The Dishonesty of Honest People:

A Theory of Self-Concept Maintenance

Nina Mazar

University of Toronto, 105 St. George Street, Toronto, ON M5S3E6,

phone: 416-946-5650, fax: 416-978-5433, nina.mazar@utoronto.ca

On Amir

University of California San Diego, Otterson Hall, 9500 Gilman Drive,

MC 0553, La Jolla, CA 92093-0553, phone: 858-534-2023, fax: 858-534-0745, oamir@ucsd.edu

Dan Ariely*

Duke University, One Towerview Road, Durham, NC 27708

phone: 919-660-7703, fax 919-681-6246, dandan@duke.edu

Author Note

*We thank Daniel Berger, Anat Bracha, Aimee Drolee, and Tiffany Kosolcharoen for their help in conducting the experiments, as well as Ricardo E. Paxson for his help in creating the matrices.

The Dishonesty of Honest People:

A Theory of Self-Concept Maintenance

ABSTRACT

People like to think of themselves as honest. On the other hand, dishonesty pays, and it often pays quite well. How do people resolve this tension? Our research shows that people behave dishonestly enough to profit, but honestly enough to delude themselves of their own integrity. A little bit of dishonesty gives a taste of profit without spoiling a positive self-view. Two mechanisms allow for such self-concept maintenance: inattention to moral standards and categorization malleability. Six experiments support our theory of self-concept maintenance and offer practical applications for curbing dishonesty in everyday life.

THE DISHONESTY OF HONEST PEOPLE: A THEORY OF SELF-CONCEPT MAINTENANCE

It is almost impossible to open a newspaper or turn on a television without being exposed to a report of dishonest behavior of one type or another. To give a few examples, "wardrobing"—the purchase, use, and then return of the used clothing—costs the U.S. retail industry an estimated $16 billion annually (Speights and Hilinski 2005); the overall magnitude of fraud in the U.S. property and casualty insurance industry is estimated to be 10% of total claims payments, or $24 billion annually (Accenture 2003); and the "tax gap," or the difference between what the IRS estimates taxpayers should pay and what they actually do pay, exceeds $300 billion annually (more than 15% noncompliance rate; Herman 2005). If this evidence is not disturbing enough, perhaps the largest contribution to dishonesty comes from employee theft and fraud that has been estimated at $600 billion a year in the U.S. alone — an amount almost twice the market capitalization of General Electric (Association of Certified Fraud Examiners 2006).

WHY ARE PEOPLE (DIS)HONEST?

Rooted in the philosophy of Thomas Hobbes, Adam Smith, and the standard economic model of rational and selfish human behavior (i.e., *homo economicus*) is the belief that people carry out dishonest acts consciously and deliberatively by trading off the expected external benefits and costs of the dishonest act (Becker 1968; Allingham and Sandmo 1972). According to this perspective, people would consider three aspects as they pass a gas station: the expected amount

of cash they stand to gain from robbing the place, the probability of being caught, and the magnitude of punishment if caught in this act. On the basis of these inputs, people reach a decision that maximizes their interests. Thus, according to this perspective, people are honest or dishonest only to the extent that the planned trade-off favors a particular action (Hechter 1990; Lewicki 1984). In addition to being central to economic theory, this external cost-benefit view plays an important role in the theory of crime and punishment, which forms the basis for most policy measures aimed at preventing dishonesty and guides punishments against those who exhibit dishonest behavior. In summary, this standard external cost-benefit perspective generates three hypotheses as to the forces that are expected to increase the frequency and magnitude of dishonesty: higher magnitude of external rewards (Ext-H1), lower probability of being caught (Ext-H2), and lower magnitude of punishment (Ext-H3).

From a psychological perspective, and in addition to financial considerations, another set of important inputs to the decision whether to be honest (or not) is based on internal rewards. Psychologists show that as part of socialization, people internalize the norms and values of their society (Campbell 1964; Henrich et al. 2001), which serve as an internal benchmark against which a person compares her behavior. Compliance with the internal values system provides positive rewards, whereas noncompliance leads to negative rewards (i.e. punishments). The most direct evidence regarding the existence of such internal reward mechanisms comes from brain imaging studies revealing that acts based on social norms, such as altruistic punishment or social cooperation (de Quervain et al. 2004; Rilling et al. 2002), activate the same primary reward centers in the brain (i.e., nucleus accumbens and caudate nucleus) that external benefits such as preferred food, drinks, and monetary gains do (Knutson et al. 2001; O'Doherty et al. 2002).

Applied to the context of (dis)honesty, we propose that one major way in which the internal reward system exerts control over behavior is by influencing people's self-concept—that is, the way individuals view and perceive themselves (Aronson 1969; Baumeister 1998; Bem 1972). Indeed, it has been shown that people typically value honesty (i.e., honesty is part of their internal reward system), that they have very strong beliefs in their own morality, and that they want to maintain this aspect of their self-concept (Griffin and Ross 1991; Sanitioso, Kunda, and Fong 1990; Greenwald 1980; Josephson Institute of Ethics 2004). This means that if a person fails to comply with her internal standards for honesty, she will have to negatively update her self-concept, which is aversive. On the other hand, if a person complies with her internal standards she avoids such negative updating and maintains her positive self-view in terms of being an honest person. Interestingly, this perspective suggests that in order to maintain their positive self-concepts, individuals will comply with their internal standards even when doing so involves investments of effort or sacrificing financial gains (e.g., Aronson and Carlsmith 1962; Harris, Mussen, and Rutherford 1976; Sullivan 1953).

In our gas station example, this perspective suggests that people who pass by a gas station will be influenced not only by the expected amount of cash they stand to gain from robbing the place, the probability of being caught, and the magnitude of punishment if caught, but also by the manner in which the act of robbing the store might make them perceive themselves.

The utility derived from behaving in line with one's self-concept could conceivably be just another part of the cost–benefit analysis (i.e., adding another variable to account for this utility). However, even if we consider this utility as just another input, it probably cannot be manifested as a simple constant, because the influence of dishonest behavior on the self-concept will

most likely depend on the particular action, its symbolic value, its context, and its plasticity. In the following sections we characterize these elements in a theory of self-concept maintenance, and test the implications of this theory in a set of six experiments.

THE THEORY OF SELF-CONCEPT MAINTENANCE

People are often torn between two competing motivations: gaining from cheating versus maintaining their positive self-concept as honest individuals (Aronson 1969; Harris, Mussen, and Rutherford 1976). If they cheat, they could, for example, gain financially, but at the expense of an honest self-concept. In contrast, if they take the high road, they might forgo financial benefits but maintain their honest self-concept. This seems to be a win–loose situation; choosing one path involves sacrificing the other.

In this work, we suggest that people typically solve this motivational dilemma adaptively by finding a balance or equilibrium between the two motivating forces, such that they derive some financial benefit from behaving dishonestly but still maintain their positive self-concept in terms of being honest individuals. To be more precise, we posit a magnitude-range of dishonesty within which people can cheat, but their behaviors, which they would usually consider dishonest[1], do not bear negatively on their self-concept (they are not forced to update their self-concept). Although many mechanisms may allow people to find such a compromise, we focus on two particular means, categorization and attention devoted to ones own moral standards. Using these mechanisms, individuals are able to record their actions (e.g., "I am claiming $x in tax exemptions") without confronting the moral meaning of their actions (e.g., "I am dishonest"). We focus on these two mechanisms because they support the role of the self-concept in decisions

about honesty and because we believe they have a wide set of important applications in the marketplace. Although not always mutually exclusive, we elaborate on each separately.

Categorization

We hypothesize that for certain types of actions and magnitudes of dishonesty, people can categorize their actions in more compatible terms and find rationalizations for their actions. As a consequence people can cheat while avoiding any negative self-signals that might affect their self-concept, and thus avoid negatively updating their self-concept altogether (Gur and Sackeim 1979).

