*Journal of Economic Perspectives—Volume 30, Number 3—Summer 2016—Pages 189–212*

# Motivated Bayesians: Feeling Moral While Acting Egoistically

Francesca Gino, Michael I. Norton, and Roberto A. Weber

A growing body of research yields ample evidence that individuals' behavior often reflects an apparent concern for moral considerations. Using a broad definition of morality—to include varied non-egoistic motivations such as fairness, honesty, and efficiency as possible notions of "right" and "good"—economic research indicates that people's behavior often reflects such motives (Fehr and Schmidt 2006; Abeler, Becker, and Falk 2014). Perhaps this should not come as a surprise to economists, given that Adam Smith prominently highlighted such motivations in *The Theory of Moral Sentiments* in 1759—17 years before *The Wealth of Nations*.

A natural way to interpret evidence of such motives using an economic framework is to add an argument to the utility function such that agents obtain utility both from outcomes that yield only personal benefits and from acting kindly, honestly, or according to some other notion of "right" (Andreoni 1990; Fehr and Schmidt 1999; Gibson, Tanner, and Wagner 2013). Indeed, such interpretations can account for much of the existing empirical evidence. However, a growing body of research at the intersection of psychology and economics produces findings inconsistent with such straightforward, preference-based interpretations for moral behavior. In particular, while people are often willing to take a moral act that imposes personal

■ *Francesca Gino is the Tandon Family Professor of Business Administration, Harvard Business School, Boston, Massachusetts. Michael I. Norton is the Harold M. Brierley Professor of Business Administration, Harvard Business School, Boston, Massachusetts. Roberto A. Weber is Professor of Economics, University of Zurich, Zurich, Switzerland. Their email addresses are fgino@hbs.edu, mnorton@hbs.edu, and roberto.weber@econ.uzh.ch.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
http://dx.doi.org/10.1257/jep.30.3.189

doi=10.1257/jep.30.3.189

material costs when confronted with a clear-cut choice between "right" and "wrong," such decisions often seem to be dramatically influenced by the specific contexts in which they occur. In particular, when the context provides sufficient flexibility to allow plausible justification that one can both act egoistically while remaining moral, people seize on such opportunities to prioritize self-interest at the expense of morality. In other words, people who appear to exhibit a preference for *being* moral may in fact be placing a value on *feeling* moral, often accomplishing this goal by manipulating the manner in which they process information to justify taking egoistic actions while maintaining this feeling of morality.

As an example of how such motivated beliefs help people easily reinterpret their egoistic behavior, consider Fritz Sander, a German engineer employed by Topf & Sons during World War II, whose work included designing—and attempting to patent—more efficient incineration devices for use in Nazi concentration camps.[1] Following the war, he justified his actions as morally consistent with his professional obligations: "I was a German engineer and key member of the Topf works, and I saw it as my duty to apply my specialist knowledge in this way to help Germany win the war, just as an aircraft construction engineer builds airplanes in wartime, which are also connected with the destruction of human beings" (as quoted by Fleming 1993). Such justifications abound in less-extreme cases as well. Enron chief executive officer Jeffrey Skilling (2002 [2011]), following the firm's bankruptcy and his own convictions for conspiracy and fraud, testified: "I was immensely proud of what we accomplished. We believed that we were changing an industry, creating jobs, helping resuscitate a stagnant energy sector, and … trying to save consumers and small businesses billions of dollars each year. We believed fiercely in what we were doing." Skilling proceeded to testify that he was "not aware of any inappropriate financing arrangements" and that he had left the company "solvent and highly profitable." Of course, one could ask whether Sanders and Skilling were simply lying—that is, knowingly attempting to exculpate their misdeeds by arguing that they were the product of nobler motives. Research on self-serving approaches to morality, however, suggests that they may have processed information and facts in a biased way, allowing them to feel that their questionable behavior was morally justifiable.

In this paper, we will argue that there is a widespread tendency for individuals to exploit justifications and uncertainties present in decision-making environments in order to act egoistically—and, possibly, dishonestly or unethically—without feeling that what they are doing is "bad." That is, people often appear concerned less with the morality of their actions or the outcomes they produce, and more with what the actions they take reveal about them as moral beings. People want to believe they are moral, and prefer actions that support this belief—sometimes independently of whether those actions are themselves actually moral. To facilitate this belief, people often acquire and process information about what is "moral" or "immoral"

[1] See "Topf & Songs as Partners of the SS—The Patent Application" (http://www.topfundsoehne.de/cms-www/index.php?id=120&l=1).

in self-serving ways—and these biased beliefs, rather than a preference for morality itself, may drive much human behavior in contexts involving morality. In decisions involving morality, we argue that people often act as "motivated Bayesians"—while they gather and process information before and during the decision-making process, they tend to do so in a way that is predictably biased toward helping them to feel that their behavior is moral, honest, or fair, while still pursuing their self-interest. Hence, while classical Bayesians will both seek out the most informative evidence and process it in an unbiased way, motivated Bayesians will also be influenced by the evidence that they encounter but will be biased both in choosing which information to acquire and in their interpretation of such information in order to facilitate beliefs in their own morality.

We begin by describing psychological research on motivated reasoning, the domain-general process by which people's goals and emotions influence the manner in which they collect and evaluate information during decision making. We then discuss two ways in which people act as motivated Bayesians when faced with moral decisions, each having the property that people interpret evidence self-servingly to facilitate egoistic behavior at the expense of some moral concern: self-serving judgments of morality and self-serving interpretations of reality. First, we argue that people often form self-serving judgments of what, exactly, constitutes fair or moral behavior or outcomes. When there is some flexibility in interpreting what is "right" and "wrong" or "moral" or "immoral," people's judgments of the morality of an act are often biased in the direction of what best suits their interests. Second, we argue that a similar but distinct phenomenon occurs when people actually alter their judgments of objective qualities—such as their own abilities or the quality of competing options—as a way of making egoistic behavior appear more moral. Finally, we argue that motivated Bayesian reasoning in moral decision making has important implications for many behaviors relevant for economics and policy. In domains including charitable giving, corruption and bribery, and discrimination in labor markets, the ability of people to pursue egoistic objectives while maintaining a belief in their own morality has important consequences for their behavior.

## Motivated Reasoning and Motivated Bayesians

Decades of research in psychology shows that people care about their self-concept and expend a great deal of effort maintaining a positive image of the self, often by engaging in motivated reasoning (Steele 1988; Kunda 1990). Kunda (1987), for example, shows that people's explanations for the successes of others tend to reflect favorably on themselves: when asked to indicate the extent to which several factors had contributed to the success of a target person's marriage, participants rated attributes that they personally possessed (such as being the youngest child, or having an employed mother) as more important than characteristics they did not possess. In other words, participants' templates of success in marriage were self-serving. Similarly, people are quick to attribute their successes to their own qualities

("I got an A because I am smart") but their failures to situational factors ("I got an F because the professor is an idiot") (Weiner 1985).

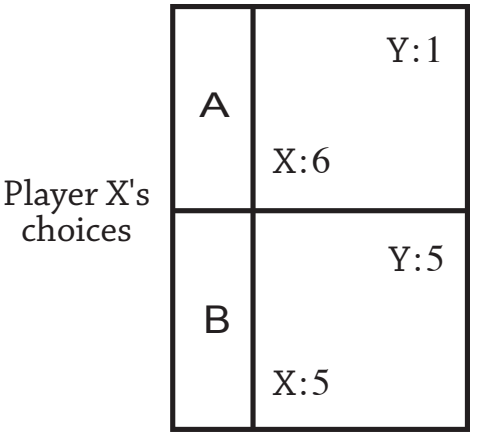
However, a crucial aspect of such motivated reasoning is that this ability to manipulate is not without limit. As noted by Kunda (1990, p. 480), people reach the conclusions they want to reach, "but their ability to do so is constrained by their ability to construct seemingly reasonable justifications for these conclusions." In short, people cannot simply believe anything they want to believe, but are instead—at least in part—constrained by the evidence they encounter and the conclusions that might plausibly be supported by such evidence.

