



# **Dishonesty in Everyday Life and its Policy Implications**

**Nina Mazar and Dan Ariely**

**Abstract:**

Dishonest acts are all too prevalent in day-to-day life. In the current review, we examine some possible psychological causes for such dishonesty that go beyond the standard economic considerations of probability and value of external payoffs. We propose a general model of dishonest behavior that includes also internal psychological reward mechanisms for honesty and dishonesty, and we point to the implications of this model in terms of curbing dishonesty.

**Keywords: performance-based incentives**

---

Nina Mazar is Post-Doctoral Associate and Lecturer in Marketing at the Sloan School of Management at MIT. Dan Ariely is Luiz Alvarez Renta Professor of Management Science at MIT and a visiting scholar at the Federal Reserve Bank of Boston's Research Center for Behavioral Economics and Decision-Making. Their email addresses are [ninam@mit.edu](mailto:ninam@mit.edu) and [ariely@mit.edu](mailto:ariely@mit.edu), respectively.

This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at <http://www.bos.frb.org/economic/wp/index.htm>.

The views expressed in this paper are those of the authors and do not necessarily represent the views of the Federal Reserve Bank of Boston or the Federal Reserve System.

**This version: January 2006**

## **DISHONESTY IN EVERYDAY LIFE**

Companies such as Enron, Tyco, WorldCom, and Adelphia are associated with some of the biggest financial scandals in corporate American history since the Great Depression. These corporations exemplified how the boom years in the 1990s were accompanied by a serious erosion in business principles, and especially ethics, in the accounting and auditing profession. The Brookings Institution estimated that the Enron and WorldCom scandals alone would cost the U.S. economy approximately \$37 to \$42 billion off Gross Domestic Product (GDP) in the first year alone (Graham, Litan, and Sukhtankar 2002). For comparison, this is approximately the amount that the U.S. federal government spends per year on homeland security (de Rugy 2005).

In addition to corporate scandals, almost all companies present their employees with the conflict between selfishly pursuing their own financial goals and being honest. Perhaps the clearest of these examples is brokerage companies. The fact that brokers are rewarded based on the volume or profitability of the business they place provides ample opportunities for conflicts of interest. Even though brokers are supposed to act in their clients' best interest, the commissions system can tempt brokers to choose personal gains over their clients' interests: they may pressure clients to buy and sell when the brokers stand to gain larger commissions, recommend stocks or funds that are suitable for the broker but not in the client's interest, delay the trades ordered by their clients to first invest their own money, or misuse knowledge of a large impending order (Davis 2004; McDonald 2002).

Companies and their employees are not alone in the playground of dishonesty (Murphy and Laczniak 1981). In addition to corporate scandals and individuals within companies who behave dishonestly, consumers also behave in ways that are ethically questionable (Bagozzi 1995; Vitell 2003). For example, in a recent survey conducted by Accenture about insurance fraud (2003), nearly 25 percent of U.S. adults approved of overstating the value of claims to insurance companies, and more than 10 percent indicated that submitting insurance claims for items that were not lost or damaged, or for treatments that were not provided, is acceptable. According to Accenture's press release, the Insurance Services Office estimates that the cost of fraud in the U.S. property and casualty industry is approximately 10 percent of total claims payments, or \$24 billion annually. Similar incidents of consumer fraud can be found in the retail industry. "Wardrobing," or the return of used clothing, was estimated to cost about \$16 billion in the U.S. in 2002, according to the National Retail Federation (Speights and Hilinski 2005).

Another domain that is central to consumers' unethical behavior has to do with intellectual property theft, such as music, film, and software piracy. While the standards and morals linked to such behaviors are not yet well established, it is clear that the economic implications of these endeavors are very large. As stated by the Office of the United States Trade Representative, intellectual property theft worldwide costs American companies at least \$250 billion a year – a staggering statistic considering that the copyright industries make up approximately 6 percent of the U.S. GDP (= \$626 billion)

and employ 4 percent of America's workforce (U.S. Department of Justice 2004). But perhaps the largest contribution to consumer dishonesty comes from tax deception, including such issues as omitting income and inflating deductions. A recent study by the IRS, based on special audits of randomly selected individual income tax returns for the 2001 tax year, estimates that the "tax gap" – the difference between what the IRS estimates taxpayers should pay and what they actually do pay – is somewhere between \$312 billion and \$353 billion annually. These numbers translate into an overall noncompliance rate of 15 percent to 16.6 percent (Herman 2005).

The examples mentioned above represent only a subset of everyday deception by companies, by individuals within companies, and by the individual consumers. Together, they contribute to the American economy's losing hundreds of billions of dollars in tax revenues, wages, and investment dollars, as well as hundreds of thousands of jobs a year.<sup>1</sup> As the damages to society's welfare become more apparent, substantial money is being invested on fierce introductions of special governmental task forces (for example, the Justice Department's task force on intellectual property), programs, and laws to fight dishonest behavior. However, little has proven successful so far. For example, even though the IRS is ramping up its audits on high-income taxpayers and corporations, focusing more attention on abusive shelters, and launching more criminal investigations, the overall tax gap has not changed much as a reaction to these measures. As IRS officials publicly state, too many audits result in no change in the amount of taxes paid. This means that these large and expensive efforts for increased compliance seem to be a huge waste of time and money for the taxpayer and the IRS (Herman 2005).

Similar disappointing stories can be reported, for example, from industry anti-piracy measures, such as investing in technologies for better copy-protection of CDs and DVDs, blocking of unauthorized downloads, improving consumer identifiability, or increasing the number of lawsuits. Evidence for the last was reported in a recent article on law.com, stating that music industry lawsuits against individuals have been ineffective in cutting peer-to-peer (P2P) music-swapping. Even though the industry measures seem to have contributed to an increased awareness of copyright laws and lawsuit-campaigns, individuals seem relatively unintimidated by them. An April 2004 survey revealed that 88 percent of children between the ages of 8 and 18 years understood that P2P music downloading is illegal, but despite this fact, 56 percent of these children admitted to continuing to download music (Lohmann 2004).

Given this limited success in curbing dishonest behavior, there are two possible approaches for understanding and limiting future dishonesty. The first approach assumes that the current strategy for curbing dishonesty is the correct one, but that it is not practiced sufficiently or with sufficient force. Under this approach, the two major variables – the probability of being caught and the magnitude of punishment – should be increased, and this would reduce or eliminate dishonesty. At the extreme, this would mean cutting off the right arm for minor crimes and amputating more important organs for more severe crimes. The second approach, which is the one we will take here, questions whether the typical path taken to reduce dishonesty (that is, increasing the probability of being caught and the magnitude of punishment) is the right path. In the

next section we will elaborate on these two approaches. The first approach will be under the general heading of the standard rational model, and the second will be under the heading of the psychological theory of honesty.

