while of some clinical interest, aren't particularly relevant to the sorts of philosophical and legal issues under discussion here. Being entirely nonconscious, cases of somnambulance and the like sit at an extreme on the spectrum between nonconscious and conscious action.

However, the sorts of nonconscious processes we're interested in—those examined in much of the social psychology literature and upon which the situationist challenge (e.g., Doris, 2002) is based—are sophisticated processes that are operative only when the individual is conscious of *something*. The question, therefore, is not whether consciousness in the sense of awareness of *anything at all* is needed for agency/responsibility; as Levy's psychiatric examples bring out, it surely is. Rather, the issue is whether conscious *deliberation*—the conscious reflection and reasoning emphasized on traditional philosophical accounts—is necessary for agency/control. It is on this point that we part ways with the traditional view and, potentially, with Levy. We believe, on grounds of the sophisticated, integrative capacities of nonconscious processes (see Suhler & Churchland, 2009, for extended discussion) that conscious deliberation and reasoning are not a necessary condition of control and agency.

Where the question of responsibility is particularly momentous is, of course, in the context of the criminal law. The law is remarkably wise and sophisticated on many of the issues regarding the mental status of the defendant, being the long-haul product of much experience and thoughtful reflection and deliberation. A study of cases reveals just how the law currently takes into account the mental state of the defendant. (See, e.g., the clear and insightful book by Bonnie, Jeffries, & Low, 1986.) New important debates are now emerging concerning whether certain kinds of evidence based on findings in neuroscience should be admitted into evidence either in the liability phase of a capital case, or in the sentencing phase, or neither. (See, e.g., Baum, 2011.)

9 Evolutionary Insights into the Nature of Choice: Evidence from Nonhuman Primates

Ellen E. Furlong and Laurie R. Santos

When faced with a decision, say, whether or not to donate money to a charity, we generally feel as if we are free to decide in a way that satisfies our own plans and preferences. If we believe in the mission of the charity, we may choose to support it, but if we have other plans for our money (i.e., a vacation or a new car), we may not. However, a growing body of empirical research suggests that both our choices and our preferences are remarkably easily manipulated. Indeed, our choices can be unconsciously swayed by a variety of factors as irrelevant as the phrasing of the request (e.g., Tversky & Kahneman, 1981), whether we are in a group or alone (e.g., Darley & Latané, 1968), and even whether we are holding a hot cappuccino or an iced latte (e.g., Williams & Bargh, 2008). In addition, empirical work in social psychology suggests that our preferences are not as stable as we often assume; whether we prefer a particular charity, for example, can depend on whether we have recently been forced to work for that charity (e.g., Festinger & Carlsmith, 1959) or whether we've been incidentally exposed to its name in Internet ads (Zajonc, 1968). Amazingly, even though these seemingly irrelevant factors have profound effects on our preferences and decisions, we are generally unaware of their power; we would never naturally explain our donation to a particular charity by the temperature of our coffee or presence of a stranger.

The fact that such apparently irrelevant situational factors have a firm grip on both our decisions and preferences poses a few serious problems for the nature of human freedom and responsibility (see discussions in Doris, 2002; Nahmias, 2007; Harman, 1999). For example, how can we assume that people are free to act on their preferences if our decisions are deeply bound to irrelevant situational factors outside of both our awareness and control (see discussion in Nahmias, 2007)? Similarly, if we grant that decisions of moral importance are strongly bound by situational influences, how does this affect our notions of moral responsibility and

evaluation (e.g., Doris, 1998, 2002)? Finally, if preferences themselves are subject to situational factors, what does that mean for notions of free will and the idea that we act toward stable goals?

Unfortunately, the current chapter won't attempt to solve any of these big problems. Instead, we will use another set of empirical findings to add an even further descriptive wrinkle to all these problems of freedom and responsibility. Specifically, we will review our own recent work on decision making in nonhuman primates to demonstrate that many of the situational factors inherent in human decision making also control the decisions of our close nonhuman evolutionary relatives. We begin by presenting three empirical cases in which nonhuman primates' decisions and preferences are swayed by the same irrelevant situational factors that affect choice in humans. First, we show that capuchin monkey choice is controlled largely by how different decision problems are framed, suggesting that the contextual factors that affect human choice may influence other primates' choices via identical cognitive mechanisms. Next, we show that orangutan cooperation is controlled by a different irrelevant feature, namely, the currency in which the cooperative payoffs are framed. We then show that this random factor can have a profound effect on both humans' and orangutans' moral decision of whether to cheat. Third, we discuss how capuchin monkeys' preferences can be twisted by their past decisions, demonstrating that an irrelevant past random decision can strongly influence a monkey's future choices.