We use the term "motivated Bayesian" to describe this general type of biased information processing. In textbook Bayesian reasoning encountered in introductory statistics courses, people have probability distributions of prior beliefs and then update these beliefs with an unbiased evaluation of any new evidence they encounter. Motivated Bayesians bias this process, for example, by ignoring or underweighting unfavorable evidence or by manipulating the inferences that they draw from the evidence.[2] For example, in a choice context involving morality, a motivated Bayesian has prior beliefs about her own moral qualities. Making, say, an egoistic choice at the expense of some moral objective or obligation should lead an unbiased Bayesian to update (in this case, by downgrading) her beliefs about her moral qualities. However, a motivated Bayesian confronting such a choice will manipulate the information she acquires and how she processes that information in order to reach the conclusion that her egoistic behavior is, in fact, not reflective of immorality. That is, people can be quite creative at manipulating their perceptions of a situation in order to make egoism appear "not that bad" from a moral perspective.

As a specific example, consider the situation analyzed by Batson, Kobrynowicz, Dinnerstein, Kampf, and Wilson's (1997) study, in which participants in a laboratory experiment distribute two tasks between themselves and another participant: a positive task (where correct responses to a task earn tickets to a raffle) and a negative task (not incentivized and described as "rather dull and boring"). Participants were informed: "Most participants feel that giving both people an equal chance—by, for example, flipping a coin—is the fairest way to assign themselves and the other participant to the tasks (we have provided a coin for you to flip if you wish). But the decision is entirely up to you." Half of participants simply assigned the tasks without flipping the coin; among these participants, 90 percent assigned themselves to the positive task. However, the more interesting finding is that among the half of participants who chose to flip the coin, 90 percent "somehow" ended up with the positive task—despite the distribution of probabilities that one would expect from a two-sided coin. Moreover, participants who flipped the coin rated their actions as

[2]This description falls within the more general perspective of treating people as "quasi-Bayesians" in behavioral economic theory (Camerer and Thaler 2003). Under this modeling approach, people make a few systematic mistakes in how they process information, but otherwise employ Bayesian inference procedures. An example of motivated quasi-Bayesian information processing that shares features with the processes we describe is provided by Rabin and Schrag's (1999) model of "confirmatory bias."

*Figure 1*
**Baseline Game from Dana, Weber, and Kuang (2007)**



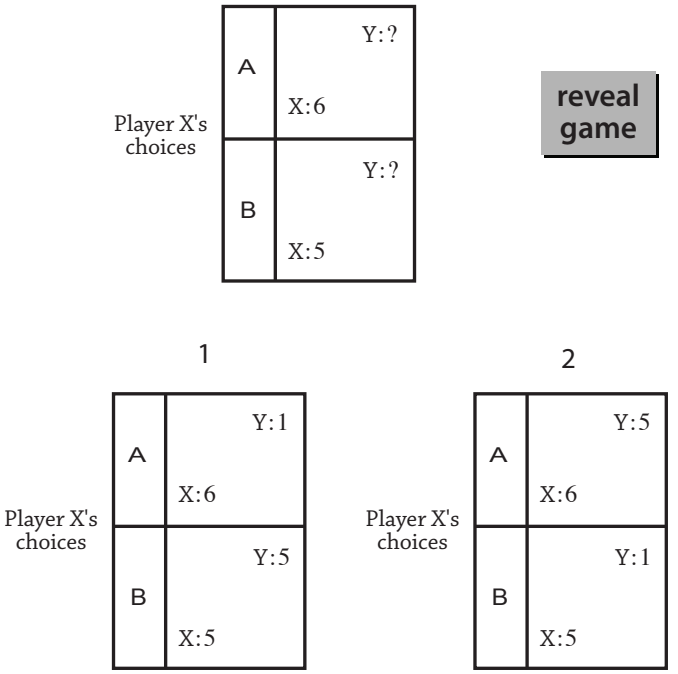

*Source:* Dana, Weber, and Kuang (2007, Figure 1: Interface for baseline treatment).

more moral than those who did not—even though they had ultimately acted just as egoistically as those who did not flip in assigning themselves the positive task. These results suggest that people can view their actions as moral by providing evidence to themselves that they are fair (through the deployment of a theoretically unbiased coin flip), even when they then ignore the outcome of that coin flip to benefit themselves. Follow-up research on children aged 6 to 11 suggests that this pattern of behavior has a developmental trend (Shaw et al. 2014). As children get older, they remain just as likely to assign themselves to the positive task: what changes with age is their likelihood of flipping the coin—that is, of attempting to gather evidence of their morality rather than actually behaving morally. Hence, consistent with motivated Bayesian reasoning, the act of flipping the coin—perhaps enough times to produce a favorable outcome—seems to provide sufficient evidence to decision makers that their egoistic behavior is in fact consistent with moral behavior.

As another example, consider the decision illustrated in Figure 1, drawn from Dana, Weber, and Kuang (2007). Choice *A* can be viewed as the egoistic act, because it gives the decision maker (*X*) \$6 instead of \$5 but the other person (*Y*) \$1 instead of \$5. Choice *B* arguably incorporates other considerations—such as equality and total welfare—that can be interpreted as moral. When confronted with this decision, 74 percent of participants in an experiment using monetary incentives selected the latter option, essentially giving up \$1 in order to act in concordance with some apparent moral consideration.

But consider the seemingly similar decision in Figure 2, also from Dana, Weber, and Kuang (2007). In this case, the decision maker again faces a choice between options that offer more (*A*) or less (*B*) money. But, now, the consequences for the other party are unknown, as reflected by the "?" symbol representing unknown payoffs. There are two possible—and equally likely—states of the world,

*Figure 2*
**Hidden Information Game from Dana, Weber and Kuang (2007)**




*Source:* Dana, Weber, and Kuang (2007, Figure 2: Interface for hidden information treatment).

depicted at the bottom of the figure, each with a different set of payoffs that might result from the decision maker's actions. In the first case, the payoffs are identical to those in the baseline case in Figure 1—choosing option *A* rewards the decision maker but harms the other party. But in the other case, acting egoistically also benefits the other person and yields the highest total earnings. That is, in the second case being egoistic is also being moral. Importantly, all decision makers had to do was to click a button ("reveal game") in order to find out the true payoffs. In roughly 50 percent of cases, actual payoffs were identical to those in the baseline, and morally motivated individuals in these cases could have easily acquired the information necessary to sacrifice self-interest for the sake of the greater good.

Despite the preponderance of individuals being willing to choose a more equal distribution with larger social payoffs in the baseline experiment, only 37 percent of participants did so in the hidden-information case in which the payoffs were identical to those in the baseline. That is, even though the resulting outcomes and the ability of individuals to implement those outcomes were identical, simply starting people off in a state of ignorance about the consequences of their actions diminished, by half, the frequency with which people sacrificed personal wealth

in pursuit of a moral objective. Moreover, roughly half of the decision makers did not bother to click the button and acquire the payoff information. In this context, decision makers appear to treat ignorance—even if it is a self-imposed absence of evidence that could easily be eliminated—as an excuse for acting egoistically. As motivated Bayesians, participants treat an action taken under willful ignorance as less indicative of an underlying egoistic motivation. This general result was subsequently replicated in other studies (Larson and Capra 2009; Matthey and Regner 2011; Grossman and van der Weele 2013; Feiler 2014).

Importantly in the Dana, Weber, and Kuang (2007) example, people cannot simply convince themselves that choosing ($6, $1) over ($5, $5) is the "moral" thing to do—otherwise, most people would do it in the baseline situation. Instead, they require some "wiggle-room" to reach the desired conclusion, provided by an informational default that allows the perception that choosing ($6, $?) over ($5, $?)—even when in a state of self-imposed ignorance—is not that bad. This is possible when they have the ability to interpret their own behavior favorably, in the manner of a motivated Bayesian.[3]

The research we review next provides further insights into the processes by which people manipulate their perceptions—of what is fair, of the likely outcomes of random processes, of perceived quality, and even of their own abilities—when doing so allows them to maintain a positive moral image while also garnering more personally desirable outcomes.