## **ECONOMIC AND PSYCHOLOGICAL THEORIES OF (DIS)HONESTY**

### **Economic Theories of the Effect of External Incentives**

The standard economic perspective of the *homo economicus* is one in which the individual is a rational, selfish human being, interested only in maximizing its own payoffs. This rational individual knows what it wants and what it does not want, and it is able to perform corresponding trade offs to select the option that is expected to deliver the greatest positive surplus (Hobbes 1968; Smith 1997; Smith 1999). From this perspective, there is nothing special about the decision to be honest. The decision to be dishonest depends solely on its expected external benefits (for example, getting money, getting a better position) and expected external costs (for example, paying a fine, losing a job) to the individual (Hechter 1990; Lewicki 1984): the higher the external rewards from being dishonest, the higher the extent to which an individual engages in dishonest behavior (see Figure 1). Such a cost-benefit tradeoff means that decisions about honesty are like every other decision that individuals face.

Because this view of dishonesty has been generally adopted, and, in particular, has been adopted in legal theorizing, efforts to limit dishonesty have generally assumed that the

only ways to make progress are restricted to influencing the external costs and benefits of a dishonest act. The ensuing emphasis on the pervasiveness of police force and magnitude of punishment are the two simplest ways to manipulate these external costs.

\*\*\* Figure 1 \*\*\*

### **Psychological Theories of the Effect of Internal Incentives**

*Internal rewards for virtuous behavior.* In contrast to this classic economic perspective, there exists ample evidence from different academic fields such as psychology, sociology, anthropology, and behavioral and experimental economics, as well as neuroeconomics and neuroscience, that in addition to the external reward mechanisms, there also exist internal reward mechanisms and that these exert influence on individuals' decisions. Economists such as Ernst Fehr (see for example, Fehr and Fischbacher 2003; Fehr and Fischbacher 2004; Fehr and Gächter 2002) and James Andreoni (see, for example, Andreoni, Harbaugh, and Vesterlund 2003; Andreoni and Miller 2005) have repeatedly demonstrated altruism and reciprocity in social dilemma games. For example, in an ultimatum game, two people have the opportunity to split a certain amount of money through a one-time-only and anonymous interaction. One of the two players, the proposer, offers a division of the money. The second player, the receiver, then has to decide whether she wants to accept or reject this proposition. If the receiver rejects the proposed division, both players go home without any money. If the receiver accepts the division, each player receives the amount that the proposer offered. From a game

theoretical, rational point of view, the proposer should split the money unequally, favoring herself. After all, it is a one-time game, so reciprocation is not an issue. In addition, offering at least 1 cent to the second player should make the receiver accept the offer, because that player will be better off with this amount relative to not getting anything. Instead, however, typically a majority of the offers are split equally, and many offers that are not split equally are rejected. Results of this nature demonstrate that individuals care about more than maximizing their own monetary payoffs and that these other considerations include social utility and the care for others' outcomes. A related recent work highlighting this point in the context of (dis)honesty has been conducted by Uri Gneezy (2005). In his paper, Gneezy uses a simple deception game with asymmetric information to demonstrate how individuals act selfishly in the sense of maximizing their own payoffs, yet are also sensitive to the costs their lies impose on others. These results are important when we consider how they play in daily deception since there are many differences in wealth (for example, between employees and employers and between consumers and corporations) and in the perception of the cost that deception creates for the other party. In essence, these results suggest that individuals will be more dishonest when they face wealthier counterparts and when the cost of the deception to these counterparts seems lower.

Thus, an important question – especially for economists – is, why do individuals consider more than “just” their material payoffs. To address this question, Joseph Henrich and colleagues (2001) undertook a large cross-cultural study of behavior in ultimatum, public good, and dictator games. They analyzed the behavior of individuals from 15 small-scale



societies in 12 countries exhibiting a wide variety of economic and cultural conditions. The sample consisted of foraging societies, societies that practice slash-and-burn horticulture, nomadic herding groups, and sedentary, small-scale agriculturalist societies. Two of their main findings supporting the hypothesis of internalized reward mechanisms are that: (1) the observed behavior varied substantially across societies, and (2) the individuals' preferences were not exogenous, as the classical model would predict, but rather shaped by their society's characteristic economic and social interactions of everyday life. In other words, socialization is key to the development of internalized reward mechanisms.

Adding to the behavioral evidence for internal rewards, recent findings from neuroscience and neuroeconomics provide further credence for the existence of internalized reward mechanisms, while also pointing to the brain structure that might be implicated in their activation. Thanks to the recent rapid diffusion of brain imaging studies like Positron Emission Tomography (PET) or event-related functional Magnetic Resonance Imaging (fMRI), there have been groundbreaking developments, adding more and more pieces to the puzzle of reward processing. Several studies in these fields identified certain regions in the dorsal and ventral striatum, especially the caudate nucleus and the nucleus accumbens, that represent a brain's pleasure center and that can be activated through very different forms of rewards (see also related studies in rats and primates; for example, Olds and Milner 1954; Schultz et al. 1992). Humans, for example, show significantly increased striatal activity during the anticipation of monetary gains (Knutson et al. 2001), pleasant tastes (O'Doherty et al. 2002), or beautiful faces (Aharon

et al. 2001). More important for our point, however, is the fact that the same brain regions are also activated in anticipation of satisfying social outcomes (that is, social rewards). Rilling and colleagues (Rilling et al. 2002) showed how the striatum “lit up” when individuals rewarded cooperators. Seemingly contrary to these findings, de Quervain and colleagues (de Quervain et al. 2004) reported similar findings when individuals punished defectors, and this result held even if the punishment was possible only at a personal material cost. Together, these latter two studies suggest that individuals feel good about complying with internalized social norms and values: Somebody who cooperates “*should*” be rewarded and somebody who defects “*should*” be punished in order to reestablish socially desirable behavior.

Psychology has also argued for a long time for internal reward mechanisms. Most notably, Sigmund Freud and disciples (see Freud 1933; Freud and Strachey 1962; Freud, Strachey, and Gay 1989) lectured extensively about the superego – a part of an individual’s self that represents a society’s moral norms and values that the individual has internalized in the course of its early life (see also Campbell 1964 on the internalization of moral norms). This superego acts as an internal judge, rewarding or punishing the individual depending on whether the individual complies with these norms and values.