In presenting these three cases, we will attempt to make two claims about the relevance of comparative work for discussions of freedom and responsibility. First, we will argue that the problematic situational aspects of human decision making *may run far deeper* than even the vast body of evidence in adult human decision making might suggest; indeed, situational factors may be phylogenetically ancient influences, ones that are woven into the cognitive architecture of our species' decision making. Second, we will use evidence from "situationism" (cf. Doris, 2002) in nonhuman primate decision making to argue that humans may be *even more strongly bound* by these pervasive situational influences than researchers have suspected. Specifically, we will argue that similarities in human and nonhuman decision-making biases hint that situational factors may act in a more automatic and encapsulated way than has previously been thought. These two factors together, we will argue, mean that situational influences may be trickier to overcome than we think, which has important normative implications for how (and also whether) people should try to rise above these factors (see discussions in Doris, 2002).

The Power of Framing in Nonhuman Primate Decision Making

Some of the most heralded situational effects on our decision making are cases in which the outcome of a choice can feel very different depending on how it is worded. To see this effect in action, imagine that you're a policy maker considering how to combat a deadly disease that is expected to kill 600 people. You can choose one of two courses of action: one remedy in which 200 people will be saved for sure, and another in which there is a one-third probability that 600 people will be saved and a two-thirds probability that nobody will be saved. If you're like most people, you might favor the first and less risky option—the one in which you can be certain that at least 200 people will be saved for sure. Now imagine you're combating a different deadly disease which is also expected to kill 600 people. Now your choices are between the following: one remedy in which 400 people will die for sure, and a second in which there is a one-third probability that nobody will die and a two-thirds probability that 600 will die. Here, you might be inclined to be a bit more risky, choosing the option in which there's a one-third chance that no one will die. Indeed, most people show just this pattern of performance (e.g., Tversky & Kahneman, 1981), choosing the safe option in the first scenario and the risky option in the second. The problem, of course, is that the problems are totally identical—the only thing that differs across the two problems is how they're worded. Nonetheless, people seem to have very different intuitions about what would be best to do in each case. In this and many other situations, people's intuitions seem to be based solely on how a particular problem is worded or framed (see review in Kahneman & Tversky, 2000). In the above problem, when outcomes were described in terms of people dying (i.e., the number of people that would die), people were inclined to avoid a sure loss; when the mathematically identical choice was described in terms of survival rates (i.e., the number of people who would be "saved"), people changed their strategy and sought out safe options. Wording alone seems to control our intuitions, even when judging what to do in a life-or-death scenario.

The power of wording, in this case, seems to tap into a set of biases originally described by Tversky and Kahneman (1981). First, people tend to think of problems in relative, not absolute, terms. Rather than objectively examining a decision, say whether to sell a stock or to hold on to it, people exhibit *reference dependence*: They evaluate outcomes relative to a reference point (see Kahneman & Tversky, 1979). Kahneman and Tversky also observed that people tend to treat changes from a reference point differently depending on whether those changes were positive (gains) or

negative (losses). As evidenced in the scenarios above, people exhibit *loss aversion*, avoiding options that lead to less than the status quo. In line with this view, people tended to be risk averse when dealing with perceived gains—they chose sure gains over 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. Human loss aversion can also be observed in the fact that most average-salaried academics would pass up the following objectively rational gamble: a 50% chance to win $1,001 and a 50% chance to lose $1,000.

Although the biases originally made famous by Kahneman and Tversky have been well-documented in both experimental and real-world situations (e.g., Kahneman & Tversky, 2000), less work to date had addressed the ease with which people were able to overcome these biases (see the discussion in Chen et al., 2006). Are strategies like loss aversion and reference dependence easily overcome with a bit of cognitive control? Or are such strategies more encapsulated, perhaps an innate part of the way we make decisions? To explore these issues, we and our colleagues (Chen et al., 2006; Lakshminarayanan et al., 2008) decided to examine whether loss aversion and reference dependence extended beyond the human species. More specifically, we examined whether capuchin monkeys share human-like framing effects.