## Self-Serving Judgments of Morality

One way in which people engage in motivated evaluations of the morality of their own behavior is through a flexible construction of beliefs regarding what is "moral." For example, as we describe below, people may be more psychologically comfortable stating something that is not factually true when it is more likely that it *could* have been true (Schweitzer and Hsee 2002; Shalvi, Dana, Handgraaf, and De Drue 2011), or when a lie also benefits someone else (Wiltermuth 2011; Gino, Ayal, and Ariely 2013). A motivated Bayesian can interpret the moral implications of a lie self-servingly, thereby making it easier to act dishonestly. Motivated judgments can also influence perceptions of what is "fair" or "just." In many contexts, it is not straightforward to conclude how much one person deserves relative to another. In such cases, people often interpret the evidence regarding fairness and justice in self-serving ways—by evaluating what is moral through the lens of what also happens to be most personally rewarding. Our goal in this section is to show that judgments of what, precisely, is moral often possess some flexibility and that a motivated Bayesian

[3] Such "self-signaling" can be captured by models in which individuals do not have complete access to their underlying moral motivations when making moral judgments about their own behavior. Instead they draw inferences about their own motivations through actions they observe themselves taking (Bénabou and Tirole 2011; Grossman and van der Weele forthcoming).

can rely on such flexibility to pursue egoistic objectives while maintaining the feeling of adherence to moral standards.

The notion that people are self-serving in how they form judgments of justice is nicely illustrated in experiments on pre-trial bargaining (Babcock, Loewenstein, Issacharoff, and Camerer 1995; Babcock and Loewenstein 1997). In one study, law students were given a civil tort case and assigned to litigate one side of the case. After reviewing the case information, they provided estimates of the award actually granted by a judge in the case, with monetary incentives for accuracy. They also provided assessments of what would constitute a fair settlement. The judgments differed dramatically—by about 50 percent of the average settlement amount—between those "lawyers" assigned to argue the plaintiff's case versus those assigned to represent the defense. Importantly for our argument that people are motivated Bayesians, the difference was much smaller when people reviewed the case material *before* finding out which side of the case they would represent. That is, being forced to process the evidence and develop initial judgments of what constitutes a fair and unbiased settlement *before* having an incentive to view certain outcomes as more or less fair, subsequently prevented participants from having the flexibility to interpret the evidence as supporting a personally favorable notion of justice.

Other studies show that when people can choose among different standards of fairness, very little information is needed for them to favor the standards that are personally beneficial. For example, consider a situation where two people are working on a joint project, but one person's work has produced $20 and another has produced $10. Now suppose that one person decides unilaterally how to divide the total $30 earned by the pair. One could divide the total $30 either with an equitable division rule ($15, $15) or with a meritocratic one that allocates rewards according to account inputs ($20, $10). Either has some justification as a "moral" or "just" way to divide jointly produced rewards. In several studies, many people appear to identify the fair distribution of rewards in such cases as the one that best suits their financial interests (Frohlich, Oppenheimer, and Kurki 2004; Messick and Sentis 1979; Konow 2000).

As a concrete example of how motivated Bayesians construct self-serving judgments of what is just or fair, Rodriguez-Lara and Moreno-Garrido (2012) had pairs of participants answer quiz questions, which yielded a shared reward based on the number of correct answers provided by the pair. Importantly, the productivity of individuals' answers, in terms of how much they contributed to the reward, varied across participants. For example, in one variant of the experiment, one person generated 150 pesetas for each correct answer, while the other generated 200 pesetas. One participant in each pair was then randomly given discretion over how to allocate the combined "earnings" produced by the pair. Rodriguez-Lara and Moreno-Garrido identified three possible allocation rules that such an allocator might employ based on different judgments of what is "fair." Under an "egalitarian" rule, the proposer and the allocator receive the same amount of money, independent of their individual productivity. Under an "accountability" rule, participants are accountable for what they can actually control, which in this case is the number of correct answers,

but not accountable for the randomly determined productivity per answer. Hence, under accountability, participants receive money in proportion to the number of their correct answers. Finally, under a "libertarian" rule, participants receive an allocation equal to the money that they generated on the quiz based on their correct answers and their random productivity. The results provide clear evidence of motivated Bayesian reasoning. When the allocator's productivity was lower than that of the recipient, allocators relied more on the accountability rule and less on the libertarian rule—that is, they were more likely to allocate according to a rule that rewarded correct answers but not the random productivity shocks. However, when allocators were randomly assigned to be the ones whose output generated more revenue, the importance of accountablity and libertarianism was reversed—allocators were more likely to incorporate these random shocks as part of the entitlements in a just reward. Importantly, only 10 percent of the participants in Rodriguez-Lara and Moreno-Garrido's study kept everything for themselves—doing so feels clearly unjust and immoral. But while perhaps acting somewhat more "morally," the remaining 90 percent tended to form self-serving judgments of fairness consistent with motivated Bayesian reasoning. As Konow (2000) shows, such self-serving judgments of fairness can even constrain one's judgments of what is fair when subsequently dividing money among others as a disinterested third party.

Motivated Bayesians can similarly convince themselves that their actions are more moral than purely egotistical behavior when the situation gives them license to do so, even when the resulting outcomes are the same as those obtained through egotistical acts. This is the case in the study, discussed above, by Dana, Weber, and Kuang (2007): people seem to be more comfortable implementing unequal and inefficient outcomes when they can do so under a veil of self-imposed ignorance.

Another way motivated Bayesians can perceive the same egoistic act as more moral is by acting through an intermediary, which seems to diminish perceptions of moral responsibility. In a study by Hamman, Lowenstein, and Weber (2010), participants could either act egoistically, at the expense of another, by making decisions themselves or by selecting someone to make such decisions on their behalf. In one experimental treatment, participants decided unilaterally how to divide $10 with an anonymous and passive recipient—in a repeated version of the well-known "dictator game." In another treatment, participants hired "agents" to make the allocation decisions on their behalf. Importantly, the subject doing the hiring had all the market power, so agents had to compete for employment by trying to implement the level of sharing that those participants desired. When participants made allocation decisions themselves, a slight majority (51 percent) shared a positive sum with the recipient. However, when acting through intermediaries, this proportion declined to 13 percent—driven by the fact that participants sought out those agents willing to share the least on their behalf. Moreover, when asked to evaluate their behavior, decision makers who acted through agents felt less responsible for the unfair outcomes they had produced and perceived them as fairer. Hence, simply being able to hand off their "dirty work" to someone else can make people evaluate their pursuit of egoistic motives as less wrong. Once again, slightly different paths

to the same egoistic outcome can seem more moral when accompanied by a superficial justification. Other studies involving intermediaries that reveal conceptually similar results include Drugov, Hamman, and Serra (2013) and Erat (2013).

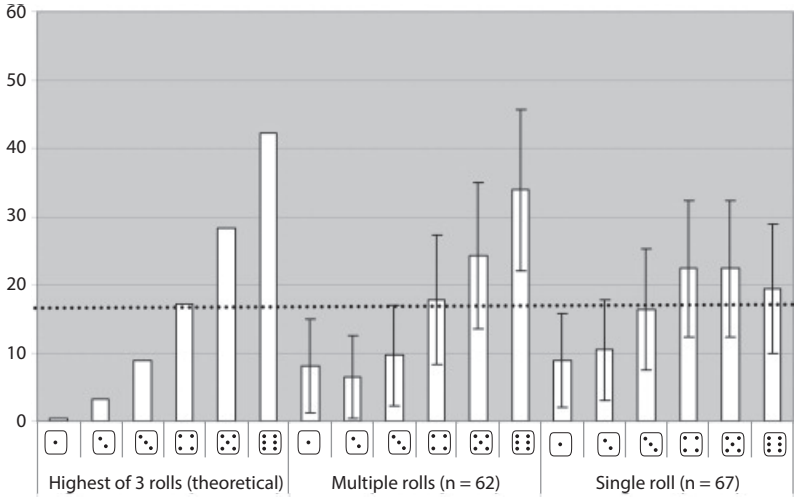
This ability to interpret evidence in a manner favorable to both one's egoism and perceptions of one's morality can be found in contexts beyond those involving sharing and distributing wealth. Many investigations of dishonesty, often led by psychologists interested in morality and behavioral ethics, provide evidence that slightly different paths of behavior toward the same egoistic end can provide individuals with flexibility to favorably interpret the morality of their behavior and the actions that they ultimately take (for examples, see Mazar, Amir, and Ariely 2008; Gino, Norton, and Ariely 2010; Shalvi, Gino, Barkan, and Ayal 2015).