*Activation of internal rewards lessens sensitivity to external incentives.* Based on the evidence of internal reward mechanisms described above, Nina Mazar, On Amir, and Dan Ariely (2005) studied how external and internal reward mechanisms work in concert in influencing an individual’s decision to be (dis)honest. The general

procedure was such that it contrasted incentives of financial rewards favoring dishonest behavior with the individual's inherent motivation for honest behavior. Taking the motivation for financial rewards as self-evident, the authors tested the extent to which the decision to be (dis)honest is driven by the possible consequences of being caught cheating (external costs) relative to the internal rewards of honesty. In one of their experiments, participants were presented with a set of 50 multiple-choice general knowledge questions and were promised 10 cents per answer that they solved correctly within 15 minutes.

There were four different between-subjects conditions that differed in the procedure once participants finished answering the questions on the original test sheets. The four conditions were: experimenter graded (control), self-graded (self), self-graded plus shredding (self+), and self-graded plus shredding plus self-payment (self++). In the control condition, participants were told that once they finished answering the questions, they should ask the experimenter for a bubble answer sheet (a sheet with circles to indicate the answers, much like the one used in exams that are scanned and graded electronically) on which to transfer their answers and then should take both forms to the experimenter, who would check their performance and pay them.

In the self condition, participants were told that once they finished answering the questions on the original test sheet, they should ask the experimenter for a pre-marked bubble answer sheet, which included the same circles as in the control condition, but with one of the answers pre-marked as being the correct answer. The respondents' task was to

transfer their own answers to the bubble answer sheet and then count and write down the number of correct answers. Once completed, they had to take both forms to the experimenter, at which point they would be paid accordingly. In this condition, subjects had the possibility of cheating at the risk that the experimenter might discover it.

The self+ condition was based on the self condition with the exception that participants were instructed that after transferring their answers onto the pre-marked answer sheet, they should walk to a shredder, shred their original test sheet, and take only the answer sheet to the experimenter, at which point they would be paid accordingly. This condition offered a lower probability of being caught cheating than the self condition. In fact, there was no legal way of proving that anyone cheated. The experimenter could only be suspicious about an atypically high number of correct answers.

Since this social interaction could make cheating participants feel uncomfortable, Mazar, Amir, and Ariely decreased the chance of being caught even further in the final condition (self++). In this condition, participants were instructed to shred both their test sheet and the answer sheet, to then walk over to a large jar with money that was sitting at the end of the table, and to take the amount that they earned.

This study had four important findings. First, participants in the external condition solved significantly fewer questions than the participants in the three self conditions. Since there was no reason to believe that the skill level of the participants was different in any of these conditions, this result implies that external reward mechanisms exist and that

individuals are dishonest when it is good for them. This result is also along the lines of Maurice Schweitzer and Christopher Hsee's findings (2002), according to which participants were more likely to be dishonest, the more ambiguous the private information they had.

Second, there was no significant difference between the three self conditions although the probability of being caught decreased dramatically from the self to the self+ to the self++ condition (and thus should have the fear of being caught). This result provides direct evidence that the external disincentives for dishonesty are only a part of a more complex picture.

Third, the magnitude of dishonesty in the three self conditions was relatively small. Participants cheated only 20 percent of the possible average magnitude and were, thus, far from exhibiting maximal dishonesty as the standard rational model would predict. This result suggests that individuals seem to possess internal reward mechanisms for honesty, because, when given the opportunity and incentives to be dishonest, individuals exhibited dishonest behavior, but this dishonest behavior was limited in its extent (limited probably by the strength of their internal reward mechanisms).

Fourth, these results suggest that the act of cheating itself can, at some level, activate the internal reward mechanism. This means that while low levels of dishonesty might go "unnoticed" from the point of view of the internal reward mechanism, at some point the

magnitude of dishonesty itself can activate this mechanism and limit one's dishonest behavior.

Together, these findings suggest that the relationship between the external and internal reward mechanisms is quite complex. In particular, the authors hypothesize that the internal reward mechanism is either active or inactive (in the same way that individuals largely categorize actions and individuals as honest or dishonest rather than in a more continuous way) and influences the tendency for acting (dis)honestly as a step function:

(1) Below a certain level of dishonesty the internal reward mechanism may not be activated at all and hence would not influence behavior – that is, the propensity for dishonesty in such cases will be a function of external cost-benefit considerations, as depicted in Figure 1. (2) Beyond the activation threshold, when the dishonest act is noticeable, the internal reward mechanism gets activated and exerts – at least within a certain range – its maximal force independently of the level of external rewards. As a consequence, once an individual's internal standards “kick in,” they override the sensitivity to external rewards, such that the propensity for dishonesty becomes independent of increased external rewards for being dishonest (within a certain range). (3) Arguably, it is likely that at some point, when the external rewards become very large, these rewards will tempt the individual and may ultimately prevail. That is, at some point one's internal standards could get overridden or de-activated such that the causes for dishonesty will preclude such internal considerations and be based solely on planned dishonesty (a cost-benefit analysis, as theorized by the standard rational model). Figure 2 sketches the hypothesized relationship of these components as illustrated by the findings

of Mazar, Amir, and Ariely. As suggested by these findings and the contrast between Figures 1 and 2, a model for honesty that assumes that people behave like a *homo economicus* would lead to very different predictions about an individual's decision to be (dis)honest and, to the extent that these assumptions are wrong, would lead to wrong prescriptions for the prevention of deception.

\*\*\* Figure 2 \*\*\*

*Changing the activation threshold for internal rewards.* It is important to note that all the research mentioned above supports the theory that internal reward mechanisms do exist. There also exists, however, research suggesting that it is possible to move the activation threshold, that is, to cause internal reward mechanisms for honesty to be more active, or to kick in earlier. Representative of this kind of evidence is the extensive body of research in psychology on objective self-awareness by Shelley Duval and Robert Wicklund (1972). According to Duval and Wicklund, objective self-awareness represents attention directed toward oneself that induces self-evaluation in relation to standards salient or accessible in the immediate situation, which in turn increases motivation to meet the standard (see also Carver and Scheier 1998). Particular situations, such as being in front of a real or implied audience (Duval and Wicklund 1972), being individualized, standing in front of a mirror (Carver and Scheier 1978), or writing short stories about oneself (Fenigstein and Levine 1984), can increase an individual's awareness of herself as an object in the world. Once the awareness of the self is increased, individuals are also more likely to be aware of discrepancies between how they want to view themselves (the

ideal self) and how they actually behave. Given this tension and how uncomfortable this tension can be, individuals who are aware of it might work actively to reduce the discrepancy, either by shifting attention away from the self or by changing their behavior in order to act more in accordance with their ideal self. In the domain of deception, this means that higher self-awareness might lead to more honest behavior.