To do so, we (Chen et al., 2006) first introduced capuchin monkeys to a token trading economy in which they could trade tokens (small metal discs) with human experimenters in exchange for food rewards. The monkeys quickly picked up on this, readily placing their tokens in the hands of human experimenters and taking the food rewards that were offered. We then introduced monkeys to a "market" in which they were presented with choices of how to spend their token budget. Our first goal was to see if monkeys were using their token budget in some of the same ways that humans use actual money. To test this, we presented the monkeys with various decisions that humans make every time they enter a shop. For example, monkeys were presented with two experimenters offering the same kind of food reward (i.e., apple slices), but one experimenter offered more apples per trade than the other (i.e., one of the experimenters was offering apple slices on sale). If monkeys were using their token budget like humans, then they should quickly learn to trade with the experimenter offering food on sale rather than the experimenter offering the same food at "full price." Monkeys did this quite readily, robustly preferring to trade with the experimenter offering more food than one offering less. In fact, the monkeys' preferences for trading with the human offering a better deal were indistinguishable from those of humans engaged in a similar buying situation.

Once monkeys demonstrated that they seemed to be paying attention to the offers of the respective experimenters, we (Chen et al., 2006) then tested whether the monkeys would also evaluate their outcomes with respect to a reference point. Capuchins were given a choice between two new experimenters. One experimenter always offered the capuchin one apple slice and either gave the monkey this apple slice or added a second unexpected bonus apple slice. If monkeys consistently traded with this experimenter, their expected payoff was 1.5 apple slices. The second experimenter always offered two apple slices, and either gave the monkeys the offered two slices or took one of the slices away, offering only one slice. Thus, just like the first experimenter, the capuchin could expect to receive on average 1.5 apple slices from the second experimenter. If monkeys, like humans, are reference dependent, they should evaluate the outcomes with respect to the initial offer—they should judge the person offering one but occasionally adding one to make two as a "better deal" than the person offering two but occasionally subtracting one, even though the average outcome is exactly the same across the two experimenters (1.5). In fact, that is exactly what the monkeys did—capuchins preferred to trade with the experimenter offering one slice and adding to it than the experimenter offering two slices and subtracting one. Even though the expected payoff of both experimenters was the same, monkeys strongly preferred to trade with the experimenter framing the rewards in terms of a gain relative to the reference point than the experimenter framing the rewards in terms of a relative loss.

We then explored whether monkeys were also susceptible to loss aversion. We (Chen et al., 2006) presented monkeys with a choice between an experimenter who offered one piece of food and always gave the monkey that one piece of food and a second experimenter who offered monkeys two pieces of food but always removed one piece, presenting the monkey with only one. Although both experimenters offered the monkey the exact same amount of food (i.e., one piece), capuchins avoided trading with the experimenter offering a loss, preferring to trade with the experimenter who offered one and gave one than the experimenter who offered two and gave only one. Much like humans, monkeys seem averse to losses, avoiding experimenters who seemed to offer less food than they originally expected.

Finally, we (Lakshminarayanan et al., 2011) investigated whether capuchins' loss aversion also affected their preference for risk. Did monkeys, like humans, become more risk seeking in the face of losses? To test this, we allowed monkeys to choose between two new experimenters. One

experimenter was safe (always doing the same thing on each trial) while the other was risky (varying his behavior from trial to trial). In the first condition, both the safe and risky experimenters framed their offer as a gain. The safe experimenter showed one piece of apple and added a second one on every trial. The risky experimenter, in contrast, always began by offering one piece of apple and on some trials gave a large gain of two apples while on other trials he gave no bonus. Although the safe and the risky experimenters both gave an average absolute offer of two pieces of apple, the capuchins strongly preferred to trade with the safe experimenter over the risky experimenter. When faced with a choice of safe or risky gains, monkeys, like people, chose to go with the safe option. We then tested how monkeys reacted to risky and safe losses. Monkeys chose between safe and risky experimenters who each began by offering three pieces of apple but delivered less than this amount. The safe experimenter always took away one piece of apple (resulting in a sure offer of two pieces) whereas the risky experimenter sometimes took two pieces away (resulting in an offer of only one apple piece) and sometimes took no pieces away. In contrast to their performance with gains, monkeys in this condition reliably traded with the risky experimenter over the safe experimenter. Like humans, monkeys seem to act more risky when dealing with losses than with gains.