As one example, a study by Shalvi et al. (2011) gave participants the opportunity to lie—by misreporting the outcome of a die roll—in order to obtain more money: higher numbers meant higher payoffs. Hence, an individual could report an outcome of six to obtain the highest possible earnings, and not even the experimenter could identify whether that individual had actually rolled a six. Shalvi et al. either had people roll the die once and report that outcome or, in a "multiple rolls" condition, roll the die three times with the instruction to report only the first roll. Panel 1 of Figure 3 shows the theoretical distribution of reporting the highest of best-of-three rolls. Panel 2 shows the distribution of reported outcomes when participants rolled the die multiple times, while Panel 3 shows the distribution for those who only rolled the die once. People appear to lie more when they roll the die multiple times with instructions to report only the first roll than when they only roll it once. Critically, the distribution of reported die rolls in the multiple rolls case is similar to a best-of-three distribution, suggesting that having observed a favorable die-roll outcome among one of the rolls *that did not count* allowed people to feel more morally justified in reporting that roll as their outcome. That is, if an outcome "could have been true"—in that the individual observed it actually happen—then lying about it seems to provide less-clear evidence of immorality than simply concocting an outcome that was never observed. Rather than treating the counterfactuals as irrelevant, these participants, like motivated Bayesians, incorporate all die roll outcomes as relevant evidence if doing so allows them to win more money by reporting a higher score.

The above examples share a common feature of motivated Bayesian reasoning. The decision maker presumably starts with a belief about his or her own concerns for egoism and morality, and then decides whether to take an action that provides evidence of the strength of these two motives.[4] However, rather than processing this evidence in an unbiased manner, a motivated Bayesian uses the context surrounding the choice to bias the inference drawn from one's own actions. Whether because a motivated Bayesian "did not know" the consequences of actions through willful

[4]For examples of models in which decision makers' actions provide signals of underlying motivations, see Bénabou and Tirole (2006, 2011), Ariely and Norton (2008), and Grossman and van der Weele (2013).

*Figure 3*
**Distributions of Reported Die Rolls**




*Source:* Shalvi, Dana, Handgraaf, and De Dreu (2011).
*Note:* In an experiment by Shalvi et al. (2011), people were either asked to roll the die once and report that outcome or, in a "multiple rolls" condition, roll the die three times with the instruction to report only the first roll. Higher reported numbers meant higher payoffs. Panel 1 of Figure 3 shows the theoretical distribution of reporting the highest of best-of-three rolls. Panel 2 shows the distribution of reported outcomes when participants rolled the die multiple times, while Panel 3 shows the distribution for those who only rolled the die once

ignorance or because the person was reporting outcomes that "could have" been true, this person, despite acting egoistically, reaches self-serving conclusions that such acts do not reflect a lack of morality.

## Self-Serving Interpretations of Reality

A separate and distinct type of motivated Bayesian reasoning involves not changing one's interpretation of the evidence regarding what is fair/unfair or moral/immoral, but instead changing one's perception of the evidence itself in order to arrive at a more positive moral impression of one's behavior. Such self-serving information processing is common in people's evaluations of their own characteristics and abilities, even in contexts that do not involve tradeoffs between egoism and morality. Several studies document that people seek out and attend to information that reinforces the belief that they are better than others in domains such as intelligence and attractiveness, overweighting positive information and underweighting negative information (Mobius, Niederle, Niehaus, and Rosenblat 2011; Eil and Rao 2011). For instance, Quattrone and Tversky (1984) found that

people who were told that greater tolerance to immersing one's body in cold water indicated longer (or shorter) longevity subsequently increased (or decreased) the amount of time for which they endured such a task. Hence, a motivated Bayesian who wants to believe in personal longevity may manipulate the evidence that is used in this judgment.

In the domain of moral behavior, people also seem to manipulate beliefs about their own abilities, particularly when doing so makes cheating seem less bad. Consider this statement by disgraced cyclist Lance Armstrong, stripped of his Tour de France victories after a doping scandal: "When you win, you don't examine it very much, except to congratulate yourself. You easily, and wrongly, assume it has something to do with your rare qualities as a person" (Armstrong and Jenkins 2003). Evidence that people misconstrue information about their morally questionable actions to instead provide evidence of their competence is provided by Chance, Norton, Gino, and Ariely (2011). They conducted a series of studies using a paradigm in which participants earned money for answering questions on an IQ test. Some participants took a standard IQ test, while others took the same test but with the answers printed at the bottom—allowing them to "check their work." Not surprisingly, those with the answers at the bottom scored higher on the test and made more money. But the key finding for our purpose occurs when both groups were then shown a second test, which had no answers at the bottom, and were incentivized to predict their performance on that test. An unbiased individual who had used the visible answers on the first test to obtain a higher score would presumably recognize this fact and anticipate lower performance on the second test. However, a motivated Bayesian might instead ignore the presence of the answers—or any effect they may have had on performance on the first test—and instead attribute good performance to personal intelligence, or assume it is driven by what Armstrong called "rare qualities as a person." Consistent with the latter account, people's predictions showed that they disregarded the presence of the answers and instead predicted that they would continue to perform well on the second test, attributing success to their innate "genius" rather than to cheating. Moreover, because payment for performance on the second test was based in part on the accuracy of predictions, these overestimations of performance resulted in motivated Bayesians making less money than people who never had the answers and were not tempted to cheat. When forced to take multiple tests without answers—a process that provided a stream of accurate feedback about their true ability—people were slow to correct their inflated beliefs; but when given another opportunity to cheat and perform well, they were quick to regain their faith in their enhanced abilities (Chance, Gino, Norton, and Ariely 2015).

People also manipulate their beliefs about the likely outcomes of random processes when doing so facilitates egoistic behavior. For example, Haisley and Weber (2010) presented participants with two options. An "other-regarding" option yielded payoffs for the decision maker and for a passive recipient that were relatively equal, for example, \$2.00 and \$1.75, respectively. The "self-interested" option gave the decision maker more money (for example, \$3.00) and gave the recipient a lower payoff involving risk—for example, a lottery paying \$0.50 with $p = 0.5$ and \$0 with $p = 0.5$.

Hence, the self-interested choice was guaranteed to make the recipient worse off, but by how much depended on the outcome of the lottery. The key manipulation in the study was the nature of the lotteries. In a simple-risk condition, the lottery was an objective $p = 0.5$ lottery, where ten red and ten blue chips were placed in a bag and one was drawn at random, with participants free to choose the winning color for the recipient. In an "ambiguous" lottery condition, the composition of the bag was unknown—participants were told that some random combination of red and blue chips had been determined prior to the experiment. Hence, the ambiguous lottery was objectively identical to the lottery involving known simple risk—in both cases there is a 0.5 probability of a ball of each color being selected—but its description created uncertainty about the precise color composition of the bag that would determine outcomes.[5]

The main hypothesis tested by Haisley and Weber (2010) was whether the vague nature of the ambiguous lottery would provide participants the flexibility to manipulate their beliefs about the likely outcome. That is, if participants can convince themselves that the ambiguous lottery is likely to yield a positive payoff with greater probability—since the probability could be anywhere between 0 and 1—then the self-regarding option appears less harmful for the recipient. Indeed, self-interested choices were selected in 73 percent of cases in the ambiguity condition, but only 59 percent of cases under simple risk. Here, the presence of ambiguous consequences for another seems to facilitate egoistic behavior.

Two pieces of evidence from the Haisley and Weber (2010) study particularly suggest a role for motivated Bayesian information processing. First, Haisley and Weber included another treatment dimension to examine whether first inducing participants to express their natural attitudes toward ambiguity, which are typically negative, would subsequently limit their flexibility to interpret ambiguity favorably. In the "constrained" treatment condition, participants started the experiment by choosing which type of lottery they preferred for themselves: one involving simple risk or one involving ambiguity. Consistent with classic evidence of "ambiguity aversion," a large majority of participants preferred the lottery involving simple risk. Importantly, only *after* expressing these attitudes toward ambiguity, did these subjects perform the main choice task, in which they chose whether to take more money for themselves and give the recipient a lottery, which involved either simple risk or ambiguity. Unlike the "unconstrained" participants discussed above, "constrained" participants did not exhibit more frequent self-interested behavior under ambiguity (59 percent) than under simple risk (63 percent). These results show that people who have just expressed an unfavorable view of ambiguity then find it difficult to switch to a favorable view when it becomes convenient to do so.