One well-known example of the effect of self-focused attention on increasing the alignment between one's behavior and one's internal standards was provided by Arthur Beaman and colleagues (1979). In this famous experiment conducted during Halloween, trick-or-treating kids entered a house and were told by the experimenter to take only one candy, after which the experimenter left the children alone in front of the candies. They found that children who were individualized by being asked their names and addresses were more likely to take only one candy. In addition, these children were even more likely to take only one candy when confronted with a mirror directly behind the candy bowl. Mazar, Amir, and Ariely (2005) used a different manipulation to influence self-awareness – a manipulation that is more directly related to honesty – in order to test whether an increased awareness would lead to more honest behavior on a math test. In their experiment, participants first had to either write down as many of the 10 Commandments as they could remember (increased self-awareness regarding honesty) or the names of 10 books that they had read in high school (control). They had 2 minutes for this task before they moved on to an ostensibly separate task: the math test. The task in the math test was to search for number combinations that added up exactly to 10. There were 20 questions and the duration of the experiment was restricted to 5 minutes. After



the time was up, students were asked to recycle the test form they had worked on and indicate on a separate collection slip how many questions they had solved correctly. For each correctly solved question they were paid 50 cents. The results showed that students who were made to think about the 10 Commandments claimed to have solved fewer questions than the controls. Moreover, the reduction of dishonesty in this condition was such that the declared performance was indistinguishable from that of another group that had their responses checked by an external examiner. This suggests that the higher self-awareness in this case was so powerful that it eliminated dishonesty completely. Yet another method that proved successful in increasing self-awareness with the consequence that students cheated less on that math test was to make students sign an honor code before starting with the test. This finding is particularly interesting, since it provides some evidence for the effectiveness of a very simple commitment device that can easily be adopted by schools, companies, or the government.

A different kind of research that can be interpreted in line with the self-awareness theory is based on Daniel Schachter and colleagues' work on the role of emotional arousal for moral behavior. Daniel Schachter and Bibb Latane (1964), for example, tested the effects of a tranquilizer in a situation in which subjects could cheat. They found that subjects cheated significantly more after having taken the tranquilizer, a sympathetic inhibitor, when compared with the placebo controls (see also Schachter and Singer 1962). These results suggest that activation of the threshold is based on arousal, and that once arousal is decreased (even by artificial means such as a tranquilizer), the threshold is less likely to be activated and hence dishonesty is likely to be more prevalent. This idea was further

developed by Richard Dienstbier and Pamela Munter (1971), who showed that it is not emotional arousal per se that influences the tendency for cheating, but one's understanding or interpretation of the meaning and significance of that arousal (see also Dienstbier 1975; Dienstbier 1972).

*Self-deception.* Thus far, the discussion has focused on deceptive acts and the activation of a threshold that ignites the internal rewards. In our conceptualization of the threshold, an important part was that the amount of deception can, by itself, activate the threshold. An added complexity comes from the idea of self-deception, whereby individuals can reframe an act to themselves in a way that makes the act not be perceived as dishonest. Dishonest acts under such reframing (self-deceptions) would not contribute to the activation of the threshold and hence would not influence the tendency for honesty. As a consequence, even if an individual has internalized standards for honesty, acts of dishonesty that do not influence the activation of the threshold (self-deceptions) would not have any influence on behavior, and the relationship between external rewards and the tendency to dishonesty would remain in zone 1 of Figure 2 (never reaching the threshold).

In general, self-deception represents a biased, self-serving information flow within an individual – an active but unconscious misrepresentation of reality to one's own conscious mind (Trivers 2000). While it seems paradoxical that one could deceive oneself, casual observation suggests that people are quite effective in maintaining unrealistically positive views of themselves: Individuals maintain beliefs in their

intelligence, competence, and moral worth in the face of their sometimes foolish, incompetent, and immoral behavior. Similarly, individuals frequently estimate that they are, for example, better, more intelligent, or more beautiful than average. Since typically far more than 50 percent of individuals estimate this “better than average” effect, aggregating these data clearly shows that the collective result violates logic (see, for example, Alicke et al. 1995). Other researchers have shown how individuals can be led to believe false or biased facts about their own past (see research on memory distortion and suppression of unwanted memories, for example, Anderson, Cohen, and Taylor 2000; Loftus 1994; Schachter and Dodson 2002), or they can convince themselves of certain motivations for their behavior, thereby hiding their true intentions. Thus, self-deception can be successful even in the most extreme cases: For example, doctors who participated in genocide in Nazi Germany managed to convince themselves of the rectitude of their actions (Lifton 1986).

While deceiving oneself can be beneficial in the short run because it allows individuals to maintain self-regard, the inflated self-perception can be problematic in the long run when that bogus self-regard has costs. Mike Norton and Dan Ariely (2005) have shown this in the context of an IQ test. In one of their experiments, individuals participated in two IQ tests, one after another. Half of the individuals in the first test were provided with an answer key on the bottom of their test sheet, while the other half did not have an answer key. When the first test was over, each of the participants learned how many questions he or she had solved correctly. After that, participants were asked to predict how many questions they would solve on a second, similar test without an answer key. The

prediction task was designed such that individuals had a monetary incentive to be accurate. Payment was more dependent on the accuracy of their prediction and less on their performance, so it was a dominating strategy to predict accurately.

The results of this experiment showed that in the first round participants who had the answer key performed better on the test than participants without the answer key, and that there was no performance difference on the second IQ test when none of the participants was provided with an answer key. This result suggests that the presence of the answers allowed participants to solve more problems correctly. More importantly, it seems that those participants who had the answer key in the first test felt that this better performance was actually due to their greater intelligence, as reflected in their higher estimates of their performance on the second test. As a consequence, participants who deceived themselves by having an inflated perception of themselves made less money. Interestingly, when Norton and Ariely asked a different group of individuals to predict how they would perform in such an experiment, individuals predicted that if they had the answer key on the first test but not on the second, they would perform better on the first test, but not on the second test. This result suggests an added danger of self-deception – that individuals believe that they are immune to it.