Taken together, capuchin monkeys appear to exhibit several of the biases that affect human choice. Monkeys are sensitive to the initial state when considering outcomes, evaluating their outcomes relative to a reference point. Monkeys also avoid outcomes framed as losses and are even willing to take on more risk to avoid the chance of a loss. Together, this work suggests that some of the classic framing effects observed in human choice may be evolutionarily old behavioral biases. In this way, work on capuchin monkey biases raises some important new questions about the extent to which humans may be more bound to these strategies than initially thought. Before turning to this issue, though, we first show that loss aversion and reference dependence aren't the only framing effects that may be evolutionarily old. Indeed, similarly deep-seated framing effects may affect primates' decisions in the moral domain as well.

How Our Cooperative Motivations Are Shaped by Unexpected Framing Effects

Humans face numerous moral situations in which we must decide whether or not to be nice to another individual. Intuitively, you might think that

such cooperative decisions come down to a set of normatively relevant decision variables, such as whether you like a potential cooperative partner, your political views, how well you think you may be able to help, and so on. Although these factors do appear to affect cooperative decisions, recent work suggests that many other less sensible factors affect our cooperative decisions as well.

One especially strange factor that appears to affect cooperative decisions is the type of currency people use to make a donation (Furlong & Opfer, 2009). Consider a strange real-world example that occurred in 2002, the year in which many European countries switched to the euro. Economists were surprised to notice that switching to the euro seemed to drastically alter some countries' donation behavior. People in some countries, such as Italy and Spain, drastically changed their donations to charity with the introduction of the euro whereas people in other countries, such as Germany and Ireland, did not change their behavior (Cannon & Cipriani, 2004). How can the currency a person uses affect how much he or she is willing to give to charity?

Insight into this strange effect may come from an unexpected source: constraints on our ability to distinguish numeric quantities. People's ability to discriminate two different numeric quantities relies on two dimensions: the size of the numbers to be discriminated and the distance between them (e.g., Banks & Hill, 1974; Moyer & Landauer, 1967; Starkey & Cooper, 1980). Generally people find it easier to discriminate differences between small numbers (i.e., 3 vs. 5) than to discriminate identical differences between large numbers (i.e., 13 vs. 15), a finding termed the *numeric size effect*. In addition to this numeric size effect, people also experience a *numeric distance effect*, in which discriminability depends on the distance between quantities. In other words, people more easily discriminate numbers with larger distances (i.e., 3 vs. 15) than numbers with smaller distances (i.e., 5 vs. 13). These numeric size and distance effects can be explained by a logarithmic representation of numbers in which we overestimate differences among small quantities and compress differences among large quantities (see Dehaene, 2007, for a review).

Understanding the logarithmic nature of our numeric representations has allowed researchers to gain some insight into changes in people's donation patterns after the introduction of the euro—donations may have changed because people's subjective sense of how much money they lost to charity changed after the introduction of the euro. For an Italian used to the lira, the change in currency made the loss of money more salient than when using a currency based on larger, less discriminable numbers.

However, the change in Irish currency was minimal, resulting in an indiscriminable change in the salience of currency lost to charity.

To explore how currency affects cooperative behavior a bit more systematically, we (Furlong & Opfer, 2009) tested the effect of currency on cooperation in a well-studied economic game known as the iterated prisoner's dilemma (IPD; e.g., Axelrod & Hamilton, 1981; Dawes & Thaler, 1988; Messick & Brewer, 1983; Rapoport & Chammah, 1965). The IPD is a cooperative game between two agents in which the most lucrative strategy is for both agents to engage in mutual cooperation because they can make more ($3 each) by mutually cooperating than by engaging in mutual defection ($1 each). However, players face a temptation to defect—if one player defects when the other player cooperates, the defector has the potential to earn ($5) while the cooperator earns nothing ($0). In this way, even though the optimal strategy in an IPD is to engage in mutual cooperation, when provided with prisoner's dilemmas like this, people often defect much more than would be optimal.