Two important aspects of categorization are its relative malleability and its limit. Behaviors with malleable categorization are ones that allow people to reinterpret them in a self-serving manner, and the degree of malleability is likely to be determined by their context. For example, intuition suggests that it is easier to steal a 10¢ pencil from a friend than to steal 10¢ out of this friend's wallet to buy a pencil, because the former scenario offers more possibilities to categorize the action in terms that are compatible with friendship (e.g., my friend took a pencil from me once; this is what friends do). This thought experiment suggests that a higher degree of categorization malleability facilitates dishonesty (stealing), but also that some actions are inherently less malleable and therefore cannot be categorized successfully in compatible terms (Dana, Weber, and Kuang 2005; for a discussion of the idea that a medium like a pen can disguise the final outcome of an action – stealing –, see Hsee et al. 2003). In other words, as the categorization malleability increases, so does the magnitude of dishonesty a person can commit without influencing his or her self-concept (Baumeister 1998; Schweitzer and Hsee 2002; Pina e Cunha and Cabral-Cardoso 2006).

The second important aspect of the categorization process pertains to its inherent limit. The ability to categorize behaviors in ways other than as dishonest or immoral can be incredibly useful for the self, but it is hard to imagine that this mechanism is without limits. Instead, it may be possible to "stretch" the truth and the bounds of mental representations only up to a certain point (what Jean Piaget [1950] calls assimilation and accommodation). If we assume that the categorization process has such built-in limits, we should conceptualize categorization as effective only up to a threshold, beyond which people can no longer avoid the obvious moral valence of their behavior.

Attention to Standards

The other mechanism that we address in the current work is the attention people pay to their own standards of conduct. This idea relates to Duval and Wicklund's (1972) theory of objective self-awareness and Langer's (1989) concept of mindlessness. We hypothesize that when people attend to their own moral standards (are mindful of them), any dishonest action is more likely to be reflected in their self-concept (they will update their self-concept as a consequence of their actions), which in turn will cause people to adhere to a stricter delineation of honest and dishonest behavior. However, when individuals are inattentive to their own moral standards (are mindless of them) their actions are not evaluated relative to their standards, their self-concept is less likely to be updated, and therefore their behavior is likely to diverge from their standards. Thus, the attention to standards mechanism predicts that in cases in which ones moral standards are more accessible, people will have to confront the meaning of their actions more readily and therefore be more honest (for ways to increase accessibility, see Bateson, Nettle, and Roberts 2006; Bering, McLeod, and Shackelford 2005; Diener and Wallbom 1976; Haley and Fessler

2005). In this sense, greater attention to standards may be modeled as a tighter range for the magnitude of dishonest actions that does not trigger updating of the self-concept, or a lower threshold up to which people can be dishonest without influencing their self-concept.

Categorization and Attention to Standards

Whereas the categorization mechanism depends heavily on stimuli and actions (i.e., degree of malleability and magnitude of dishonesty), the attention to standards mechanism relies on internal awareness or salience. From this perspective, these two mechanisms are distinct: the former focuses on the outside world, whereas the latter on the inside world. However, they are related in that they both involve attention, are sensitive to manipulations, and relate to the dynamics of acceptable boundaries of behavior.

Thus, while the dishonesty that both self-concept maintenance mechanisms allow stems from different sources, they both tap the same basic concept. Moreover, in many real-world cases, these mechanisms may be so interrelated that it would be hard to clearly distinguish whether the source of this type of dishonesty comes from the environment (categorization) or the individual (attention to standards). In sum, the theory of self-concept maintenance that considers both external and internal rewards-systems suggests the following hypotheses:

Ext&Int-H1: Dishonesty increases as attention to standards for honesty decreases.

Ext&Int-H2: Dishonesty increases as categorization malleability increases.

Ext&Int-H3: Given the opportunity to be dishonest, individuals will be dishonest up to a certain level that does not force them to update their self-concept.

EXPERIMENT 1: INCREASING ATTENTION TO STANDARDS FOR HONESTY THROUGH RELIGIOUS REMINDERS

The general setup of all our experiments involved a multiple-question task, in which participants got paid according to their performance. We compared the performance of respondents in the control conditions, in which they had no opportunity to be dishonest, with "cheating" conditions, in which they had such an opportunity. In Experiment 1, we tested the prediction that increasing people's attention to their standards for honesty would make them more honest, by contrasting the magnitude of dishonesty in a condition in which participants were reminded of their own standards for honesty with a condition in which they were not.

On the face of it, the idea that any reminder can decrease dishonesty seems strange—after all, don't people know that it is wrong to be dishonest even without such reminders? However, from the self-concept maintenance perspective, the question is not whether a person knows it is wrong to behave dishonestly, but whether she thinks of these standards and compares her behavior to them in the moment of temptation. In other words, if a mere reminder of standards for honesty has an effect, we may assert that people don't naturally attend to those standards. In Experiment 1 we implemented this reminder through a simple recall task.

Method

Two hundred twenty-nine students participated in this experiment, which consisted of a two-task paradigm as part of a broader experimental session with multiple, unrelated paper-and-pencil tasks that appeared together in a booklet. In the first task, we asked respondents to either write down the names of 10 books they had read in high school (no moral reminder) or the Ten Commandments (moral reminder). They had two minutes to complete this task. The idea of the

Ten Commandments recall task was that independent of people's religion, whether people believed in God, or knew any of the commandments, knowing that the Ten Commandments are about moral rules would be enough to increase attention to their own moral standards and increase the likelihood of behavior consistent with these standards (for a discussion of reminders of God in the context of generosity, see Shariff and Norenzayan 2007). The second, ostensibly separate task, consisted of two sheets of paper: a test sheet and an answer sheet. The test sheet consisted of 20 matrices, each based on a set of 12 three-digit numbers. Participants had four minutes in which to find two numbers per matrix that added up to 10 (see Figure 1). We selected this type of task because it is a search task, and though it can take some time to find the right answer, once found, the respondents could unambiguously evaluate whether they had solved the question correctly (assuming that they could add two numbers to 10 without error), without the need for a solution sheet and the possibility of a hindsight bias (Fischhoff and Beyth 1975). Moreover, we used this task based on a pre-test showing that participants did not think of this task as one that reflected on their math ability or intelligence. The answer sheet was used to report the total number of correctly solved matrices. We promised that at the end of the session, two randomly selected participants would earn $10 for each correctly solved matrix.

••• Figure 1 •••

In the two control conditions (after the 10 books and Ten Commandments recall task, respectively), at the end of the four minutes matrix task, participants continued with the next task in the booklet. At the end of the entire experimental session, the experimenter verified their answers on the matrix task and wrote down the number of correctly solved matrices on the answer sheet in the booklet. In the two recycle conditions (after the 10 books and Ten Commandments

recall task, respectively), at the end of the four minutes matrix task, participants indicated the total number of correctly solved matrices on the answer sheet, and then tore out the original test sheet from the booklet and placed it in their belongings (to recycle it later), thus providing them with an opportunity to cheat. The entire experiment represented a 2 (type of reminder) × 2 (ability to cheat) between-subjects design.

Results and Discussion

The results of Experiment 1 confirmed our predictions. The type of reminder had no effect on participants' performance in the two control conditions ($M_{\text{Books/control}} = 3.1$ vs. $M_{\text{Ten Commandments/control}} = 3.1$; $F(1,225) = .012$, $p = .91$), which suggest that the type of reminder did not influence ability or motivation. Following the book recall task however, respondents cheated when they were given the opportunity to do so ($M_{\text{Books/recycle}} = 4.2$) but they did no cheat after the Ten Commandments recall task ($M_{\text{Commandments/recycle}} = 2.8$; $F(1,225) = 5.24$, $p = .023$), creating a significant interaction between type of reminder and ability to cheat ($F(3, 225) = 4.52$, $p = .036$). Interestingly, the level of cheating remained far below the maximum. In fact, participants cheated on average "only" 6.7% of the possible magnitude. Most important, and in line with our self-concept maintenance idea, reminding participants of standards for morality eliminated cheating completely: In the Ten Commandments/recycle condition participants' performance was undistinguishable from those in the control conditions ($F(1,225) = .49$, $p = .48$).

We designed Experiment 1 to focus on the attention to standards mechanism (Ext&Int-H1), but one aspect of the results—the finding that the magnitude of dishonesty was limited and well below the maximum possible level in the two recycle conditions—suggested that the categorization mechanism (Ext&Int-H2) might have been at work as well.