Researchers have explained this paradox of not knowing about one's self-deception with the assumption that a person can simultaneously store both true and false knowledge, with a bias toward the true knowledge being stored in the unconscious and the false (that is, misrepresented) knowledge stored in the conscious (Greenwald 1997). Robert Trivers

(2000) pointed out that from an evolutionary standpoint this way of organizing knowledge has the advantage that self-deception must not only act primarily in the service of fooling oneself but can also act in the service of deceiving others. The latter is true since an outside observer gets in touch first and foremost with the conscious mind of the deceiver, and if the deceiver is also deceiving herself, the conscious mind will be aware of only the false information. The true information would be hidden in the unconscious. Cues like higher-pitched voice, increased pupil size, or lip-pressing that generally accompany attempted, conscious deception should then not be available to the observer, and hence the deceit would be harder to detect (DePaulo and Morris 2004). As a consequence, even if individuals are fully self-aware (that is, the internal reward mechanism is active) and the net utility of deception is negative (that is, the costs loom larger than the benefits),<sup>2</sup> deception that eludes the person committing the dishonest act – self-deception – might not completely vanish.

## **POLICY GUIDELINES FOR REDUCING DISHONESTY**

The standard rational model of decisions about honesty and dishonesty assumes that individuals trade off only external costs and benefits of an outcome. In contrast, the psychological model sketched here assumes that decisions about honesty also include considerations of internal reward mechanisms. In addition, we have argued that decisions about honesty can sometimes be not cognizant, for example, at levels of activation below the threshold, or when self-deception is involved. As a consequence, making the right policy recommendation to decrease dishonesty depends particularly on analysis of what

is driving the deceit in a particular situation. Our literature review suggests that there are four general drivers of dishonesty: (1) lower external costs and relatively higher benefits of deception, (2) lack of social norms, resulting in a weak internal reward mechanism, (3) lack of self-awareness that primes the activation of the internal reward mechanism, or (4) self-deception. In the next sections we elaborate on how these different perspectives on dishonesty would translate into different approaches for trying to reduce or curb dishonesty.

### **When Dishonest Behavior Is Caused by External Rewards**

If the cause of deception lies solely in greater external benefits than costs of the dishonest act, the solution is simple: One has to set the costs for dishonest actions such that they are larger than the expected benefits. This can be achieved either by increasing the probability of being caught or by increasing the severity of the punishment. Thus, if the cause of dishonesty is based solely on an imbalance of external costs and benefits, then the standard legal approach of controlling the external costs is appropriate. This theory implies that it is appropriate to introduce governmental task forces like the Department of Justice's task force on intellectual property, which, among other combat strategies, invests in increasing the number of specially trained prosecutors. The same is true for the IRS's increase in audits as well as for the music industry's aggressive increase in filing lawsuits against individual deceivers.

Even if the cause of deception has to do with the cost-benefit-analysis, there might be ways to increase the effectiveness and efficiency of measures to combat dishonest behavior. For example, if the probability of being caught and the magnitude of punishment are evaluated differently, and research suggests that the probability of punishment is more important than the severity of the punishment (Nagin and Pogarsky 2003), it might be best to allocate efforts accordingly. However, even if the legislator decides to invest more effort in the probability of detection and decrease the magnitude of punishment (for example, moving from a \$500 fine for not stopping at a stop sign with 10 percent probability of being caught to a \$100 fine with a 50 percent probability of being caught), there is still the question of what is the optimal probability for deterrence.

Informing the question of what is the optimal probability for deterrence, research by Rachel Barkan and colleagues (1998) suggests that the best approach is to eliminate altogether the probability of being caught, that is, to move to non-probabilistic punishments. Their main argument is that events that have low probability are unlikely to occur (by definition) and hence can have a perverse effect on learning: Individuals who violate the rule and are not caught get a positive reward for the violation that causes them to underestimate the probability of being caught and over time increases their tendency to behave in this undesirable way (see also Erev et al. 2003). According to this perspective, a person who expects that running a red light would involve a \$500 fine in 5 percent of the cases is more likely to go through a red light than a person who has the same expected value but with certainty (that is, \$25 for sure). More importantly, over time the person in the probabilistic punishment setting is going to further discount the probability of the

punishment (as long as she is not caught), which will lead to an even higher tendency for violation. Obviously, eliminating the probabilistic component from all undesirable behaviors is impossible, but there are clearly some cases (like crossing an intersection on a red light) where this is possible and desirable.

### **When Dishonest Behavior Is Caused by the Internal Reward Mechanism**

If the reason for dishonest actions lies in a lack of internalized social norms, then our primary recommendation would be to invest in educational efforts and socialization to increase the strength of the internal reward mechanism. The key questions in this context are then how this can best be done and whether there is a critical age period for the internalization of such mechanisms (as in language and visual development). Educational efforts can be integrated, for example, in schools, social clubs, or religious institutions. Another possibility that is increasingly exercised by the government as well as the music, film, and software industries, is to feature public messages in an attempt to build a social norm where a particular type of behavior (such as illegally downloading music or movies) is considered socially undesirable and is frowned upon. Other efforts may illustrate how such acts can hurt the rank-and-file workers, and not just the big corporations, by reducing their job security or pay (see, for example, TV ads and movie trailers launched by the Motion Picture Association of America – MPAA – in 2003 and 2004).



Once we understand the effects of such efforts on the development of socially based internal reward mechanisms, it is important to go a step further and ask ourselves what should be the limits of such efforts and whether society should allow all ideologies to participate in the creation of such internal rewards (for example, what about racial prejudices or particular cults?). The question about the types of internal reward mechanisms that society might want to develop or not develop is particularly important if we consider the likely possibility that this process might have a critical period in which younger individuals are much more sensitive to such influences and that once the critical age is reached, these mechanisms remain relatively stable for the rest of their lives. Given the higher sensitivity of younger adults to social influence and advertising, society might want to place careful boundaries on the development of such socially based internal representation by different interested parties such as religious, financial, and social institutions.

If it is not a lack of social norms but simply a lack of self-awareness and, hence, the degree to which these internalized social norms are activated, then it is important to make use of contextual cues that increase awareness when deception is about to happen: at the point of temptation. The IRS could, for example, make a slight change to its forms by making them more personal or asking people to sign something like an honor code before they start filling out the forms. Another possibility worth trying might be to include a survey asking taxpayers questions such as how much they care about their country, how important was honesty to their parents, how many people they think lie on their taxes, or what is the typical profile of tax deceivers.