To test the effect of currency on people's intuition to defect, we (Furlong & Opfer, 2009) presented people with a standard IPD and investigated the effect of different currency units on cooperative decisions. People were either presented with their IPD payoff matrix in dollars (i.e., $3 for mutual cooperation) or in cents (i.e., 300¢). Because the same payoff amounts would be subjectively easier to discriminate when presented in dollars than when presented in cents, we predicted that people may be more tempted to defect when making decisions in dollars than when making an economically identical decision presented in cents. As found in previous studies, participants playing for dollars generally engaged in low rates of cooperation and high rates of defection. However, when playing for an equal amount of cents, cooperative behavior changed quite drastically; people in the cents condition engaged *in four times as much cooperation* as in the dollars condition. Even though the payoff structure was identical across the dollars and cents conditions, people drastically increased their rate of cooperation when playing for cents rather than for dollars. Even when all of the relevant aspects of their decision were identical across conditions, people's intuition about whether to defect was drastically shaped by the units used to describe their payoffs.

As in the case of the framing biases reviewed earlier, our tendency to shift cooperative motivations based on the units of a problem raises interesting questions about the nature of responsibility. Before turning to these, however, it's worth exploring how fundamental such biases are. Is our susceptibility to the units of a problem a strange feature of only some

human decisions, or does this tendency affect decision making in other primates as well?

To test how deeply this bias extends, we (Furlong et al., 2012) decided to explore whether numerical biases also change the cooperative decisions of other primate species. Specifically, we tested whether one nonhuman primate, the orangutan, would be biased to cooperate less when dealing with more discriminable units. Human and orangutan participants were given the choice to cooperate or defect based on different payoffs. We then varied the units in which different payoffs were presented while keeping the overall payoff value constant. As before, human participants were shown their payoffs in dollars ($3) or cents (300¢), while orangutan participants were shown payoffs in either grapes (3 grapes) or grape pieces, where each grape was cut into 10 pieces (30 grape pieces). Like humans, orangutans showed a robust effect of unit; orangutans engaged in low rates of cooperation when paid in grapes but showed higher rates of cooperation when given the exact same payoff value in grape pieces.

These data suggest that nonhuman cooperative tendencies may be just as susceptible to numerical effects as those of humans. In this way, how our minds subjectively compare numeric values seems to affect cooperation in a rather deep way. Indeed, this work suggests that even a morally relevant decision is subject to biases that may be evolutionarily quite old, and likely a deep part of our decision processes. We now turn to a final decision bias that appears to be evolutionarily old, the tendency to reevaluate our preferences based on our choices.

How Preferences Can Change Based on Our Decisions

One of the most common assumptions about the nature of free will is that we use our actions to achieve our goals and preferences. Within this notion, though, are a set of assumptions about the nature of preferences. First, it's assumed that we have access to our preferences—we can use them to guide our own actions. Second, it's assumed that preferences are in some sense stable; we have a set of reasonably consistent likes and dislikes that guide the choices we make.

Unfortunately, recent work in social psychology suggests that preferences might not be so straightforward. Indeed, much empirical work in judgment and decision making demonstrates that preferences are not stable, coherent features of the mind but rather are malleable, fragile, and in some cases may even be constructed on the fly (see review in Ariely & Norton, 2008). In one classic demonstration of this, Brehm (1956) gave

participants the chance to rate a set of household items. Afterwards, participants were given a choice between two of the items they had rated. The trick was that the items presented during this choice phase were two objects that the participant had liked equally; in this way, participants would presumably have to choose between the two items randomly. Brehm then explored how the act of making a random decision between the two items affected participants' subsequent preferences. Under most accounts of human preference, it would be crazy to think that the simple act of making a choice would influence what subjects liked about the household items—none of the items' features had changed after the decision and no new information about the objects became available through the act of choosing between them. Nevertheless, participants' preferences for the items changed drastically after making a choice. Critically, when asked to rerate all of the items, participants' ratings of the object they chose against went down. The act of choosing against an object seemed to make it less appealing. Indeed, the mere act of choice seems to affect what we like, even in cases where doing so gives us no new information about the objects in question.

The phenomenon of choice shaping our preferences has now been widely documented in social psychology, even in surprising cases where it's obvious that our decisions are random. Sharot and colleagues (2010), for example, gave people the opportunity to rate different potential vacation locations. Participants were then shown two of the vacation options and asked to choose between them. The trick, however, was that the names of the vacations were perceptually "masked" by a set of nonsense letters, making it impossible to tell which item was which; participants were thus asked to make a choice between items when they knew they had no idea which item was which. Sharot and colleagues then explored how making this clearly blind choice affected participants' preferences. When asked to rerate all of the items, people tended to prefer items they chose against less than they had originally. Even in the case of a clearly blind choice, people allow their decisions to alter their future preferences.