One possible alternative interpretation of the books/recycle condition is that over their lifetime, participants had developed standards for moral behavior according to which over-claiming by a few questions on a test or in an experimental setting was not considered dishonest. If so, our participants could have been completely honest from their point of view. Similarly, in a corrupt country in which a substantial part of the citizenry overclaims on taxes, the very act of overclaiming is generally accepted and therefore not necessarily considered immoral. However, if this accounted for our findings, increasing people's attention to morality (Ten Commandments /recycle condition) would not have decreased the magnitude of dishonesty. Therefore, we interpreted these findings as providing initial support for the self-concept maintenance theory.

It is also interesting to note that participants, on average, remembered only 4.3 of the Ten Commandments, and we found no significant correlation between the number of commandments recalled and the number of matrices the participants claimed to have solved correctly ($r = -.14$, $p = .29$). If we use the number of commandments remembered as a proxy for religiosity, the lack of relationship between religiosity and magnitude of dishonesty suggests that the efficacy of the Ten Commandments is based on increased attention to ones internal honesty standards, leading to a lower tolerance for dishonesty (i.e., decreased self-concept maintenance threshold).

Finally, it is worth contrasting these results with people's lay theories about such situations. A separate set of students (n = 75) correctly anticipated that participants would cheat when given the opportunity to do so, but they anticipated that the level of cheating would be higher than what it really was ($M_{\text{pred\_Books/recycle}} = 9.5$) and they anticipated that reminding participants of the Ten Commandments would not significantly decrease cheating ($M_{\text{pred\_Commandments/recycle}} = 7.8$; $t(73)=1.61$, $p = .11$). The contrast of the predicted results with the actual behavior that we found

suggests that participants understand the economic motivation for overclaiming, but that they overestimate its influence on behavior, and that they underestimate the effect of the self-concept in regulating honesty.

EXPERIMENT 2: INCREASING ATTENTION TO STANDARDS FOR HONESTY THROUGH COMMITMENT REMINDERS

Another type of reminder, an honor code, refers to a procedure that asks participants to sign a statement in which they declare their commitment to honesty before taking part in a task (Dickerson et al. 1992; McCabe and Trevino 1993, 1997). While many explanations have been proposed for the effectiveness of honor codes used by many academic institutions (McCabe, Trevino, and Butterfield 2002; see http://www.academicintegrity.org), the self concept maintenance idea may shed light on the internal process underlying its success. In addition to manipulating the awareness of honesty standards through commitment reminders at the point of temptation, Experiment 2 represented an extension of the Experiment 1 by manipulating the financial incentives for performance (i.e., external benefits), and by doing so also tested the external cost/benefit hypothesis that dishonesty increases as the expected magnitude of reward from the dishonest act increases (Ext-H1).

Method

Two hundred seven students participated in the Experiment 2. Using the same matrix task, we manipulated two factors between participants: the amount earned per correctly solved matrix (50¢ and $2 – paid to each participant) and the attention to standards (control, recycle, and recycle+honor code).

In the two control conditions at the end of five minutes, participants handed both the test and answer sheets to the experimenter, who verified their answers and wrote down the number of correctly solved matrices on the answer sheet. In the two recycle conditions, participants indicated the total number of correctly solved matrices on the answer sheet, folded the original test sheet, and placed it in their belongings (to recycle it later), thus providing them an opportunity to cheat. Only after that did they hand the answer sheet to the experimenter. The recycle+honor code condition was similar to the recycle condition except that at the top of the test sheet there was an additional statement that read: "I understand that this short survey falls under MIT's [Yale's] honor system." Participants printed and signed their names below the statement. Thus, the honor code statement appeared on the same sheet as the matrices, and this sheet was recycled before participants submitted their answer sheets. In addition, to provide a test for Ext-H1, we manipulated the payment per correctly solved matrix (50¢ and $2) and contrasted performance levels between these two incentive levels.

Results and Discussion

Figure 2 depicts the results. An overall ANOVA revealed a highly significant effect of the attention to standards manipulation ($F(2,201) = 11.94$, $p < .001$), no significant effect of the level of incentive manipulation ($F(1,201) = .99$, $p = .32$), and no significant interaction ($F(2,201) = .58$, $p = .56$). When given the opportunity, respondents in the two recycle conditions (50¢ and $2) cheated ($M_{recycle} = 5.5$) relative to those in the two control conditions (50¢ and $2: $M_{control} = 3.3$; $F(1,201) = 15.99$, $p < .001$), but again, the level of cheating fell far below the maximum (i.e., 20); participants cheated "only" 13.5% of the possible average magnitude. In line with our

findings in Experiment 1, this latter result supports the idea that we were also observing the workings of the categorization mechanism.

Between the two levels of incentives (50¢ and $2 conditions), we did not find a particularly large difference in the magnitude of cheating; in fact, cheating was slightly more common (by approximately 1.16 questions), though not significantly so, in the 50¢ condition ($F(1,201) = 2.1$, $p = .15$). Thus, we did not find support for Ext-H1. One possible interpretation of this decrease in dishonesty with increased incentives is that the magnitude of dishonesty and its effect on the categorization mechanism depended on both the number of questions answered dishonestly (which increased by 2.8 in the 50¢ condition and 1.7 in the $2 condition) and on the amount of money inaccurately claimed (which increased by $1.4 in the 50¢ condition and $3.5 in the $2 condition). If categorization malleability was affected by a mix of these two factors, we would have expected the number of questions that participants reported as correctly solved to decrease with greater incentives (at least as long as the external incentives were not too high).

Most important for Experiment 2, we found that the two recycle+honor code conditions (50¢ and $2: $M_{\text{recycle+honor code}} = 3.0$) eliminated cheating to the extent that the performance in these conditions was undistinguishable from the two control conditions (50¢ and $2: $M_{\text{control}} = 3.3$; $F(1,201) = .19$, $p = .66$) but significantly different from the two recycle conditions (50¢ and $2: $M_{\text{recycle}} = 5.5$; $F(1,201) = 19.69$, $p < .001$). The latter result is interesting given that the two recycle+honor code conditions were procedurally very similar to the two recycle conditions. Moreover, it is worth noting that the two institutions in which we conducted this experiment did not have an honor code system at the time and therefore, objectively, the honor code had no implications of external punishment. When we replicated the experiment in an institution that had a

very strict honor code, the results were identical, suggesting that it is not the honor code per-se and its implied external punishment, but rather the reminder of morality that was at play.

••• Figure 2 •••

Again we asked a separate set of students (n = 82) at the institutions without an honor code system to predict the results, and while they predicted that the increased payment would marginally increase dishonesty ($M_{pred\_\$2}$ = 6.8 vs. $M_{pred\_\$0.5}$ = 6.4, $F(1,80) = 3.3$, p=.07), in essence predicting Ext-H1, they did not anticipate that the honor code would significantly decrease dishonesty ($M_{pred\_recylce+honor\ code}$ = 6.2 vs. $M_{pred\_recycle}$ = 6.9, $F(1,80) = .74$, p=.39). The contrast of the predicted results with the actual behavior suggests that people understand the economic motivation for overclaiming, that they overestimate its influence on behavior, and that again they underestimate the effect of the self-concept in regulating honesty. In addition, the fact that predictors did not expect the honor code to decrease dishonesty suggests that they did not perceive the honor code manipulation to have implications of external punishment.

EXPERIMENT 3: INCREASING CATEGORIZATION MALLEABILITY

Making people mindful by increasing their attention to their honesty standards can curb dishonesty, but the theory of self-concept maintenance also implies that increasing the malleability to interpret ones actions should increase the magnitude of dishonesty (Schweitzer and Hsee 2002). To test this hypothesis, in Experiment 3, we manipulated whether the opportunity for dishonest behavior occurred in terms of money or in terms of an intermediary medium (tokens). We posited that introducing a medium (Hsee et al. 2003) would offer participants more room for in-

terpretation of their actions, thereby making the moral implications of dishonesty less accessible, and hence making it easier for participants to cheat at higher magnitudes.