The consideration of internal rewards also suggests that the theory of optimal punishment (optimal punishment trades off the benefits of deterrence and the cost of punishing innocent individuals) should be reconsidered with these inputs in mind. If the punishment magnitude is determined in a way that makes the costs slightly higher than the benefits, and if these costs also include internal costs, the optimal punishment will be lower by that amount. For example, if the expected benefit for a particular crime is  $Y$ , and the internal reward for honesty is  $X$ , the standard rational model would prescribe a punishment with an expected magnitude of  $-(Y+\epsilon)$ , while the model that includes internal rewards would prescribe  $-(Y+\epsilon)+X$ . The complexity, of course, is that not everyone has the same level of internal reward mechanisms, and to the degree that these are unobservable it is hard to assess the true level of optimal punishment (although it is possible that one day we will have a test for this). On the other hand, signs of repeated criminal behavior, for example, can be taken as an indication of a lower level of internal reward mechanisms, causing the magnitude of  $X$  to be updated as lower. This type of framework, in which  $X$  is an individual-specific variable, has the potential to help us build a theory of repeated punishment with the same desired principles of optimal punishment, but with more effectiveness (right now, repeated crime is punished more severely, as in California's "three strikes and you're out" approach, but there is no theory or logical guideline for the magnitude of these policies).

## **When Dishonest Behavior Is Caused by Self-Deception**

Finally, if the deception is of the self-deception variety, it is hard to fight. Self-deception due to a self-serving bias is very robust. Several academics have shown that neither paying people to be more realistic in their views nor making people write essays arguing the other side's point of view or educating individuals about the self-serving bias is successful in debiasing individuals (see, for example, Babcock and Loewenstein 1997; Babcock, Loewenstein, and Issacharoff 1997). The most successful way of fighting self-deception might therefore be to eliminate altogether the incentives that spawn the bias: Simply eliminate the situations that can give rise to this type of behavior. Max Bazerman, George Loewenstein, and Don Moore (2002; see also Bazerman and Loewenstein 2001), for example, argue that if deceptive audits by accounting professionals are mainly due to self-deception, it might be more effective to pass laws or enforce standards among the accounting profession that bar auditors from offering both consulting and tax services to clients, prohibit hiring of accountants through clients, and allow only contracts limited in time. As these examples show, fighting deception caused by self-deception requires very serious interventions, limiting substantially the freedom and self-determination of individuals in certain situations.

## **CONCLUSION**

In sum, there is no question that dishonesty is prevalent in daily life. The standard economics perspective considers one cause for dishonesty – external reward mechanisms

– and, as a consequence, emphasizes probability of being caught and magnitude of punishment as the only ways to overcome dishonesty. In contrast, the psychological perspective presented here suggests that dishonesty is also influenced by internal reward mechanisms and that these should be taken into account when considering effective measures for limiting dishonesty in daily life. Moreover, the psychological perspective suggests that internal and external rewards are not simply additive, but take a particular functional form (see Figure 2). With this functional form in mind, the psychological approach for reducing dishonesty could be based on increasing the long-run effectiveness of internal rewards (education), on increasing the short-run effectiveness of internal rewards (contextual cues), or on eliminating the possibility of dishonest acts in cases in which the cause could be attributed to self-deception. Once we understand better the role of internal rewards, both prevention and punishment of dishonesty could become more effective and efficient.

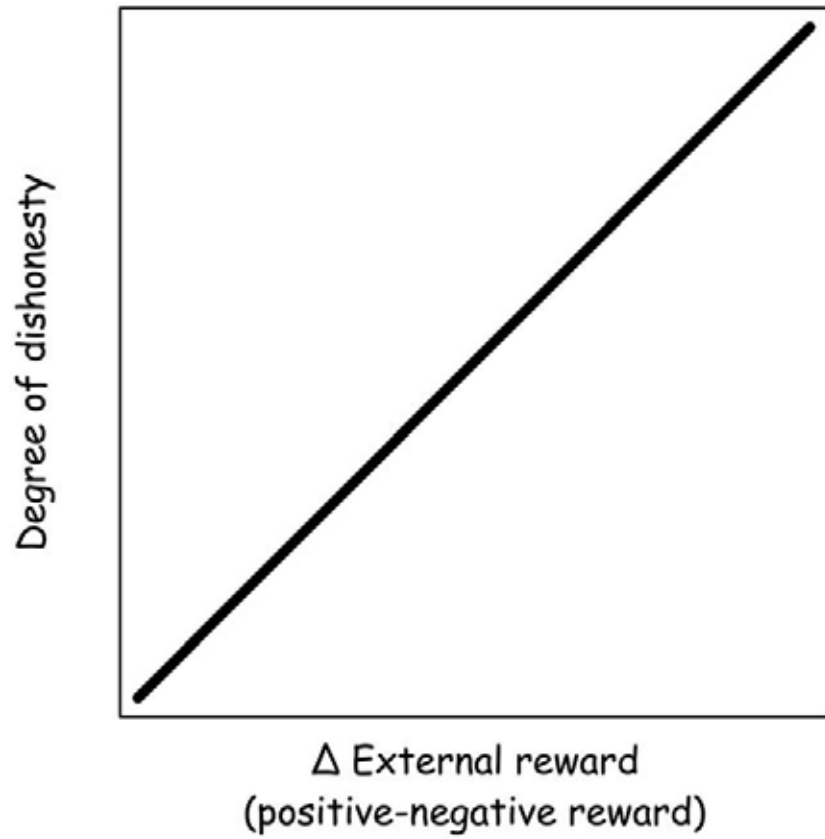
## Endnotes

<sup>1</sup> Of course not everybody loses in every dishonest act – those who committed this dishonest act can gain.

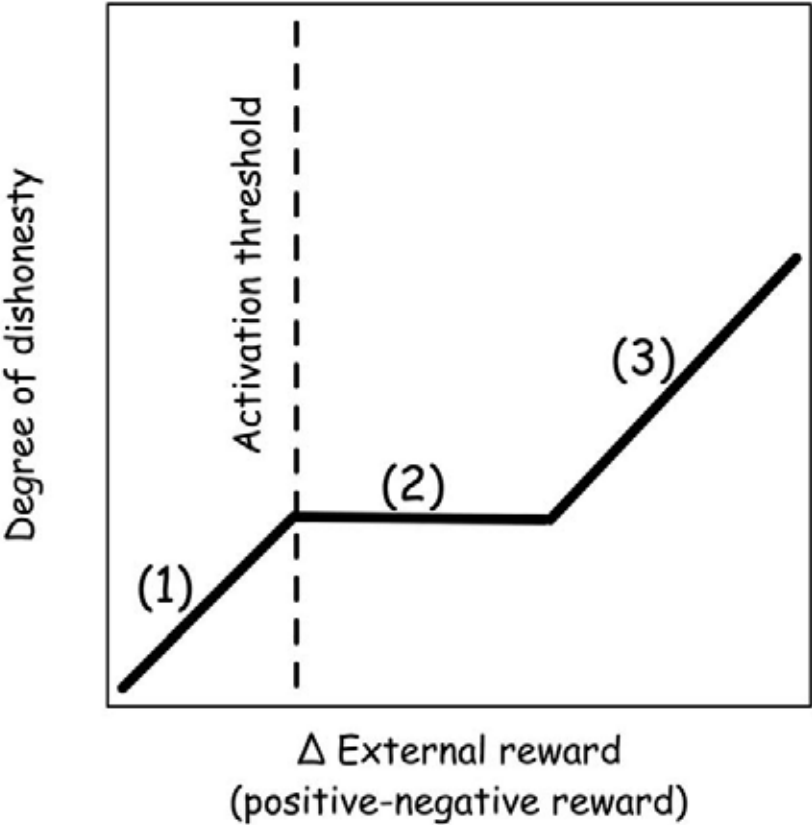
<sup>2</sup> We assume that the external reward mechanism as postulated by classical economics is always active, at least in normally healthy individuals.