A second piece of evidence comes from asking participants to estimate the expected value of the payoff to the recipient produced by their choices, with

[5] Having less information about the actual composition of the bag typically induces "ambiguity aversion," whereby the ambiguous lottery is perceived as less desirable (Fox and Tversky 1995; Sarin and Weber 1993).

incentives for accuracy. Participants in the experiment make four choices that potentially affected the payoffs for a recipient. This part of the experiment also included a group of participants who made hypothetical choices, which they knew had no real consequences, so there was no incentive to engage in belief manipulation. Each participant played the game four times, resulting in four choices. Haisley and Weber (2010) calculated the degree to which the different types of participants over- or underestimated the expected value for the recipient resulting from their choices. Figure 4 shows the average estimate bias, cumulative across four choices, for the different groups of participants. The greatest degree of overestimation (by $0.89 across four choices) was demonstrated by "unconstrained" participants making choices under ambiguity; in no other case does ambiguity produce significantly greater overestimation of the value of lotteries, relative to simple risk. Thus, the only group that seems to adopt a strongly favorable view of the likely consequences of their choices is the group that has both an incentive to do so and the flexibility to manipulate their beliefs (having not been recently constrained by stating which kind of lottery they would choose for themselves).

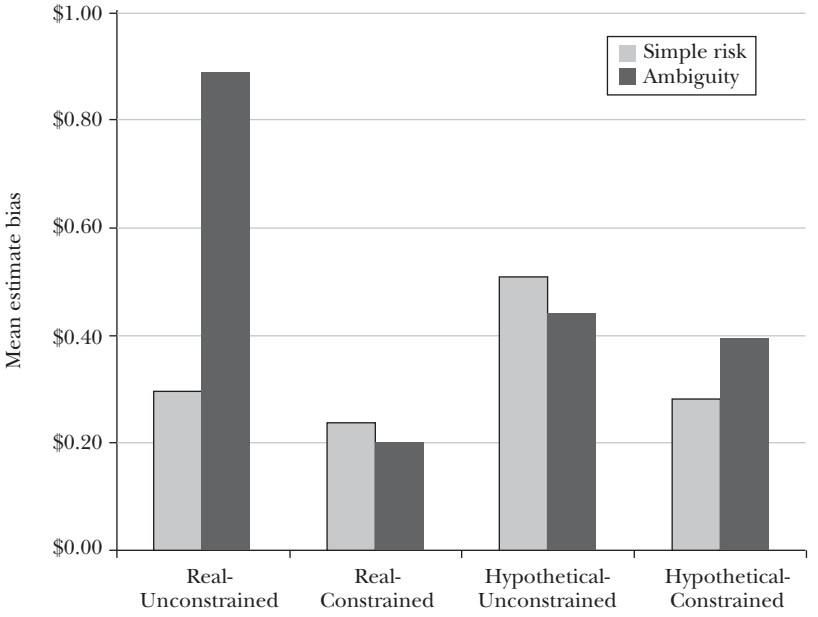
In the study by Haisley and Weber (2010), the choice confronting participants is one in which acting egoistically gives the other participant an unfavorable lottery. Hence, an individual sufficiently concerned with not prioritizing egoism over fairness may find it difficult to take such an action from a moral perspective. However, a convenient opportunity to satisfy both objectives arises if one can reinterpret the evidence to suggest that the unattractive lottery for the other party is, in fact, more attractive than it actually is.

Recent work provides additional evidence of motivated Bayesian reasoning in which people change their beliefs or preferences in order to facilitate egoistic acts. For example, one participant who may benefit by taking money from another may feel better about doing so when the first participant has some reason to feel convinced that the other intends to act unkindly as well (Di Tella, Perez-Truglia, Babino, and Sigman 2015). In the next section, we discuss some additional examples that are particularly relevant for policy questions of interest to economists.

## Why the Psychology of Self-Serving Moral Judgments Matters for Economists

As we have shown, self-serving judgments of morality and self-serving interpretations of reality are two common ways in which people act as motivated Bayesians. Much of the pioneering evidence of this phenomenon—and a large part of the existing knowledge—comes from laboratory experiments in psychology, where the idea that people are self-serving in information processing has long been of central interest (Hastorf and Cantril 1954; Festinger 1957). An important question is the extent to which motivated Bayesian reasoning is relevant for the domains that typically interest economists. Below, we discuss several economic contexts in which the kind of motivated reasoning we describe above likely plays an important role.

*Figure 4*
**Overestimation of Consequences for Another**




*Source:* Haisley and Weber (2010).
*Note:* This experiment involved choosing between two options: one yielding relatively egalitarian payoffs and another yielding more money for the decision maker, less for the other, and making the others' payoff the result of a lottery. The experiment varied whether the lottery involved simple risk (a known 0.5 probability) or ambiguity (a probability anywhere from 0 to 1). In the "constrained" treatment condition, participants started the experiment by choosing which type of lottery they preferred for themselves. The "unconstrained" treatment did not have this component. Some participants made hypothetical choices, which they knew had no real consequences, while others made real choices, which they knew would affect another person. The participants played the game four times, making four choices. The participants were asked to estimate the expected value for the recipient resulting from their choices, with incentives for accuracy. The figure shows the mean estimate bias, cumulative across four choices. See text for details.

**Charitable Giving**

A natural application of the insights on how motivated Bayesians confront tradeoffs between egoism and sharing wealth is to the domain of charitable giving, which constitutes both a sizeable portion of economic activity and an active area of economic research. Part of the interest among economists lies in understanding why people voluntarily donate to help others—a behavior potentially consistent with a moral motivation such as valuing the well-being of aid recipients or feeling pleasure from the act of giving (Andreoni 1990; Dunn, Aknin, and Norton 2008). However, if people prefer to act selfishly while at the same time believing that they are concerned with fairness and morality—and can employ motivated reasoning to satisfy both objectives—then we might observe them relying on excuses and justifications to avoid making costly charitable donations. Indeed, research suggests that avoiding charitable donation requests is easier for participants than declining the

requests once they are made and that, therefore, participants may go out of their way to avoid the request altogether (Flynn and Lake 2008; Lazear, Malmendier, and Weber 2012; DellaVigna, List, and Malmendier 2012; Andreoni, Rao, and Trachtman forthcoming). As with the research reviewed above on willful ignorance, such behavior is consistent with people having some flexibility in how they judge the morality of their actions—and choosing a course of action, when it is available, that yields less giving without a direct challenge to their moral standing.

People may also manipulate their beliefs about the attractiveness of a charitable donation—similarly to the phenomenon observed by Haisley and Weber (2010)—when doing so gives them justifications for acting egoistically. For instance, Exley (2016) examines people given the option to make a donation to a charity, but with some risk that the charity may not receive the donation—as when there is potential waste or corruption. Specifically, she compares situations involving a "self–charity tradeoff," in which people choose between a monetary allocation to be received personally or a monetary allocation to a charity where one of the two allocations involves risk, with other situations involving "no self–charity tradeoff," in which people choose between either a certain amount of money or a risky lottery for themselves, or a certain amount of money or a risky lottery for a charity. By varying the certain amount against which a risky lottery is compared, Exley can observe how much subjects appear to value risky lotteries for themselves or for a charity, and how this is influenced by the presence or absence of a self–charity tradeoff.

Figure 5 shows that when there is no tradeoff between egoism and helping the charity, in the left panel, people treat risk equivalently whether it affects their earnings or those of the charity—that is, they discount the "value" of a given amount of risky money similarly based on the probability that the money might not be received. However, when it comes to decisions involving a tradeoff between egoism and helping the charity, in the right panel, attitudes toward risk diverge considerably. In cases that involve, for example, a choice between keeping money for oneself or giving a risky lottery for the charity, choices reflect a much greater devaluation of lotteries involving risk for the charity than for oneself. In fact, in the right panel, for choices in which one can either give riskless money to the charity or allocate money to a risky lottery for oneself ("self lottery"), people appear to become risk-loving—overvaluing lotteries relative to their expected value—presumably because doing so creates the justification for keeping more money.

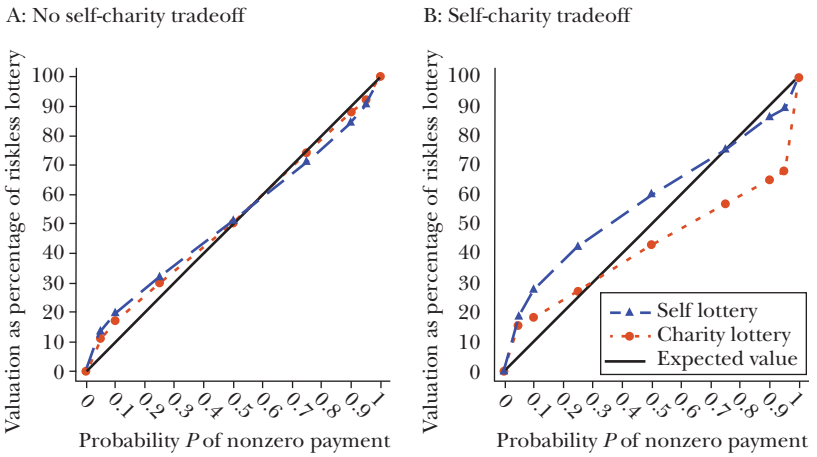
Hence, although participants' donation decisions reflect concern for the charity, when they can justify giving less by altering their attitudes toward risk to make donation relatively less attractive, they do so. Statements such as, "I would donate, but it would just go to waste" or "the charity's overhead is too high," may reflect motivated Bayesian information processing in action, coming up with justifications for not giving.