Method

Four hundred fifty students participated in our experiment. Participants had five minutes to complete the matrix task, and were promised 50¢ for each correctly solved matrix. We used three between-subjects conditions: the same control and recycle conditions as in Experiment 2, and a recycle+token condition. The latter condition was similar to the recycle condition, except that participants knew that each correctly solved matrix would earn them 1 token, which they would exchange for 50¢ a few seconds later. Once the five minutes elapsed, participants in the recycle+token condition recycled their test sheet and submitted only their answer sheet to an experimenter, who gave them the corresponding amount of tokens. Participants then went to a second experimenter, who exchanged the tokens for money (this experimenter also paid the participants in the other conditions). We counterbalanced the roles of the two experimenters.

Results and Discussion

Similar to our previous findings, participants in the recycle condition solved significantly more questions than participants in the control condition ($M_{recycle} = 6.2$; $M_{control} = 3.5$; $F(1,447) = 34.26$, $p < .001$), which suggests that they cheated. What is more, participants' magnitude of cheating was well below the maximum—only 16.5% of the possible average magnitude. Most interestingly, and in line with our hypothesis Ext&Int-H2, introducing tokens as the medium of immediate exchange further increased the magnitude of dishonesty ($M_{recyle+token} = 9.4$) such that it was significantly larger than in the recycle condition ($F(1,447) = 47.62$, $p < .001$)—presumably without any changes in the probability of being caught or the severity of the punishment.

Our findings support the idea that claiming more tokens instead of claiming more money offered more categorization malleability such that people could interpret their dishonesty in a more self-serving manner -- reducing the negative self-signal that they otherwise would have received. In terms of our current account, the recycle+token condition increased the threshold for the acceptable magnitude of dishonesty. The finding that a medium could be such an impressive facilitator of dishonesty may explain the incomparably excessive contribution of employee theft and fraud (e.g., stealing office supplies and merchandise, putting inappropriate expenses on expense accounts) to dishonesty in the marketplace, as reported in the introduction.

Finally, it is worth pointing out that our results differ from what a separate set of students (n = 59) predicted we would find. The predictors correctly anticipated that participants would cheat when given the opportunity to do so ($M_{\text{pred\_recycle}} = 6.6$, $t(29) = 5.189$, $p < 0.001$), but they anticipated that being able to cheat in terms of tokens would not be any different than being able to cheat in terms of money ($M_{\text{pred\_recycle+token}} = 7$, $t(57) = 4.5$, $p = .65$). This again suggests that individuals underestimate the effect of the self-concept in regulating honesty.

EXPERIMENT 4: RECOGNIZING ONES ACTIONS BUT NOT UPDATING THE SELF-CONCEPT

Our account of self-concept maintenance suggests that by engaging only in relatively low level of cheating, our participants stayed within the threshold of acceptable magnitudes of dishonesty and thereby benefited from being dishonest without receiving a negative self-signal (i.e., their self-concept remained unaffected). To achieve this balance, we posit that people recorded their actions correctly (i.e., they knew that they were overclaiming), but the categorization and/or

attention to standards mechanisms prevented this factual knowledge from being morally evaluated. Thus, people did not necessarily confront the true meaning or implications of their actions (e.g., "I am dishonest"). We test this prediction (Ext&Int-H3) in Experiment 4.

To test the hypothesis that people know of their actions but do not update their self-concepts, we manipulated participants' ability to cheat on the matrix task and measured their predictions about their performance on a second matrix task that did not allow cheating. If participants in a recycling condition did not recognize that they overclaimed, they would base their predictions on their exaggerated (i.e., dishonest) performance in the first matrix task. Their predictions therefore would be higher than the predictions of those who could not cheat on the first task. If, however, participants who overclaimed were cognizant of their exaggerated claims, their predictions for a situation that does not allow cheating would be attenuated, and theoretically would not differ from their counterparts' in the control condition. In addition, to test whether dishonest behavior influenced people's self-concept, we asked participants about their honesty after they completed the first matrix task. If participants in the recycling condition (who were cheating) had lower opinions about themselves in terms of honesty than those in the control condition (who were not cheating), this would mean that they had updated their self-concept. But if cheating did not influence their opinions about themselves, this would suggest that they had not fully accounted for their dishonest behaviors, and consequently, that they had not paid a price for their dishonesty in terms of their self-concept.

Method

Forty-four students participated in this experiment, which consisted of a four-task paradigm, administered in the following order: a matrix task, a personality test, a prediction task, and

a second matrix task. In the first matrix task, we repeated the same control and recycle conditions from Experiment 2. Participants randomly assigned to either of these two conditions had five minutes to complete the task and received 50¢ per correctly solved matrix. The only difference from Experiment 2 was that we asked all participants (not just those in the recycle condition) to report on the answer sheet the total number of matrices they had correctly solved (participants in the control condition then submitted both the test and the answer sheets to the experimenter, who verified each of their answers on the test sheets in order to determine participants' payments).

In the second, ostensibly separate task, we handed out a 10-item test with questions ranging from political ambitions to preferences for classical music to general abilities. Embedded in this survey were two questions about their self-concept as it relates to honesty. One question asked how honest they considered themselves to be (absolute honesty) on a scale from 0 (not at all) to 100 (very). The other question asked participants to rate their perception of themselves in terms of being a moral person (relative morality) on a scale from –5 (much worse) to 5 (much better) at the time of the survey in contrast to the day before.

In the third task, we surprised our participants by announcing that they would next participate in a second five-minute matrix task, but before taking part in it, their task was to predict how many matrices they would be able to solve and indicate how confident they were with their predictions on a scale from 0 (not at all) to 100 (very). Before making these predictions, we made it clear that this second matrix task left no room to overclaim as the experimenter would check the answers given on the test sheet (as was done in the control condition). Furthermore, we informed participants that this second test would consist of a different set of matrices, and the payment would depend on both the accuracy of their prediction and their performance. If their

prediction was 100% accurate, they would earn 50¢ per correctly solved matrix, but for each matrix they solved more or less than what they predicted, their payment per matrix would be reduced by 2¢. We emphasized that this payment scheme meant that it was in their best interest to predict as accurately as possible and to solve as many matrices as they could (i.e., they would make less money if they gave up solving some matrices, just to be accurate in their predictions).

Finally, the fourth task was the matrix task with different number-sets and without the ability to overclaim (i.e., only control condition). The entire experiment thus represented a two-condition, between-subjects design, differing only in the first matrix task (possibility to cheat). The three remaining tasks (personality test, prediction task, second matrix task) were the same.

Results and Discussion

The mean number of matrices "solved" in the first and second matrix tasks appear in Table 1. Similar to our previous experiments, on the first task, participants who had the ability to cheat (recycle condition) solved significantly more questions than those in the control condition ($t(42) = 2.21$, $p = .033$). However, this difference disappeared in the second matrix task, for which neither of the two groups had an opportunity to cheat ($t(42) = .43$, $p = .67$), and the average performance on the second task ($M_{2ndMatrixTask} = 4.5$) did not differ from the control condition's performance on the first task ($M_{1stMatrixTask/conrol} = 4.2$; $t(43) = .65$, $p = .519$). These findings implied that, as in the previous experiments, participants cheated when they had the chance to do so. Furthermore, the level of cheating was relatively low (on average, two to three matrices); participants cheated "only" 14.8% of the possible average magnitude.

In terms of the predictions of performance on the second matrix task, we found no significant difference ($t(42) \sim 0$, *ns.*) between those participants who were able to cheat and those

who were not able to cheat in the first matrix task ($M_{control}$ = 6.3, recycle: $M_{recycle}$ = 6.3). Moreover, participants in the control and recycle conditions were equally confident about their predictions ($M_{forecast\_control}$ = 72.5, $M_{forecast\_recycle}$ = 68.8, $t(42) = .56$, $p = .57$). Together with the difference in performance in the first matrix task, these findings suggest that those who cheated in the first task knew they had overclaimed.