## Figures

**Figure 1:** The relationship between expected net benefits of dishonesty and the propensity for dishonesty, according to a theory that includes only external rewards.



**Figure 2:** The relationship between expected net benefits of dishonesty and the propensity for dishonesty, based on the results of Mazar, Amir, and Ariely: namely, that dishonesty itself can activate the internal reward mechanisms, which, in turn, interrupt the standard relationship between external rewards and the propensity for being dishonest (as postulated by classical economics).



## References

Accenture, Inc. (2003), "One-Fourth of Americans Say it's Acceptable to Defraud Insurance Companies, Accenture Survey Finds," press release, Feb 12, [http://www.accenture.com/xd/xd.asp?it=enweb&xd=\\_dyn%5Cdynamicpressrelease\\_577.xml](http://www.accenture.com/xd/xd.asp?it=enweb&xd=_dyn%5Cdynamicpressrelease_577.xml).

Aharon, Itzhak, Nancy Etcoff, Dan Ariely, Christopher F. Chabris, Ethan O'Connor, and Hans C. Breiter (2001), "Beautiful Faces Have Variable Reward Value: fMRI and Behavioral Evidence," *Neuron*, 32 (3), 537-51.

Alicke, Mark D., M. L. Klotz, David L. Breitenbecher, Tricia J. Yurak, and Debbie S. Vredenburg (1995), "Personal contact, individuation, and the better-than-average effect," *Journal of Personality and Social Psychology*, 68 (5), 804-25.

Anderson, Stephen J., Gillian Cohen, and Stephanie Taylor (2000), "Rewriting the past: Some factors affecting the variability of personal memories," *Applied Cognitive Psychology*, 14 (5), 435-54.

Andreoni, James, William T. Harbaugh, and Lise Vesterlund (2003), "The carrot or the stick: Rewards, punishments, and cooperation," *The American Economic Review*, 93 (3), 893-902.



---- and John H. Miller (2005), "Analyzing choice with Revealed Preference: Is Altruism Rational?" in Handbook of Experimental Economics Results, Volume 1, Charles Plott and Vernon L. Smith, eds. Elsevier Science.

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

----, ----, and Samuel Issacharoff (1997), "Creating Convergence: Debiasing Biased Litigants," Law & Social Inquiry, 22 (4), 913-25.

Bagozzi, Richard P. (1995), "Reflections on Relationship Marketing in Consumer Markets," Journal of the Academy of Marketing Science, 23 (4), 272-77.

Barkan Rachel, Dov Zohar, and Ido Erev (1998), "Accidents and Decision Making Under Risk: A Comparison of Four Models", Organizational Behavior and Human Decision Processes, 74 (2), 118-144.

Bazerman, Max H. and George Loewenstein (2001), "Taking the bias out of bean counting," Harvard Business Review, 79 (1), 28.

----, ----, and Don A. Moore (2002), "Why good accountants do bad audits," Harvard Business Review, 80 (11), 96-102.

Beaman, Arthur L., Bonnel Klentz, Edward Diener, and Soren Svanum (1979), "Self-awareness and transgression in children: Two field studies," *Journal of Personality and Social Psychology*, 37 (10), 1835-46.

Campbell, Ernest Q. (1964), "The Internalization of Moral Norms," *Sociometry*, 27 (4), 391-412.

Carver, Charles S. and Michael F. Scheier (1998), *On the self-regulation of behavior*. New York: Cambridge University Press.

---- and ---- (1978), "Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence," *Journal of Personality and Social Psychology*, 36 (3), 324-32.

Davis, Ann (2004), "Open Secrets; Head of the Line: Client comes first? On Wall Street, it isn't always so; investing own money, firms can misuse knowledge of a big impending order; mischief in the 'Back Books'," *The Wall Street Journal*, Dec 16, A.1.

de Quervain, Dominique J.-F., Urs Fischbacher, Valerie Treyer, Melanie Schelthammer, Ulrich Schnyder, Alfred Buck, and Ernst Fehr (2004), "The neural basis of altruistic punishment," *Science*, 305 (5688), 1254-58.

de Rugy, Veronique (2005), "What does homeland security spending buy?" working paper, American Enterprise Institute for Public Policy Research, [http://www.aei.org/publications/pubID.21483/pub\\_detail.asp](http://www.aei.org/publications/pubID.21483/pub_detail.asp).

DePaulo, Bella M. and Wendy L. Morris (2004), "Discerning lies from truth: behavioral cues to deception and the indirect pathway of intuition," in *The Detection of Deception in Forensic Contexts*, Pär Anders Granhag and Leif A. Strömwall, eds. New York: Cambridge University Press, 15-40.

Dienstbier, Richard A. (1975), "An emotion-attribution approach to moral behavior: Interfacing cognitive and avoidance theories of moral development," *Psychological Review*, 82 (4), 299-315.

---- (1972), "The role of anxiety and arousal attribution in cheating," *Journal of Experimental Social Psychology*, 8 (2), 168-79.

---- and Pamela O. Munter (1971), "Cheating as a function of the labeling of natural arousal," *Journal of Personality and Social Psychology*, 17 (2), 208-13.

Duval, Shelley and Robert A. Wicklund (1972), *A theory of objective self awareness*. New York: Academic Press.

Erev, Ido, Paul Ingram, Ornit Raz, and Dror Shany (2003), "On the possibility of gentle rule enforcement," working paper, Department of Industrial Engineering and Management, Technion.