Are such choice-induced preference changes specific to the kind of complex decisions humans make, or are these processes a more fundamental aspect of the way preferences work in general? To get at this issue, we teamed up with colleagues (Egan et al., 2007; Egan et al., 2010) to investigate whether similar choice-based preference reversals take place in a nonhuman species, the brown capuchin monkey. Our goal was to develop a version of Brehm's classic study that could be used with nonverbal subjects. Our method presented capuchins with a novel food—differently

colored M&M's candy. Because differently colored M&M's taste the same, we assumed that capuchins might not initially have a preference for any particular colors. The question, then, was whether capuchins would develop such preferences merely through the act of choosing against one of the colors. Would capuchins also begin to dislike an M&M color that they randomly chose against? To test this, we presented monkeys with a choice between two M&M colors, say green and blue. Once subjects made their choice (e.g., they picked blue), we gave them a subsequent set of choices between the color they rejected (green) and a novel but equal tasting color (e.g., red). We found that capuchins tended to choose the novel M&M color, thereby derogating the option they had previously chosen against. Like humans tested in similar paradigms (Brehm, 1956), capuchins liked an M&M color less after they had previously chosen against it. Importantly, we observed this sort of derogation *only* after subjects had made their own choices; when monkeys were merely given one M&M color over another by a human experimenter, monkeys did not show a tendency to avoid the unreceived option.

In later studies (Egan et al., 2010), we also saw that choice can affect monkeys' preferences even in cases that only seem like a real choice, as in the case of Sharot et al. (2010). To test this, we presented capuchins with a situation that made them feel as if they had a real choice even though we had constrained their actual decision. Capuchins were allowed to choose one of two items that appeared to be placed in a box filled with wood shavings. Monkeys could make their choice by searching for and picking one of the items. What monkeys didn't realize, though, was that only one of the two options had been placed inside the box. Although it felt like an intentional choice to the monkeys, their choice was in actuality determined by the experimenter. The question was whether this constrained choice would still affect monkeys' future preferences. To test this, we again gave monkeys a choice between the item they appeared to reject and a novel one. As before, monkeys avoided they option they thought they chose against, despite the fact that we had completely constrained their choice. These results suggest that even forced choices can affect monkeys' preferences.

The capuchin work on choice-induced preference changes suggests that the act of making a decision can affect monkeys' preferences in much the same way as it affects human preferences—the mere act of making a decision can affect what a monkey likes and dislikes. The monkey results therefore demonstrate that the odd choice-induced preference changes observed in humans aren't the result of strange social psychology

experimental setups. Instead, our results suggest this tendency might be a deep feature of decision making, one that extends beyond the human species and might be pervasive across the decisions of many organisms.

What Do Nonhuman Primate Decision Biases Mean for Human Free Will and Responsibility?

The goal of this chapter was to review recent work on nonhuman primate decision-making biases in an attempt to see what such work has to say for philosophers interested in the nature of freedom and responsibility. Across three experimental domains, we've reviewed cases that violate our lay assumption that human choice operates in a rational way. People don't seem to make decisions in ways that willfully satisfy a set of stable preferences. We first learned that people's preferences can be affected by how a problem is worded or framed. Merely making a decision outcome seem like a loss can change people's preference for how much risk to take. We also saw that similar framing biases—in this case, the currency units in which a problem is presented—can affect people's intuitions about moral choices, namely, how much money to donate or how much to cheat in a prisoner's dilemma game. Finally, we saw that even the act of making a decision itself can mess with one's preferences; even blindly choosing between two unknown options can affect the extent to which people like those options later. These experimental findings in humans raise the possibility that choice and decisions don't work in the way that we've assumed. And this should be a very worrying prospect for most accounts of free will and responsibility. Indeed, these data have led a number of philosophers to propose more "situationist" accounts of choice and moral responsibility (e.g., see review in Doris, 2002).

