

- (2003) 'Introspection and Internalism', in Susana Nuccetelli (ed.), *New Essays on Semantic Externalism, and Self-Knowledge* (Boston), 257–76.
- (2004) 'Epistemic Probabilty', *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 Sceptics* (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*<sup>1</sup> 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.

<sup>1</sup> *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