Fehr, Ernst and Urs Fischbacher (2003), "The nature of human altruism," *Nature*, 425, 785-91.

---- and ---- (2004), "Social norms and human cooperation," *Trends in Cognitive Sciences*, 8 (4), 185-90.

---- and Simon Gächter (2002), "Altruistic punishment in humans," *Nature*, 415, 137-40.

Fenigstein, Allan and Michael P. Levine (1984), "Self-attention, concept activation, and the causal self," *Journal of Experimental Social Psychology*, 20 (3), 231-45.

Freud, Sigmund (1933), *New introductory lectures on psycho-analysis*. New York: W. W. Norton & Company.

---- and James Strachey (1962), *The ego and the id*. New York: Norton.

----, ----, and Peter Gay (1989), *An outline of psycho-analysis (Standard ed.)*. New York: W.W. Norton.

Gneezy, Uri (2005), "Deception: The role of consequences," *American Economic Review*, 95 (1), 384-94.

Graham, Carol, Robert E. Litan, and Sandip Sukhtankar (2002), "The bigger they are, the harder they fall: An estimate of the costs of the crisis in corporate governance," working paper, The Brookings Institution,  
<http://www.brookings.edu/views/papers/graham/20020722Graham.pdf>.

Greenwald, Anthony G. (1997), "Self-knowledge and self-deception: Further consideration," in *The mythomanias: The nature of deception and self deception*, Michael S. Myslobodsky, ed. Mahwah: Lawrence Erlbaum Associates, 51-71.

Hechter, Michael (1990), "The Attainment Of Solidarity In Intentional Communities," *Rationality and Society*, 2 (2), 142-55.

Henrich, Joseph, Robert Boyd, Sam Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath (2001), "In search of Homo economicus: Behavioral experiments in 15 small-scale societies," *American Economic Review*, 91 (2), 73-78.

Herman, Tom (2005), "Study Suggests Tax Cheating Is on the Rise; Most Detailed Survey in 15 Years Finds \$250 Billion-Plus Gap; Ramping Up Audits on Wealthy," *Wall Street Journal*, Mar 30, D.1.

Hobbes, Thomas (1968), *Leviathan*. Baltimore: Penguin Books.

Knutson, Brian, Charles M. Adams, Grace W. Fong, and Daniel Hommer (2001), "Anticipation of increasing monetary reward selectively recruits nucleus accumbens," *Journal of Neuroscience*, 21 (16), RC159 (1-5).

Lewicki, R. J. (1984), "Lying and deception: A behavioral model," in *Negotiation in Organizations*, Max H. Bazerman and Roy J. Lewicki, eds. Beverly Hills: Sage Publication, 68-90.

Lifton, Robert J. (1986), "Reflections on genocide," *Psychohistory Review*, 14 (3), 39-54.

Loftus, Elizabeth F. (1994), "The repressed memory controversy," *American Psychologist*, 49 (5), 443-45.

Lohmann, Fred von (2004), "Is suing you're your customers a good idea," *Law.com Practice Center*, Sep 29, <http://www.law.com/jsp/article.jsp?id=1095434496352>.

Mazar, Nina, On Amir, and Dan Ariely (2005), "(Dis)Honesty: a combination of internal and external rewards," working paper, Sloan School of Management, Massachusetts Institute of Technology.

McDonald, Ian (2002), "Brokers get extra incentive to push funds," *The Wall Street Journal*, Apr 8, C.17.

Murphy, Patrick E. and Gene R. Laczniak (1981), "Marketing Ethics: A Review with Implications for Managers, Educators and Researchers," in *Review of Marketing*, Ben M. Enis and Kenneth J. Roering, eds. Chicago: American Marketing Association, 251-66.

Nagin, Daniel S. and Greg Pogarsky (2003), "An experimental investigation of deterrence: Cheating, self-serving bias, and impulsivity," *Criminology*, 41 (1), 501-27.

Norton, Mike and Dan Ariely (2005), "Self-Deception: How we come to believe we are better than we truly are," working paper, Sloan School of Management, Massachusetts Institute of Technology.

O'Doherty, John P., Ralf Deichmann, Hugo D. Critchley, and Raymond J. Dolan (2002), "Neural Responses during Anticipation of a Primary Taste Reward," *Neuron*, 33 (5), 815-26.

Olds, James and Peter Milner (1954), "Positive Reinforcement Produced By Electrical Stimulation Of Septal Area And Other Regions Of Rat Brain," *Journal of Comparative and Physiological Psychology*, 47, 419-27.

Rilling, James K., David A. Gutman, Thorsten R. Zeh, Giuseppe Pagnoni, Gregory S. Berns, and Clinton D. Kilts (2002), "A Neural Basis for Social Cooperation," *Neuron*, 35 (2), 395-405.

Schachter, Stanley and Bibb Latane (1964), "Crime, cognition, and the autonomic nervous system," *Nebraska Symposium on Motivation*, 12, 221-75.

---- and Jerome Singer (1962), "Cognitive, social, and physiological determinants of emotional state," *Psychological Review*, 69 (5), 379-99.

---- and Chad S. Dodson (2002), "Misattribution, false recognition and the sins of memory," in *Episodic memory: New directions in research*, Alan Baddeley, Martin Conway, and John Aggleton, eds. London: Oxford University Press, 71-85.

Schultz, Wolfram, Paul Apicella, Eugenio Scarnati, and Tomas Ljungberg (1992), "Neuronal activity in monkey ventral striatum related to the expectation of reward," *Journal of Neuroscience*, 12 (12), 4595-610.

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.



Smith, Adam (1999), *The wealth of nations: books IV-V*. London: Penguin Books.

---- (1997), *The wealth of nations. Books I-III*. New York: Penguin Books USA.

Speights, David and Mark Hilinski (2005), "Return fraud and abuse: How to protect profits," *Retailing Issues Letter*, 17 (1), 1-6.

Trivers, Robert (2000), "The elements of a scientific theory of self-deception," in *Evolutionary perspectives on human reproductive behavior*, Peter Moller and Dori LeCroy, eds. New York: New York Academy of Sciences, 114-131.

U.S. Department of Justice, Office of the Attorney General (2004), "Report of the Department of Justice's Task Force on Intellectual Property," [www.usdoj.gov/criminal/cybercrime/IPTaskForceReport.pdf](http://www.usdoj.gov/criminal/cybercrime/IPTaskForceReport.pdf).

Vitell, Scott J. (2003), "Consumer Ethics Research: Review, Synthesis and Suggestions for the Future," *Journal of Business Ethics*, 43 (1-2), 33-47.