As for the 10 personality questions-survey, after the first task, participants in both conditions had equally high opinions of their honesty in general ($t(42) = .97$, $p = .34$) and their morality in comparison with the previous day ($t(42) = .55$, $p = .58$), which suggests that cheating in the experiment did not affect their reported self-concepts in terms of these characteristics. Together, these results support our self-concept maintenance theory and indicate that people's limited magnitude of dishonesty "flies under the radar"; they do not update their self-concept in terms of honesty even though they do recognize their actions (i.e., that they overclaim).

In addition, we asked a different group of 39 students to predict the responses to the self-concept questions (absolute honesty and relative morality). In the control condition, we asked them to imagine how an average student who solved four matrices would answer these two questions. In the recycle condition, we asked them to imagine how an average student who solved four matrices but claimed to have solved six, would answer these two questions. As can be seen in Table 1, they predicted that cheating would decrease both a person's general view of herself as an honest person ($t(37) = 3.77$, $p < .001$) as well as her morality compared with the day before the test ($t(37) = 3.88$, $p < .001$)[2]. This finding provides further support for the idea that individuals do not accurately anticipate the self-concept maintenance mechanism.

••• Table 1 •••

EXPERIMENT 5: NOT CHEATING DUE TO OTHERS

Thus far, we have accumulated evidence for a magnitude of cheating, which seems to depend on the attention one pays to own standards for honesty as well as categorization malleability. Moreover, the results of Experiment 4 provide some evidence that cheating can take place without an associated change in the self-concept. Overall, these findings are in line with our theory of self-concept maintenance: When people are torn between the temptation to benefit from cheating and the benefits of maintaining a positive view about themselves, they solve this dilemma by finding a balance between these two motivating forces such that they can engage to some level in dishonest behavior without updating their self-concept. Although these findings are consistent with our theory of self-concept maintenance, there are several other alternative accounts for these results. In the final two experiments, we try to deal with a few of these.

One possible alternative account that comes to mind posits that participants were driven by self-esteem only (e.g., John and Robins 1994; Tesser, Millar, and Moore 1988; Trivers 2000). From this perspective, a person might have cheated on a few matrices so that he or she did not appear stupid in comparison with everybody else (we used the matrix task partially because it is not a task that our participants seemed to relate to IQ, but this account might still be possible).

A second alternative for our findings might argue that participants were driven only by external, not internal, rewards and cheated up to the level that they believed their dishonest behavior could not be detected. From this perspective, participants cheated just by a few questions not because some internal force stopped them, but because they estimated that the probability of being caught and/or the severity of punishment would be negligible (or zero), if they only cheat

by a few questions. As a consequence, they cheated up to this particular threshold – in essence estimating what they could get away with and cheating up to that level.

A third alternative explanation is that the different manipulations (e.g., moral reminders) influenced the type of social norms that participants apply to the experimental setting (see Reno, Cialdini, and Kallgren 1993; for focusing effects, see Kallgren, Cialdini, and Reno 2000). According to this norm compliance argument, a person who solves three matrices but knows that on average people report having solved six should simply go ahead and do what others are doing: report six solved matrices (i.e., cheat by three matrices).

What these three accounts have in common is that all of them are sensitive to the (expected) behavior of others. In contrast, our self-concept maintenance theory implies that the level of dishonesty is set without reference to the level of dishonesty exhibited by others (at least in the short-term). This contrast suggests a simple test where we manipulate participants' beliefs about others' performance levels. If the level of cheating is driven by the desire for achievement, external costs, or norm compliance, then the number of matrices that participants claim to have solved should increase when they believe that the average performance of others is higher. If, however, the level of cheating is driven by self-concept maintenance considerations, then believing that others solve many more matrices should have no effect on the level of dishonesty.

Method

One hundred eight students participated in a matrix task experiment, where we manipulated two factors between participants: the ability to cheat (control and recycle, as in Experiments 2) and beliefs about the number of matrices that the average student solves in the given condition in the time allotted (four matrices, which is the accurate number, or eight matrices

which was an exaggeration). As before, the dependent variable was the number of matrices reported as being solved correctly. The experiment represented a 2x2 between-subjects design.

Results and Discussion

Participants in the two control conditions solved on average 3.3 and 3.4 matrices, while those in the corresponding recycle conditions solved 4.5 and 4.8 matrices (in the 4 and 8 believed standard performance conditions, respectively). A two-factorial ANOVA of the number of matrices solved as a function of the ability to cheat and the belief about others' performances showed a main effect of the ability to cheat ($F(1,104) = 6.89$, $p = .01$) but no main effect of the beliefs about average performance levels ($F(1,104) = .15$, $p = .7$), and no interaction ($F(1,104) = .09$, $p = .76$). That is, when participants had a chance to cheat, they cheated, but the level of cheating was independent of information about the average reported performance of others. This finding argues against drive toward achievement, threshold due to external costs, or norm compliance as alternative explanations for our findings.

EXPERIMENT 6: SENSITIVITY TO EXTERNAL REWARDS

Since the external costs of dishonest acts are central to the standard economic cost-benefit view of dishonesty, we wanted to test its influence more directly. In particular, following Nagin and Pogarsky's (2003) suggestion that increasing the probability of getting caught is much more effective than increasing the severity of the punishment, we aimed to manipulate the former type of external cost, that is, the likelihood of getting caught on three levels and measure the amount of dishonesty across these three cheating conditions. If only external cost-benefit trade-offs were at work in our setup, we should find that the level of dishonesty increases as the prob-

ability of being caught decreases (Ext-H2). On the other hand, if self-concept maintenance limits the magnitude of dishonesty, we should find some cheating, but the level of dishonesty should be roughly of the same magnitude, regardless of the probabilities of getting caught.

Method

This experiment entailed multiple sessions with each participant sitting in a private booth ($N = 326$). At the start of each session, the experimenter explained the instructions for the entire experiment. The first part of the experimental procedure remained the same for all conditions, but the second part varied across conditions. All participants received a test with 50 multiple-choice, general-knowledge questions (e.g., How deep is a fathom? How many degrees does every triangle contain? What does 3! equal?), had 15 minutes to answer the questions, and were promised 10¢ for each question they solved correctly. After the 15 minutes, participants received a "bubble sheet" onto which they transferred their answers. Similar to Scantron sheets used with multiple-choice tests, for each question, the bubble sheet provided the question number with three circles labeled a, b, and c, and participants were asked to mark the corresponding circle. The manipulation of our conditions pertained to the bubble sheet and to what participants did with it after transferring their answers.

In the control condition, participants received a standard bubble sheet. When they finished transferring their answers, they handed both the test and the bubble sheet to the experimenter, who checked their answers, summed up the number of correct answers, and paid the participants 10¢ for each correct answer. In the no-recycle condition (first cheating condition), the bubble sheet had the correct answers premarked, such that the circles representing the correct answers were shaded in gray. This design prompted a dilemma for participants when they faced a

question they had answered incorrectly on their test sheet: they could be honest and fill in the corresponding incorrect bubble or be dishonest and fill in the correct bubble. After participants finished transferring their answers, they summed up the number of their correct answers, wrote that number at the top of the bubble sheet, and handed both the test and the bubble sheet to the experimenter, who paid them according to their self-summed score. In this condition, subjects could cheat with some risk that the experimenter might discover it, if she compared the answers on the bubble sheet to the answers on the test sheet. The recycle condition (second cheating condition) was similar to the no-recycle condition, with the difference that participants were instructed to transfer their answers to the premarked bubble sheet and then walk to a shredder, shred their original test sheet, and take only the bubble sheet to the experimenter, at which point they would be paid accordingly. Because of the shredding, this condition offered a lower probability of being caught cheating than the no-recycle condition. Finally, the recycle+ condition (third cheating condition) further decreased the probability of being caught by instructing participants to shred both their test sheet and the bubble sheet, walk over to a large jar with money at the corner of the room, and take the amount they earned. In addition, by making the payment "self-service," the recycle+ condition eliminated any interactions with the experimenter, thereby decreasing social concerns with cheating[3]. At the start of each experimental session of the recycle+ condition, the jar was filled with different denominations that totaled $100. After each session (out of the sight of students), we collected the jar and measured the amount of money in it[4].

Results and Discussion

On average, participants in the control condition solved 32.6 questions, while those in the no-recycle, recycle, and recycle+ conditions solved 36.2, 35.9, and 36.1 questions, respectively.