**Discrimination**

Another domain in which motivated Bayesians may find creative ways around doing the "right" thing is discrimination. If people are adept at altering the values that they subjectively place on seemingly objective criteria in order to justify ethically

*Figure 5*

**Valuation for Money to Oneself or to a Charity, Based on Risk and on Whether the Decision Involves a Tradeoff between Egoism and Helping the Charity**




*Source:* Exley (2016).

*Note:* Four situations are compared: 1) a certain monetary allocation or one involving risk, both for oneself (A: Self lottery); 2) a certain monetary allocation or one involving risk, both for a charity (A: Charity lottery); 3) a certain monetary allocation for a charity or one involving risk for oneself (B: Self lottery); and 4) a certain monetary allocation for oneself or one involving risk for a charity (B: Charity lottery). The experiment varies the certain amount against which a risky lottery is compared.

questionable preferences, this may allow them to reach the conclusion that a minority applicant for a position is worse on such "objective" criteria without believing that they themselves are actively discriminating or doing anything morally wrong.

Norton, Vandello, and Darley (2004) capture this alteration of decision criteria directly: men were asked to choose between male and female candidates for a stereotypically male job. Some participants read that the man was better educated but had less experience; others, that he had more experience but less education. Across both conditions, the majority of men selected the male applicant. Most relevant for our account, males claimed that gender played no role in their decision, instead citing education (but only when the male had more education) or experience (but only when the male had more experience) as the basis for their decision. Similar apparent manipulation of preferences is observed in a study by Snyder, Kleck, Strenta, and Mentzer (1979) in which participants chose which of two rooms to sit in to watch a movie. In one room, a person in a wheelchair was also waiting to watch the film; the other room was empty. There were two conditions: the film was either the same in both rooms (offering no excuse to avoid the disabled person) or different (offering a justification for choosing to watch the movie alone). Participants were more likely to choose to watch the movie alone when the two movies were different, presumably because this difference allowed them to claim that the movie in the "solo" room was objectively better—rather than admit to bias against sitting with the handicapped person.

By allowing motivated information processing to influence their perceptions of what constitutes an "attractive" option or candidate, individuals may find it easy to discriminate without believing they are doing so. Therefore, apparent and striking inconsistency between employers' claims that they do not engage in racial discrimination and their clearly race-based hiring decisions (Pager and Quillian 2005) may seem perfectly justifiable to the motivated Bayesian engaged in such discrimination.[6]

Such motivated Bayesian information processing may also provide an explanation for the finding that the returns to qualifications are lower for employment applicants from minority groups against which there is discrimination and that this can be partially explained by how prospective employers search for information on applicants (Bertrand and Mullainathan 2004; Bartoš, Bauer, Chytilová, and Matejka forthcoming). A motivated Bayesian employer who wants to discriminate, but feels wrong doing so blatantly, may search for reasons to favor a nonminority candidate over one from a minority group. Indeed, in interviews with 55 hiring managers, Pager and Karafin (2009) show that although managers held strong beliefs about the relative performance of black and white employees, they were often unable to generate any instances in their experience to support those impressions, suggesting that rather than updating beliefs with an unbiased evaluation of new evidence, as a classic Bayesian would, these managers were selectively weighting and interpreting information that supported their biased views.

**Corruption and Bribery**

Situations in which individuals are tempted to accept a bribe or favor a family member for a lucrative appointment also create the ideal conditions for motivated information processing (Hsee 1996). A motivated Bayesian may be quite adept at reaching a conclusion that the familiar candidate is the best qualified or that the vendor offering the highest bribe also offers the best use of public funds. Hence, an official awarding a prestigious sports tournament to a country that has also offered a lucrative personal payment may be able to convince himself that the country is really the most deserving based on "objective" criteria.

The application of this kind of reasoning to corruption is demonstrated by Gneezy, Saccardo, Serra-Garcia, and van Veldhuizen (2016). They use a task in which two participants compete over who can write the best joke (about economists), with the winner receiving a $10 prize. The prize is awarded according to the judgment of a third participant "referee" who picks the winner. The two competing participants can attempt to bribe the referee by sending part of the show-up fee that they receive in cash at the beginning of the experiment to the referee. In a "Before" condition, the referee receives any bribe in the same envelope as the written joke. Therefore,

[6] Such motivated Bayesian "nondiscrimination" can also occur in charitable donations. Fong and Luttmer (2011) find that varying the perceived race of the recipient of a charitable donation does not affect giving directly. However, nonblack donors who are led to believe that recipients are more likely to be black evaluate those recipients as less worthy of aid—for example, by choosing to believe the recipients are more likely responsible for their poverty—and, in turn, give less.

the referee observes the bribe at the same time as opening the envelope to read the jokes. In an "After" condition, the bribes and the jokes are in separate envelopes and the referee sees the bribes only after first reading the jokes. Note that these two versions change very little in terms of the tradeoff between morality and egoism. Someone who wishes to ignore the bribe and simply go with the best joke can do so in either case, which is also true for someone who wishes to simply select the egoistic option and ignore the quality of the jokes. However, a motivated Bayesian's judgments of the quality of the jokes may be swayed by which one is accompanied by the greatest personal benefit. At the same time, a motivated Bayesian who has already read the jokes and formed beliefs about their quality, before learning of the bribes, should find it harder to retroactively convince herself that the joke with the higher bribe is "better."

Consistent with motivated Bayesian reasoning, the timing of knowledge of the bribe appears to affect participants' willingness to be swayed by it. Eighty-four percent of participants in the "Before" condition selected the joke accompanied by the larger bribe, even though only 56 percent of these jokes were rated better by evaluators with no incentive. However, learning of the bribes only after reading the jokes constrains referees' judgments of joke quality: In the "After" condition, a lower proportion (73 percent) selected the joke accompanied by the larger bribe, and a much higher proportion selected the joke rated objectively better (81 percent). Hence, people are unsurprisingly swayed by bribes—but more so when they have the ability to interpret joke quality in a self-serving way.

**Attitudes Toward Market Outcomes**

Wealthier people often hold less-favorable attitudes toward redistribution (Alesina and Giuliano 2011). For example, Di Tella, Galiani, and Schargrodsky (2007) found that squatters in settlements in Argentina who were exogenously assigned property rights subsequently changed their perceptions of the inherent justice of a free-market system. In particular, these "lucky" individuals were more willing to support statements endorsing the belief that success results from hard work and that money is valuable for happiness. The correlation between personal circumstances and beliefs about the morality of the free-market system and potential resulting inequality might simply reflect self-interest: people express support for those policies that they believe to be most personally rewarding. However, motivated reasoning offers an alternative interpretation. Specifically, if motivated Bayesians can process information in a manner that allows them to reach the conclusion that what is personally rewarding is also that which is moral, then the above relationship may arise without people believing that they are compromising their morality. Instead, they may convince themselves—based on the information to which they attend and that they deem important—that the appropriate notions of fairness and justice are those that also happen to correspond to their own self-interest.

Relatedly, notions of what constitutes fair market wages may reflect self-serving biases and motivated information processing (Babcock, Wang, and Loewenstein

1996). For example, in a study by Paharia, Kassam, Greene, and Bazerman (2009) participants reported being relatively unwilling to hire a domestic worker to clean their house at a below-poverty level wage even when the worker was willing to accept this wage. When the decision was framed as hiring the worker through a placement agency ("Super Cleaners"), however, participants were far more likely to hire the worker. As in the study by Hamman, Loewenstein, and Weber (2010) that we reviewed earlier, inserting a third-party intermediary offers a degree of moral cover for what constitutes a "fair" wage.

