require agents to spend unusually large amounts of time computing difficult-to-calculate expected payoffs, organisms may sometimes do better by simply "satisificing" with a rough yet quick solution to complex problems. Prospect theory, then, may provide not just a better descriptive account of organisms' actual reasoning, but it might also be a normatively better way to make decisions than classic models because

relative judgments require less time and computation. In this way, an organism might be better off bearing the cost of irrationality's imprecision and making relative judgments than they would be bearing the cost of time and calculation energy required to make a rational estimate. If Gigerenzer and colleagues' view of biased thinking is correct, philosophers and psychologists may need to radically rethink the status of traditional rationality as a *normatively viable* decision-making model.

The first goal of this article was to provide yet another empirical challenge to the once (and—in some circles—*still*) dominant descriptive account of human and animal decision-making, the idea that organisms make decisions in ways that serve to maximize their expected utility. To do so, I describe new work demonstrating that humans are not the only species to violate the predictions of expected utility models. Capuchin monkeys, who lack cultural training and experience in economic markets, exhibit biases that are analogous to those of human subjects in classic behavioral economic studies. The second goal of this article was to briefly consider what these findings mean for the nature and development of human decision-making strategies. My lab's capuchin results suggest that the decision-making heuristics used by human adults may emerge in the absence of culture, pedagogy, and experience. As such, I have interpreted these findings as evidence that our human decision-making biases result from innate cognitive strategies, ones with a lengthy evolutionary history. My hope is this new evidence for irrational decision-making in closely related non-human primates will help to inform and constrain the way that philosophers come to think about the broader question of rationality and its evolution.

REFERENCES

Brosnan, S. F., and F. B. M. de Waal (2003) 'Monkeys Reject Unequal Pay', *Nature*, 425: 297–9.

____ and ____ (2004) 'Socially Learned Preferences for Differentially Rewarded Tokens in the Brown Capuchin Monkey (Cebus apella)', *Journal of Comparative Psychology*, 118: 133–9.

Camerer, C. F., G. Loewenstein, and M. Rabin (eds.) (2003) *Advances in Behavioral Economics* (Princeton).

Chen, M. K., V. Lakshminarayanan, and L. R. Santos (2006) 'The Evolution of our Preferences: Evidence from Capuchin Monkey Trading Behavior', *Journal of Political Economy*, 114: 517–37.
Genesove, D., and C. Mayer (2001) 'Loss Aversion and Seller Behavior: Evidence from the Housing Market', *Quarterly Journal of Economics*, 116: 1233–60.
Gigerenzer, G., and D. G. Goldstein (1996) 'Reasoning the Fast and Frugal Way: Models of Bounded Rationality', *Psychological Review*, 103: 650–69.
—— and R. Selten (eds.) (2001) *Bounded Rationality: The Adaptive Toolbox* (Cambridge).
—— and P. M. Todd (1999) *Simple Heuristics that Make us Smart* (Oxford).
Glimcher, P. W. (2003) *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics* (Cambridge).
Hardie, B. G. S., E. J. Johnson, and P. S. Fader (1993) 'Modeling Loss Aversion and Reference Dependence Effects on Brand Choice', *Marketing Science*, 12: 378–94.
Hastie, R., and R. M. Dawes (2001) *Rational Choice in an Uncertain World* (Thousand Oaks, Calif.).
Kahneman, D., and A. Tversky (1979) 'Prospect Theory: An Analysis of Decision under Risk', *Econometrica*, 47: 263–92.
—— and —— (2000) *Choices, Values, and Frames* (Cambridge).
——, J. L. Knetsch, and R. H. Thaler (1990) 'Experimental Tests of the Endowment Effect and the Coase Theorem', *Journal of Political Economy*, 98: 1325–48.
—— —— and —— (1991) 'Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias', *Journal of Economic Perspectives*, 5: 193–206.
——, P. Slovic, and A. Tversky (eds.) (1982) *Judgment under Uncertainty: Heuristics and Biases* (Cambridge).
Krebs, J. R., and N. B. Davies (1993) 'An Introduction to Behavioural Ecology', *Blackwell Scientific Publications*, 57.
Lakshminaryanan, V., L. R. Santos, and M. K. Chen (in preparation) 'The Evolution of Risky Choices: Framing Effects in Non-Human Economic Behavior', *Current Biology*.
Liv, C., G. C. Westergaard, and S. J. Suomi (1999) 'Exchange and Value in Cebus apella', *American Journal of Primatology*, 49: 74–5.
Marsh, B., and A. Kacelnik (2002) 'Framing Effects and Risky Decisions in Starlings', *Proceedings from the National Academy of Sciences*, 99: 3353–5.
Odean, T. (1998) 'Are Investors Reluctant to Realize their Losses?', *Journal of Finance*, 5: 1775–98.
Santos, L. R., and L. Lakshminarayan (in press) 'Innate Constraints on Judgment and Choice: Insights from Children and Non-Human Primates, in P. Carruthers (ed.), *The Innate Mind Foundations and the Future* (Oxford).

Smith, L. B. (1999) 'Do Infants Possess Innate Knowledge Structures: The Con Side', *Developmental Science*, 2: 133–44.
Spelke, E., K. Breinlinger, J. Macomber, and K. Jacobson (1992) 'Origins of Knowledge', *Psychological Review*, 99: 605–32.
Thaler, R. H. (1980) 'Toward a Positive Theory of Consumer Choice', *Journal of Economic Behavior and Organization*, 1: 39–60.
Tversky, A., and D. Kahneman (1974) 'Judgment under Uncertainty: Heuristics and Biases', *Science*, 185: 1124–31.
—— and —— (1981) 'The Framing of Decisions and the Psychology of Choice', *Science*, 211: 453–8.
—— and —— (1986) 'Rational Choice and the Framing of Decisions', *Journal of Business*, 59: 251–78.
—— and —— (1991) 'Loss Aversion in Riskless Choice: A Reference-Dependent Model', *Quarterly Journal of Economics*, 106: 1039–61.
Westergaard G. C., C. Liv, T. J. Chavanne, and S. J. Suomi (1998) 'Token Mediated Tool-Use by a Tufted Capuchin Monkey (Cebus apella)', *Animal Cognition*, 1: 101–6.
—— —— A. M. Rocca, A. Cleveland, and S. J. Suomi (2004) 'Tufted Capuchins (*Cebus apella*) Attribute Value to Foods and Tools during Voluntary Exchanges with Humans', *Animal Cognition*, 7: 19–24.