Here, we've tried to take this work one step further—showing that it's not just human choice that works in this unusual way. As our work demonstrates, each of the strange phenomena observed in human choice seems to be present in the decision making of nonhuman primates—primates exhibit framing effects that can change their intuitions about how to behave, even on moral games, and also show choice-induced preference changes. In this way, the problems implicit in human choice appear much more fundamental than a few small effects observed in human laboratory studies. Instead, these situationist issues may be a more fundamental aspect of the way choices work writ large, the way all decisions work across species.

The fact that other species' decision making is as problematic for accounts of free will as that of humans in laboratory settings, we feel,

makes the human findings all the more difficult to sidestep. For example, one might have been tempted to ignore some of the human findings on the grounds that they involve relatively contrived decisions that take place in strange settings (e.g., stating whether you'd like to gamble on a verbal survey). One might therefore assume that while these biases could affect choice in theory, they don't really affect choices in ways that matter for real-world decisions (though see Danziger, Levav, & Avnaim-Pesso, 2011, for at least one real-world case where choices are affected by different frames). Our primate findings complicate this interpretation, however, as our work suggests that organisms may show similar biases on completely different tasks, often ones with real-world relevance (i.e., foraging decisions). One could also come up with a different alternative sidestep of the human results, perhaps assuming that the human social psychological findings should be discounted in part because they're almost exclusively observed in Western populations (for similar logic, see Henrich, Heine, & Norenzayan, 2010). Again, the monkey work poses a problem for this account; the fact that some classic human framing effects are present in capuchin monkeys suggests that these biases are shared across species separated by over 35 million years of evolution. As such, it's unlikely that such effects can be culturally idiosyncratic in the ways one might have expected just from the initial human studies. Instead, the primate work suggests that the decision-making biases we've reviewed are likely to be universal features of human choice, ones that transcend educational level, political preferences, or cultural background.

Perhaps most importantly, the primate work hints that some strange aspects of human choice may be *deep features* of the way our decisions are made. Our work suggests that situations and frames can change preferences across species and thus that these processes are a fundamental aspect of the way decision making has evolved. The fact that situational influences are a phylogenetically ancient part of human decision making suggests that typical notions of free will—ones that assume freely chosen decisions across stable preferences—may be relatively untenable. Decisions don't seem to work that way in people, and they might not have worked that way across our primate ancestors either. This additional descriptive wrinkle we think adds even more credence to views of the nature of free will that argue we're more affected by unconscious situational factors than we realize, and that such factors need to be more adequately taken into account (e.g., Nahmias, 2007).

Second, and perhaps even more critically for philosophers, primate situational biases provide some important new insight into the extent to

which we are likely to override such biases. Our work suggests that some of our biases may be evolutionarily old tendencies, ones that natural selection shaped into unconscious cognitive mechanisms over many millions of years. Evolved tendencies of this sort tend to be rather tricky to override. Consider, for example, how tricky it is to overcome the preference that natural selection gave us to seek out sweet and fatty foods. In the same way, it's possible that the decision-making strategies we've observed might be pretty encapsulated—they might be hard to overcome even in cases where we recognize them operating. In this way, the primate findings reviewed hint that it's unlikely people will easily overcome the situational influences that affect their decisions. The fact that we may be more trapped in these biases than we think raises some interesting questions about our responsibility for such decisions. Should people be morally praised for being generous in a situation where we have reason to suspect they couldn't easily discriminate the amount of money they'd receive as a payoff? Should we blame people for making risky decisions when we know they were thinking of their choices in a loss frame? Such issues about responsibility quickly surface when dealing with the power of situational influences (see elegant discussions in Doris, 1998, 2002). The primate work we've reviewed suggest that situational influences that affect human choice might not just be powerful factors that are hard to overcome; instead the primate work suggests that such tendencies may be deeply encapsulated, perhaps even impractical to overcome. In this way, the primate work suggest that some aspects of the situation may affect us so fundamentally that it would be unreasonable to expect people to behave in the absence of their influences. As such, the findings we've reviewed demonstrating similarities in human and nonhuman biases have deep and important normative implications for the nature of human responsibility.

Although we haven't solved (surely any of) the philosophical questions surrounding issues related to free will and responsibility, we hope we've provided thinkers with an important new set of descriptive data relevant to these big questions. By incorporating data on nonhumans into the mix, we hope that those who ponder the nature of human freedom will be able to gain even more insight into how situational influences can and must be incorporated into a reasonable account of human free will and moral responsibility.