An overall ANOVA of the number of questions reported as solved revealed a highly significant effect of the conditions ($F(3,322) = 19.99$, $p < .001$). The average reported performance in the three cheating conditions was significantly higher than in the control condition ($F(1,322) = 56.19$, $p < .001$), but there was no difference in dishonesty across the three cheating conditions ($F(2,209) = .11$, $p = .9$), and the average magnitude of dishonesty was approximately 20% of the possible average magnitude, which was far from the maximal possible dishonesty in these conditions (similar to findings by Goldstone and Chin 1993). These latter results suggest that participants in all three cheating conditions seemed to have used the same threshold to reconcile the motivations to benefit financially from cheating and maintain their positive self-concept.

Experiment 6 is also useful in testing one other possible alternative explanation, which is that the increased level of cheating we observed in the three cheating conditions was due to a "few bad apples" (a few people who cheated a lot) rather than due to a general shift in the number of answers reported as correctly solved (many people cheating just by a little bit). As can be seen in Figure 3 however, the dishonesty seemed to be due to a general increase in the number of "correct responses," which resulted in a rightward shift of the response distribution[5]. To test this stochastic dominance assumption, we subjected the distributions to a series of quantile regressions and found that the cheating distributions dominated the control distribution at every possible point (e.g., at the 10th, 20th, 30th, 40th, 50th, 60th, 70, 80th, and 90th percentiles, the number of questions solved was significantly higher in the cheating conditions than in the control condition: $t(210) = 3.65$, 3.88, 4.48, 4.10, 2.92, 3.08, 2.11, 2.65, and 3.63, ps $< .05$), but the distributions across the cheating conditions did not differ from each other (no ps $< .35$).

While Experiment 6 was particularly useful for this analysis (because it included multiple cheating conditions), a stronger test would be to see whether this conclusion also holds across all our six experiments. To do so, we converted the performance across all the experiments to proportional, that is, the number of questions reported solved relative to the maximum possible. Analyzing all conditions across our experiments (n=1408), we find again strict stochastic dominance of the performance distributions in conditions that allowed cheating over conditions that did not (β =.15, $t(1406) = 2.98$, $p = .003$). We obtain similarly reliable differences for each quantile of the distributions, suggesting that the overall mean difference (β =.134, $t(1406) = 9.72$, $p<.001$) was indeed caused by a general shift in the distribution rather than a large shift of a small portion of the distribution.

••• Figure 3 •••

GENERAL DISCUSSION

People in almost every society value honesty and maintain very high beliefs about their own morality, yet examples of significant dishonesty can be found everywhere in the marketplace. The standard cost-benefit model, which is central to legal theory surrounding crime and punishment, assumes that dishonest actions are performed by purely selfish, calculating people, who only care about external rewards. The psychological perspective, in contrast, assumes that people largely care about internal rewards because they want, for example, to maintain their self-concept. On the basis of these two extreme starting points, we proposed and tested a theory of self-concept maintenance that considers the motivation from both external and internal rewards. According to this theory, people who think highly of themselves in terms of honesty make use of

various mechanisms that allow them to engage in limited amount of dishonesty while retaining their positive views of themselves. In other words, there is a band of acceptable dishonesty which is limited by internal reward considerations. We focus in particular on two related but psychologically distinct mechanisms that influence the size of this band: categorization and attention to standards, which we argue have a wide set of important applications in the marketplace.

Across a set of six experiments we found support for our theory by demonstrating that when people had the ability to cheat, they cheated, but the per person magnitude of dishonesty was relatively low (relative to the possible maximum amount). We also found that people were generally insensitive to the expected external costs and benefits associated with the dishonest acts, but they were very sensitive to contextual manipulations related to the self-concept. In particular, the level of dishonesty dropped when people paid more attention to honesty-standards and climbed with increased categorization malleability (Dana, Weber, and Kuang 2005).

Some of the results provided more direct evidence for the self-concept maintenance mechanism (Experiment 4) by showing that even though our participants knew they were overclaiming, their actions did not affect their self-concept in terms of honesty. It is also interesting to note that predictors, in contrast, expected dishonest actions to have a negative effect on the self-concept. This misunderstanding of the workings of the self-concept also manifested in respondents' inability to predict the effects of moral reminders (Ten Commandments and honor code) and mediums (tokens)[6] – suggesting that people generally expect others to behave in line with the standard economic perspective of an external cost-benefit trade-off and are unappreciative of the regulative effectiveness of the self-concept.

The theory we propose can in principle be incorporated into economic models. Some formalization related to it appears in recent economic theories of utility maximization based on models of self-signaling (Bodner and Prelec 2001) and identity (Bénabou and Tirole 2004, 2006). These models can be adopted to account for self-concept maintenance by incorporating attention to personal standards for honesty (meta-utility function and salience parameter s1, respectively) and categorization malleability (interpretation function and probability 1-λ, respectively). These approaches convey a slowly spreading conviction among economists that to study moral and social norms, altruism, reciprocity, or antisocial behavior, we must understand the underlying psychological motivations that vary endogenously with the environment (see also Gneezy 2005). The data presented herein offer further guidance on the development of such models. In our minds, the interplay between these formal models and the empirical evidence we provide represents a fruitful and promising research direction.

Some insights regarding the functional from in which the external and internal rewards work together emerge from the data, and these findings could also provide interesting paths for further investigations in both economics and psychology. For example, the results of Experiment 2 showed that increasing external rewards in the form of increasing benefits (monetary incentive) decreased the level of dishonesty (though insignificantly). This observation matches findings from another matrix experiment in which we manipulated two factors between 234 participants: the ability to cheat (control and recycle) and the amount of payment to each participant per correctly solved matrix (10¢, 50¢, $2.50, and $5). In this 2 × 4 design, we found limited dishonesty in the 10¢ and 50¢ conditions but no dishonesty in the $2.50 and $5 conditions. Furthermore, the magnitude of dishonesty was approximately the same for 10¢ and 50¢. Together these observa-

tions raise the possibility of a step function like relationship: constant, limited amount of dishonesty up to a certain level of positive external rewards; increasing the external rewards beyond that level could limit categorization malleability, leaving no room for ”under-the-radar” dishonesty. In this way, dishonesty may actually decrease with external rewards.

Finally, it is worthwhile pointing to some of the limitations of our results. The first limitation relates directly to the relationship between external and internal rewards. Arguably, at some point at which the external rewards become very high, they should tempt the person sufficiently to prevail (because the external reward of being dishonest is much larger than the internal reward of maintaining a positive self-concept). From that point on, we predict that behavior would be largely influenced by external rewards as the standard economic perspective would predict (i.e., ultimately the magnitude of dishonesty would increase with increasing, high external rewards).

Another limitation relates to the fact that our results did not support a sensitivity to others' reported behaviors, implying that, for example, self-esteem or norm compliance considerations do not have an influence on individuals' decisions about being dishonest. We do not imply that such effects are not prevalent or perhaps even powerful in the marketplace. It could be, for example, that the sensitivity to others operates slowly towards changing one's global internal standards for honesty, instead of having a large influence on the local instances of dishonesty – such as the ones that took place within our experiments.

From a practical perspective, one of the two main questions about “under-the-radar” dishonesty pertains to its magnitude in the economy. By its very nature, the level of dishonesty in the marketplace is difficult to measure, but if we take our studies as an indication, it may far ex-

ceed the magnitude of dishonesty committed by "standard run of the mill" criminals, that only consider the external rewards in their decision. Across the six experiments (excluding the recycle+token condition), among the 791 participants who could cheat, we encountered only 5 (0.6%), who cheated by the maximal amount (and thus presumably engaged in external cost-benefit trade-off analysis -- leading to standard rational dishonesty), whereas a large majority cheated only slightly (and thus presumably engaged in a trade-off of external and internal rewards leading them to engage in limited dishonesty that flies under the self-concept radar). Furthermore, the total costs we incurred due to the limited dishonesty were much greater than those associated with the maximal dishonesty. Taken at face value, these results suggest that the effort that society at large applies to deterring dishonesty—especially standard rational dishonesty—might be misplaced.