Similar self-serving justifications may influence *consumers*' desire for products that raise ethical questions, such as those produced with sweatshop labor or those that may harm the environment. When presented with undesirable products produced with sweatshop labor, participants reported being uninterested in purchasing such unethical products; when products were desirable, on the other hand, purchase interest increased hand in hand with justifications for that increased interest, evidenced by greater agreement with sentiments such as "sweatshops are the only realistic source of income for workers in poorer countries" (Paharia, Vohs, and Deshpandé 2013). Moreover, Ehrich and Irwin (2005) show that people who care about a particular issue—such as the environment—are often *less* likely to seek out product information on that attribute. Because learning about negative environmental impact would constrain purchase, motivated Bayesian consumers avoid the chance of learning in order to allow them to feel good about purchasing behavior. These experiments again show that people motivated by egoistic concerns demonstrate remarkable celerity in using and misusing information to meet self-serving goals while continuing to feel moral.

## Conclusion

Economists have developed extensive literatures on topics related to the trade-offs people make between self-interest and moral considerations such as equality, social welfare, and honesty (Hoffman, McCabe, and Smith 1996; Charness and Rabin 2002; Frey and Meier 2004; Gneezy 2005; Fischbacher and Föllmi-Heusi 2013; Abeler, Becker, and Falk 2014), and have devoted considerable attention to corruption and its potential influence on economic development (Shleifer 2004; Bertrand, Djankov, Hanna, and Mullainathan 2007; Olken 2007). These streams of research have advanced our understanding of both the characteristics of individuals likely to lead them to compromise morality in pursuit of personal gain and the conditions under which such behavior is most likely.

We argue that an underexplored element in much of this research is the frequent tendency of decision makers to engage in motivated information processing—acting as motivated Bayesians—thereby resolving the apparent tension between acting egoistically and acting morally. Individuals' flexibility and creativity in how they acquire, attend to, and process information may allow them to reach the desirable conclusion that they can be both moral and egoistic at the same time. The

extensive literature in psychology and growing literature in economics reviewed above provide compelling evidence that behavior in many domains with a moral component is often driven by such self-serving information processing, suggesting that incorporating the underlying psychology into economic models is a worthwhile endeavor for future investigation.

## References


**Abeler, Johannes, Anke Becker, and Armin Falk.** 2014. "Representative Evidence on Lying Costs." *Journal of Public Economics* 113(May): 96–104.

**Alesina, Alberto, and Paola Giuliano.** 2011. "Preferences for Redistribution." Chap. 4 in *Handbook of Social Economics*, vol. 1, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson. North Holland.

**Andreoni, James.** 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal* 100(401): 464–77.

**Andreoni, James, Justin M. Rao, and Hannah Trachtman.** Forthcoming. "Avoiding The Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving." *Journal of Political Economy*.

**Ariely, Dan, and Michael I. Norton.** 2008. "How Actions Create—Not Just Reveal—Preferences." *Trends in Cognitive Sciences* 12(1): 13–16.

**Armstrong, Lance, and Sally Jenkins.** 2003. "Every Second Counts." New York: Doubleday.

**Babcock, Linda, and George Loewenstein.** 1997. "Explaining Bargaining Impasse: The Role of Self-Serving Biases." *Journal of Economic Perspectives* 11(1): 109–26.

**Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin Camerer.** 1995. "Biased Judgments of Fairness in Bargaining." *American Economic Review* 85(5): 1337–43.

**Babcock, Linda, Xianghong Wang, and George Loewenstein.** 1996. "Choosing the Wrong Pond: Social Comparisons in Negotiations that Reflect a Self-Serving Bias." *Quarterly Journal of Economics* 111(1): 1–19.

**Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matejka.** Forthcoming. "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition." *American Economic Review*.

**Batson, Daniel C., Diane Kobrynowicz, Jessica L. Dinnerstein, Hannah C. Kampf, and Angela D. Wilson.** 1997. "In a Very Different Voice: Unmasking Moral Hypocrisy." *Journal of Personality and Social Psychology* 72(6): 1335–48.

**Bénabou, Roland, and Jean Tirole.** 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96(5): 1652–78.

**Bénabou, Roland, and Jean Tirole.** 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126(2): 805–855.

**Bertrand, Marianne, Simeon Djankov, Rema Hanna, and Sendhil Mullainathan.** 2007. "Obtaining a Driver's License in India: An Experimental Approach to Studying Corruption." *Quarterly Journal of Economics* 122(4): 1639–76.

**Bertrand, Marianne, and Sendhil Mullainathan.** 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94(4): 991–1013.

**Camerer, Colin, and Richard H. Thaler.** 2003. "In Honor of Matthew Rabin: Winner of the John Bates Clark Medal." *Journal of Economic Perspectives* 17(3): 159–76.

**Chance, Zoë, Michael I. Norton, Francesca Gino, and Dan Ariely.** 2011. "Temporal View of the Costs and Benefits of Self-Deception." *PNAS* 108(S3): 15655–59.

**Chance, Zoë, Francesca Gino, Michael I. Norton, and Dan Ariely.** 2015. "The Slow Decay and Quick Revival of Self-Deception." *Frontiers in Psychology*, vol. 6, Article 1075.

**Charness, Gary, and Matthew Rabin.** 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117(3): 817–69.

**Dana, Jason, Roberto A. Weber, and Jason Xi Kuang.** 2007. "Exploiting 'Moral Wiggle Room': Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33(1): 67–80.

**DellaVigna, Stefano, John A. List, and Ulrike**

**Malmendier.** 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127(1): 1–56.

**Di Tella, Rafael, Sebastian Galiani, and Ernesto Schargrodsky.** 2007. "The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters." *Quarterly Journal of Economics* 122(1): 209–41.

**Di Tella, Rafael, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman.** 2015. "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism." *American Economic Review* 105(11): 3416–42.

**Dunn, Elizabeth W., Lara B. Aknin, and Michael I. Norton.** 2008. "Spending Money on Others Promotes Happiness." *Science* 319(5870): 1687–88.

**Drugov, Mikhail, John Hamman, and Danila Serra.** 2013. "Intermediaries in Corruption: An Experiment." *Experimental Economics* 17(1): 78–99.

**Ehrich, Kristine R., and Julie R. Irwin.** 2005. "Willful Ignorance in the Request for Product Attribute Information." *Journal of Marketing Research* 42(3): 266–77.

**Eil, David, and Justin M. Rao.** 2011. "The Good News–Bad News Effect: Asymmetric Processing of Objective Information about Yourself." *American Economic Journal: Microeconomics* 3(2): 114–38.

**Erat, Sanjiv.** 2013. "Avoiding Lying: The Case of Delegated Deception." *Journal of Economic Behavior & Organization* 93(September): 273–78.

**Exley, Christine L.** 2016. "Excusing Selfishness in Charitable Giving: The Role of Risk." *Review of Economic Studies* 83(2): 587–628.

**Fehr, Ernst, and Klaus M. Schmidt.** 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114(3): 817–68.

**Fehr, Ernst, and Klaus M. Schmidt.** 2006. "The Economics of Fairness, Reciprocity and Altruism—Experimental Evidence and New Theories." Chap. 8 in *Handbook of the Economics of Giving, Altruism and Reciprocity*, vol. 1, edited by Serge-Christophe Kolm and Jean Mercier Ythier, 615–91. Elsevier.

**Feiler, Lauren.** 2014. "Testing Models of Information Avoidance with Binary Choice Dictator Games." *Journal of Economic Psychology* 45 (December): 253–67.

**Festinger, Leon.** 1957. *A Theory of Cognitive Dissonance*. Evanston, IL: Row & Peterson.

**Fischbacher, Urs, and Franziska Föllmi-Heusi.** 2013. "Lies in Disguise—An Experimental Study on Cheating." *Journal of the European Economic Association* 11(3): 525–47.

**Fleming, Gerald.** 1993. "Engineers of Death." *New York Times*, July 18, sec. E.

**Flynn, Francis J., and Vanessa K. B. Lake.** 2008. "'If You Need Help, Just Ask': Underestimating Compliance with Direct Requests for Help." *Journal of Personality and Social Psychology* 95(1): 128–43.

**Fong, Christina M., and Erzo F. P. Luttmer.** 2011. "Do Fairness and Race Matter in Generosity? Evidence from a Nationally Representative Charity Experiment." *Journal of Public Economics*, Charitable Giving and Fundraising Special Issue 95(5–6): 372–94.