9.1 Is Human Free Will Prisoner to Primate, Ape, and Hominin Preferences and Biases?

Brian Hare

As someone focused on understanding the evolution of human psychology, I do not spend much of my time thinking about free will. It is not that free will is an uninteresting psychological concept, it is just it traditionally has not lent itself to empirical study from an evolutionary perspective. Furlong and Santos's pioneering set of studies show how evolutionary tests relevant to issues of human free will are now very possible. In fact their comparative approach seems to challenge a number of assumptions regarding the very origins of our preferences. In my comments I want to build on their article by illustrating that there are even more ways that an evolutionary approach can help in testing many of our ideas about our free will.

Free to Take an Evolutionary Perspective

Over 30 million years ago a population of primates split in two. One group evolved into the monkeys and the other into the apes. Six million to 7 million years ago there were dozens of species of apes living across Africa and Asia. Again a population from one of the African species split and evolved into the chimpanzee and hominin subfamilies. Today there are two chimpanzee species (bonobos and chimpanzees) and just one remaining hominin species. However, only a few million years ago there were at least half a dozen species of hominins (Steiper & Young, 2006). Paleoanthropologists begin to recognize several archaic forms of *Homo sapiens* in the fossil record at around 200,000 years ago with fully anatomically modern *Homo sapiens* being recognized just over 50,000 years ago. However, of course, our species shared the planet with Neanderthals, Denisovans, and Flores people for millennia, only becoming Earth's sole human occupant around 12,000 years ago (Churchill, 1999; Falk et al., 2005; Meyer et al., 2012).

PHILOSOPHY / COGNITIVE SCIENCE

Moral Psychology

VOLUME 4 Free Will and Moral Responsibility

edited by **WALTER SINNOTT-ARMSTRONG**

Traditional philosophers approached the issues of free will and moral responsibility through conceptual analysis that seldom incorporated findings from empirical science. In recent decades, however, striking developments in psychology and neuroscience have captured the attention of many moral philosophers. This volume of *Moral Psychology* offers essays, commentaries, and replies by leading philosophers and scientists who explain and use empirical findings from psychology and neuroscience to illuminate old and new problems regarding free will and moral responsibility.

The contributors—who include such prominent scholars as Patricia Churchland, Daniel Dennett, and Michael Gazzaniga—consider issues raised by determinism, compatibilism, and libertarianism; epiphenomenalism, bypassing, and naturalism; and rationality and situationism. These writings show that although science does not settle the issues of free will and moral responsibility, it has enlivened the field by asking novel, profound, and important questions.

WALTER SINNOTT-ARMSTRONG is Chauncey Stillman Professor of Practical Ethics at Duke University and the editor of the previous volumes of *Moral Psychology*, all published by the MIT Press.

"This is an outstanding collection of work at the intersection between science and traditional approaches to moral responsibility (and free will). The contributors are among the very best people working in these areas. The book is strong evidence for the contention that progress in understanding freedom of the will and moral responsibility is often enhanced when philosophers and scientists work collaboratively. Highly recommended."
—John Martin Fischer, University of California, Riverside

"Clear, logical, and cogent are not often words associated with discussions of free will or morality—yet the chapters in this book are just that. Experts who have done a lot of thinking about these issues raise points that I would guess not many readers will have thought of, and the diverse array of perspectives is explained so well that readers who are not immersed in the science, philosophy, and debates about free will and moral responsibility will learn a lot and keep reading. I certainly did—and then immediately started telling colleagues to read this book."
—Kathleen Vohs, Co-editor of *Free Will and Consciousness: How Might They Work?*

"With this volume, Sinnott-Armstrong has created a much-needed point of entry into the free will debate. Drawing heavily on the neuroscience of choice, he deftly balances a dizzying array of perspectives built in large part on empirical facts. You experience free will emerging as an exciting yet profoundly complicated scientific challenge."
—Scott Grafton, University of California, Santa Barbara

The MIT Press
Massachusetts Institute of Technology
Cambridge, Massachusetts 02142
http://mitpress.mit.edu

978-0-262-52547-3

Moral Psychology VOLUME 4 Free Will and Moral Responsibility SINNOTT-ARMSTRONG, ed.

Moral Psychology

VOLUME 4

Free Will and Moral Responsibility

edited by
WALTER SINNOTT-ARMSTRONG