Another important applied speculation involves the medium experiment. As society moves away from cash, and electronic exchanges become more prevalent, mediums are rapidly more available in the economy. Again, if we take our results at face value, we should pay particular attention to dishonesty in these new mediums (e.g., backdating stocks), because they provide more opportunities for under-the-radar dishonesty. In addition, we observed that the medium experiment did not only allow people to cheat more, but it also increased the level of maximal cheating. In the medium experiment we observed 24 participants who cheated maximally, which indicated that the tokens not only allowed people to elevate their acceptable magnitude of dishonesty but also liberated some participants from the shackles of their morality altogether.

When we consider the applied implications of these results, we must emphasize that our findings stem from experiments not with criminals but with students at elite universities, people who will likely play important roles in the advancement of this country and who are a lot like us and those we know. The prevalence of dishonesty among these people and the finding that on an individual level, people were mostly honest rather than completely dishonest suggests to the generalizability of our results. As Goldstone and Chin (1993) concluded, people seem to be moral relativists in their everyday lives.

Naturally the next question, from a practical perspective, relates to approaches for curbing under-the-radar dishonesty. The results of the honor code, Ten Commandments, and token manipulations are promising, because they suggest that increasing people's attention to their own standards for honesty and decreasing the categorization malleability could be effective remedies. However, the means by which to incorporate such manipulations into everyday scenarios in which people might be tempted to be dishonest (e.g., returning used clothes, filling out tax returns or insurance claims), determining how abstract or concrete these manipulations must be in order to be effective (see Hayes and Dunning 1997), and discovering methods for fighting adaptation to these manipulations remain interesting and open questions.

REFERENCES

Accenture (2003), "One-Fourth of Americans Say it's Acceptable to Defraud Insurance Companies," February 12, (accessed December 1, 2006), [available at http://www.accenture.com/xd/xd.asp?it=enweb&xd=_dyn%5Cdynamicpressrelease_577.xml].

Allingham, Michael G. and Agnar Sandmo (1972), "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1, 323-338.

Aronson, Elliot (1969), "A Theory of Cognitive Dissonance: A Current Perspective," in *Advances in Experimental Social Psychology*, Vol. 4, Leonard Berkowitz, ed. New York: Academic Press, 1-34.

——— and J. Merrill Carlsmith (1962), "Performance Expectancy as a Determinant of Actual Performance," *Journal of Abnormal and Social Psychology*, 65 (3), 178-182.

Association of Certified Fraud Examiners (2006). "2006 ACFE Report to the Nation on Occupational Fraud & Abuse," [available at http://www.acfe.com/documents/2006-rttn.pdf].

Bateson, Melissa, Daniel Nettle, and Gilbert Roberts (2006), "Cues of Being Watched Enhance Cooperation in a Real-World Setting," *Biology Letters*, 2 (June), 412-414.

Baumeister, Roy F. (1998), "The Self," in *Handbook of Social Psychology,* Daniel T. Gilbert, Susan T. Fiske, and Gardner Lindzey, eds. New York: McGraw-Hill, 680-740.

Becker, Gary S. (1968), "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76 (2), 169-217.

Bem, Daryl J. (1972), "Self-Perception Theory," in *Advances in Experimental Social Psychology*, Vol. 6, Leonard Berkowitz, ed. New York: Academic Press, 1-62.

Bénabou, Roland and Jean Tirole (2004), "Willpower and Personal Rules," *Journal of Political Economy*, 112 (4), 848-886.

——— and ——— (2006), "Identity, Dignity and Taboos," Working Paper, Department of Economics and Woodrow Wilson School, Princeton University, December.

Bering, Jesse M., Katrina McLeod, and Todd K. Shackelford (2005), "Reasoning about Dead Agents Reveals Possible Adaptive Trends," *Human Nature*, 16 (4), 360-381.

Bodner, Ronit and Drazen Prelec (2001), "Self-Signaling and Diagnostic Utility in Everyday Decision Making," Working Paper, MIT Sloan School of Management, June.

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

Cross, Patricia K. (1977), "Not Can but Will College Teaching be Improved?" in *Reviewing and Evaluating Teaching. New Directions in Higher Education*, No. 17, J. A. Centra, ed. San Francisco: Jossey-Bass, 1-15.

Dana, Jason, Roberto A. Weber, and Jason Xi Kuang (2005), "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness," Working Paper, Department of Psychology, University of Illinois Urbana-Champaign.

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 (August 27), 1254-1258.

Dickerson, Chris A., Ruth Thibodeau, Elliot Aronson, and Dayna Miller (1992), "Using Cognitive Dissonance to Encourage Water Conservation," *Journal of Applied Social Psychology*, 22 (11), 841-854.

Diener, Edward and Marc Wallbom (1976), "Effects of Self-Awareness on Antinormative Behavior," *Journal of Research in Personality*, 10 (1), 107-111.

Duval, Thomas S. and Robert A. Wicklund (1972), *A Theory of Objective Self Awareness.* New York: Academic Press.

Fischhoff, Baruch and Ruth Beyth (1975), "I Know It Would Happen: Remembered Probabilities of Once-Future Things," *Organizational Behavior and Human Performance*, 13, 1-16.

Gilovich, Thomas (1991), *How We Know It Isn't So?* New York: The Free Press.

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

Goldstone, Robert L. and Calvin Chin (1993), "Dishonesty in Self-Report of Copies Made: Moral Relativity and the Copy Machine," *Basic and Applied Social Psychology,* 14 (1), 19-32.

Griffin, Dale W. and Lee Ross (1991), "Subjective Construal, Social Inference, and Human Misunderstanding," in *Advances in Experimental Social Psychology*, Vol. 24, Mark P. Zanna, ed. New York: Academic Press, 319-359.

Greenwald, Anthony G. (1980), "The Totalitarian Ego. Fabrication and Revision of Personal History," *American Psychologist*, 35 (7), 603-618.

Gur, Ruben C. and Harold A. Sackeim (1979), "Self-Deception: A Concept in Search of a Phenomenon," *Journal of Personality and Social Psychology*, 37 (2), 147-169.

Haley, Kevin. J. and Daniel M. T. Fessler (2005), "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game," *Evolution and Human Behavior*, 26 (3), 245-256.

Harris, Sandra L., Paul H. Mussen, and Eldred Rutherford (1976), "Some Cognitive, Behavioral, and Personality Correlates of Maturity of Moral Judgment," *Journal of Genetic Psychology*, 128 (1), 123-135.

Hayes, Andrew F. and David Dunning (1997), "Trait Ambiguity and Construal Processes: Implications for Self-Peer Agreement in Personality Judgment," *Journal of Personality and Social Psychology*, 72 (3), 664-677.

Hechter, Michael (1990), "The Attainment of Solidarity in Intentional Communities," *Rationality and Society*, 2 (2), 142-155.

Henrich, Joseph, Robert Boyd, Samuel 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," *The Wall Street Journal*, (March 30), D1.

Hsee, Christopher K., Fang Yu, Jiao Zhang, and Yan Zhang (2003), "Medium Maximization," *Journal of Consumer Research*, 30 (1), 1-14.

John, Oliver P. and Richard W. Robins (1994), "Accuracy and Bias in Self-Perception: Individual Differences in Self-Enhancement and the Role of Narcissism," *Journal of Personality and Social Psychology*, 66 (1), 206-219.

Josephson Institute of Ethics (2006, October 15). *Report card on the ethics of American youth*. Available at http://www.josephsoninstitute.org/reportcard/.

Kallgren, Carl. A., Robert B. Cialdini, and Raymond R. Reno (2000), "A Focus Theory of Normative Conduct: When Norms Do and Do not Affect Behavior," *Personality and Social Psychology Bulletin*, 26 (8), 1002-1012.

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.

Kunda, Ziva (1990), "The Case for Motivated Reasoning," *Psychological Bulletin*, 108 (3), 480-498.

Langer, Ellen J. (1989), "Minding Matters: The Consequences of Mindlessness-Mindfulness," in *Advances in Experimental Social Psychology,* Leonard Berkowitz, ed. San Diego, CA: Academic Press, 137-173.