**Fox, Craig R., and Amos Tversky.** 1995. "Ambiguity Aversion and Comparative Ignorance." *Quarterly Journal of Economics* 110(3): 585–603.

**Frey, Bruno S., and Stephan Meier.** 2004. "Social Comparisons and Pro-Social Behavior: Testing 'Conditional Cooperation' in a Field Experiment." *American Economic Review* 94(5): 1717–22.

**Frohlich, Norman, Joe Oppenheimer, and Anja Kurki.** 2004. "Modeling Other-Regarding Preferences and an Experimental Test." *Public Choice* 119(1–2): 91–117.

**Gibson, Rajna, Carmen Tanner, and Alexander F. Wagner.** 2013. "Preferences for Truthfulness: Heterogeneity among and within Individuals." *American Economic Review* 103(1): 532–48.

**Gino, Francesca, Shahar Ayal, and Dan Ariely.** 2013. "Self-Serving Altruism? The Lure of Unethical Actions that Benefit Others." *Journal of Economic Behavior & Organization* 93(September): 285–92.

**Gino, Francesca, Michael I. Norton, and Dan Ariely.** 2010. "The Counterfeit Self: The Deceptive Costs of Faking It." *Psychological Science* 21(5): 712–20.

**Gneezy, Uri.** 2005. "Deception: The Role of Consequences." *American Economic Review* 95(1): 384–94.

**Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia, and Roel van Veldhuizen.** 2016. "Motivated Self-Deception, Identity and Unethical Behavior." Working paper.

**Grossman, Zachary, and Joël van der Weele.** 2013. "Self-Image and Strategic Ignorance in Moral Dilemmas." University of California at Santa Barbara, Economics Working Paper Series qt0bp6z29t. Department of Economics, UC Santa Barbara. https://ideas.repec.org/p/cdl/ucsbec/qt0bp6z29t.html.

**Haisley, Emily C., and Roberto A. Weber.** 2010. "Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior." *Games and Economic Behavior* 68(2): 614–25.

**Hamman, John, George Loewenstein, and Roberto A. Weber.** 2010. "Self-interest through Delegation: An Additional Rationale for the Principal–Agent Relationship." *American Economic Review* 100(4): 1826–46.

**Hastorf, Albert H., and Hadley Cantril.** 1954. "They Saw a Game; A Case Study." *Journal of Abnormal and Social Psychology* 49(1): 129–34.

**Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith.** 1996. "Social Distance and Other-Regarding Behavior in Dictator Games." *American Economic Review* 86(3): 653–60.

**Hsee, Christopher K.** 1996. "Elastic Justification: How Unjustifiable Factors Influence Judgments." *Organizational Behavior and Human Decision Processes* 66(1): 122–29.

**Konow, James.** 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American Economic Review* 90(4): 1072–91.

**Krupka, Erin L., and Roberto A. Weber.** 2013. "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association* 11(3): 495–524.

**Kunda, Ziva.** 1987. "Motivated Inference: Self-Serving Generation and Evaluation of Causal Theories." *Journal of Personality and Social Psychology* 53(4): 636–47.

**Kunda, Ziva.** 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108(3): 480–98.

**Larson, Tara, and C. Monica Capra.** 2009. "Exploiting Moral Wiggle Room: Illusory Preference for Fairness? A Comment." *Judgment and Decision Making* 4(6): 467–74.

**Lazear, Edward P., Ulrike Malmendier, and Roberto A. Weber.** 2012. "Sorting in Experiments with Application to Social Preferences." *American Economic Journal: Applied Economics* 4(1): 136–63.

**Matthey, Astrid, and Tobias Regner.** 2011. "Do I Really Want to Know? A Cognitive Dissonance-Based Explanation of Other-Regarding Behavior." *Games* 2(1): 114–35.

**Mazar, Nina, On Amir, and Dan Ariely.** 2008. "The Dishonesty of Honest People: A Theory of Self-Concept Maintenance." *Journal of Marketing Research* 45(6): 633–44.

**Messick, David M., and Keith P. Sentis.** 1979. "Fairness and Preference." *Journal of Experimental Social Psychology* 15(4): 418–34.

**Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat.** 2011. "Managing Self-Confidence: Theory and Experimental Evidence." NBER Working Paper 17014.

**Norton, Michael I., Joseph A. Vandello, and John M. Darley.** 2004. "Casuistry and Social Category Bias." *Journal of Personality and Social Psychology* 87(6): 817–31.

**Olken, Benjamin A.** 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(2): 200–249.

**Pager, Devah, and Diana Karafin.** 2009. "Bayesian Bigot? Statistical Discrimination, Stereotypes, and Employer Decision Making." *Annals of the American Academy of Political and Social Sciences* 621(January): 70–93.

**Pager, Devah, and Lincoln Quillian.** 2005. "Walking the Talk? What Employers Say versus What They Do." *American Sociological Review* 70(3): 335–80.

**Paharia, Neeru, Karim S. Kassam, Joshua D. Greene, and Max H. Bazerman.** 2009. "Dirty Work, Clean Hands: The Moral Psychology of Indirect Agency." *Organizational Behavior and Human Decision Processes* 109(2): 134–41.

**Paharia, Neeru, Kathleen D. Vohs, and Rohit Deshpandé.** 2013. "Sweatshop Labor Is Wrong Unless the Shoes are Cute: Cognition Can Both Help and Hurt Moral Motivated Reasoning." *Organizational Behavior and Human Decision Processes* 121(1): 81–88.

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

**Rabin, Matthew, and Joel L. Schrag.** 1999. "First Impressions Matter: A Model of Confirmatory Bias." *Quarterly Journal of Economics* 114(1): 37–82.

**Rodriguez-Lara, Ismael, and Luis Moreno-Garrido.** 2012. "Self-interest and Fairness: Self-serving Choices of Justice Principles." *Experimental Economics* 15(1): 158–75.

**Sarin, Rakesh K., and Martin Weber.** 1993. "Effects of Ambiguity in Market Experiments." *Management Science* 39(5): 602–15.

**Schweitzer, Maurice E., and Christopher K. Hsee.** 2002. "Stretching the Truth: Elastic Justification and Motivated Communication of Uncertain Information." *Journal of Risk and Uncertainty* 25(2): 185–201.

**Shalvi, Shaul, Jason Dana, Michel J. J. Handgraaf, and Carsten K. W. De Dreu.** 2011. "Justified Ethicality: Observing Desired Counterfactuals Modifies Ethical Perceptions and Behavior." *Organizational Behavior and Human Decision Processes* 115(2): 181–90.

**Shalvi, Shaul, Francesca Gino, Rachel Barkan, and Shahar Ayal.** 2015. "Self-serving Justifications: Doing Wrong and Feeling Moral." *Current Directions in Psychological Science* 24(2): 125–30.

**Shaw, Alex, Natalia Montinari, Marco Piovesan, Kristina R. Olson, Francesca Gino, and Michael I. Norton.** 2014. "Children Develop a Veil of Fairness." *Journal of Experimental Psychology: General* 143(1): 363–75.

**Shleifer, Andrei.** 2004. "Does Competition Destroy Ethical Behavior?" *American Economic Review* 94(2): 414–18.

**Skilling, Jeffrey K.** 2002 [2011]. "Jeff Skilling's Congressional Testimony." Testimony given February 7, 2002 to the Subcommittee on

Oversight and Investigations. Posted April 24, 2011 at *Enron Online: The Enron Blog*. http://enron-online.com/2011/04/24/jeff-skillings-congressional-testimony/.

**Snyder, Melvin L., Robert E. Kleck, Angelo Strenta, and Steven J. Mentzer.** 1979. "Avoidance of the Handicapped: An Attributional Ambiguity Analysis." *Journal of Personality and Social Psychology* 37(12): 2297–2306.

**Steele, Claude M.** 1988. "The Psychology of Self-Affirmation: Sustaining the Integrity of the Self." In *Advances in Experimental Social Psychology*, Vol. 21: *Social Psychological Studies of the Self: Perspectives and Programs*, 261–302. San Diego, CA, US: Academic Press.

**Weiner, Bernard.** 1985. "An Attributional Theory of Achievement Motivation and Emotion." *Psychological Review* 92(4): 548–73.

**Wiltermuth, Scott S.** 2011. "Cheating More When the Spoils Are Split." *Organizational Behavior and Human Decision Processes* 115(2): 157–68.