Lewicki, Roy J. (1984), "Lying and Deception: A Behavioral Model," in *Negotiation in Organizations,* Max H. Bazerman and Roy J. Lewicki, eds. Beverly Hills, CA: Sage Publications, 68-90.

McCabe, Donald L. and Linda Klebe Trevino (1993), "Academic Dishonesty: Honor Codes and Other Contextual Influences" *Journal of Higher Education*, 64 (5), 522-538.

——— and ——— (1997), "Individual and Contextual Influences on Academic Dishonesty: A Multicampus Investigation," *Research in Higher Education*, 38 (3), 379-396.

——— and ——— and Kenneth D. Butterfield (2002), "Honor Codes and Other Contextual Influences on Academic Integrity: A Replication and Extension to Modified Honor Code Settings" *Research in higher Education*, 43 (3), 357-378.

Nagin, Daniel S. and Greg Pogarsky (2003), "An Experimental Investigation of Deterrence: Cheating, Self-Serving Bias, and Impulsivity. *Criminology*, 41 (1), 167-194.

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), 815-826.

Piaget, Jean (1950), *The Psychology of Intelligence*. New York: Harcourt Brace & Co.

Pina e Cunha, Miguel and Carlos Cabral-Cardoso (2006), "Shades of Gray: A Liminal Interpretation of Organizational Legality-Illegality," *International Public Management Journal*, 9 (3), 2099-225.

Reno, Raymond R., Robert B. Cialdini, and Carl A. Kallgren (1993), "The Transsituational Influence of Social Norms," *Journal of Personality and Social Psychology*, 64 (11), 104-112.

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 (July, 18), 395-405.

Sanitioso, Rasyid, Ziva Kunda, and Geoffrey T. Fong, (1990), "Motivated Recruitment of Autobiographical Memories," *Journal of Personality and Social Psychology*, 59 (2), 229-241.

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.

Shariff, Azim F. and Ara Norenzayan (2007), "God Is Watching You: Supernatural Agent Concepts Increase Prosocial Behavior in an Anonymous Economic Game," Working paper, Department of Psychology, University of British Columbia, January.

Speights, David and Mark Hilinski (2005), "Return Fraud and Abuse: How to Protect Profits," *Retailing Issues Letter*, 17 (1), 1-6.

Sullivan, Harry S. (1953), *The Interpersonal Theory of Psychiatry*. New York: Norton.

Svenson, Ola (1981), "Are We All Less Risky and More Skillful than our Fellow Drivers?" *Acta Psychologica*, 47 (2), 143-148.

Tesser, Abraham, Murray Millar, and Janet Moore (1988), "Some Affective Consequences of Social Comparison and Reflection Processes: The Pain and Pleasure of Being Close," *Journal of Personality and Social Psychology*, 54 (1), 49-61.

Trivers, Robert (2000), "The Elements of a Scientific Theory of Self-Deception," in *Evolutionary Perspectives on Human Reproductive Behavior,* Dori LeCroy and Peter Moller, eds. New York: New York Academy of Sciences, 114-131.

FOOTNOTES

1. Our self-concept maintenance theory is based on how people define honesty and dishonesty for themselves, regardless of whether their definition matches the objective definition or not.
2. We replicated these findings in two other prediction tasks (within and between subjects). Students anticipated a significant deterioration in their own self-concept if they (not another hypothetical student) were to overclaim by two matrices.
3. In a separate study we asked participants to estimate the probability of being caught across the different conditions and found that these conditions were indeed perceived in the appropriate order of the likelihood of being caught (i.e. no-recycle > recycle > recycle+).
4. The goal of the recycle+ condition was to guarantee participants that their individual actions of taking money from the jar would not be observable. Therefore, it was impossible to measure how much money each respondent took in this condition. We could only record the sum of money missing at the end of each session. For the purpose of statistical analysis we assigned the average amount taken per recycle+ session to each participant in that session.
5. This analysis did not include the recycle+ condition, because we were not able to measure individual-level performance and instead were limited to measuring performance per session.
6. Note that our manipulations in their general form may be thought of as priming. In that sense, our results may generalize to a much larger class of manipulations that would curtail cheating behavior and may be useful when, for example, the 10 Commandments or honor codes might not be a feasible solution, such as purchasing environments.

TABLES

Table 1: Number of matrices reported as correctly solved in the first and second matrix task, as well as predicted and actual self-reported measures of absolute honesty and relative morality in the personality test after the control and recycle conditions, respectively, of the first matrix task.

1st Matrix Task Condition	Matrix Task		Personality Test			
	Matrices Solved (0 – 20)		Absolute Honesty (0 – 100)		Relative Morality (-5 – +5)	
	1st Task	2nd Task	Predicted	Actual	Predicted	Actual
Control	4.2	4.6	67.6	85.2	0.4	0.4
Recycle	6.7	4.3	32.4	79.3	-1.4	0.6

FIGURES

Figure 1: A sample matrix of the adding-to-10 task.

1.69	1.82	2.91
4.67	4.81	3.05
5.82	5.06	4.28
6.36	5.19	4.57

Figure 2: Experiment 2: Mean number of "solved" matrices in the control condition (no ability to cheat), the recycle and reycle+honor code (HC) conditions (ability to cheat). The payment scheme was either $0.50 or $2 per correct answer. Error bars are based on standard errors of the means.

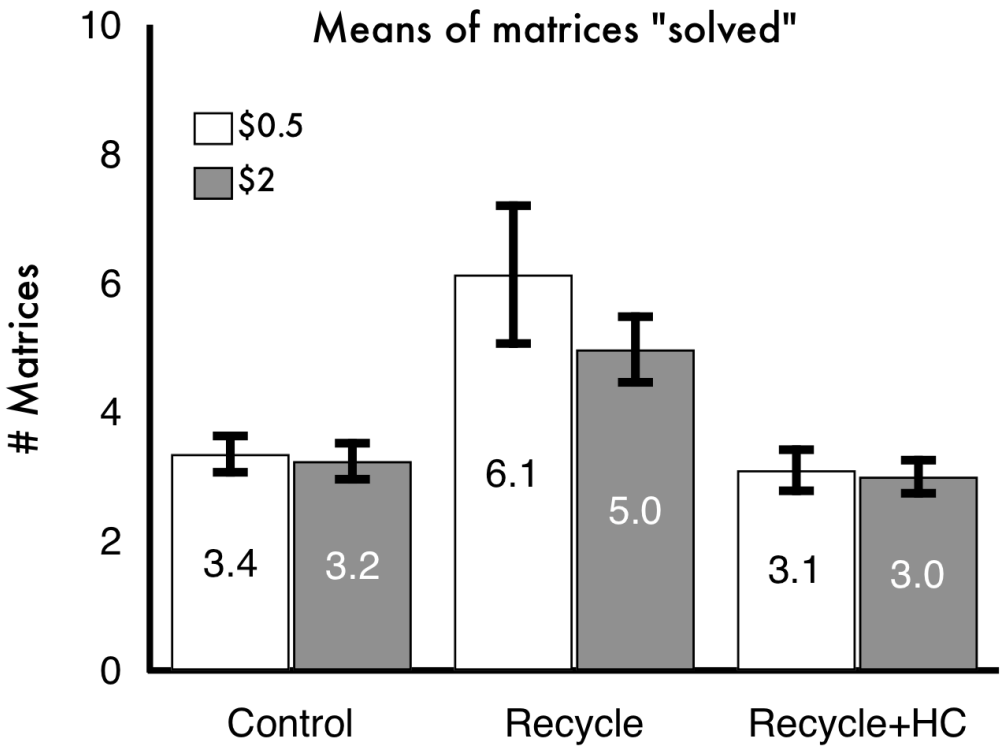

Figure 3: Experiment 6: Frequency distribution of number of "solved" questions in the control condition (no ability to cheat) and two cheating conditions: no-recycle and recycle. The values on the y-axis represent the percentage of participants having "solved" a particular number of questions; the values on the x-axis represent ±1 ranges around the displayed number (e.g., 21 = participants having solved 20, 21, or 22 questions).