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

Dishonesty plays a significant role in the economy. Here, we investigate how external and internal rewards work in concert to produce (dis)honesty. The proposed theory of self-concept maintenance posits that people typically engage in dishonest behaviors and achieve external benefits from dishonesty, but only to the extent that their dishonest acts allow them to maintain a positive view of themselves in terms of being honest. We focus on two mechanisms that people employ to maintain their positive self-concept: categorization and attention to standards. The results show that (1) given the opportunity, people will engage in dishonest behaviors; (2) increasing attention to internal honesty standards decreases the tendency for dishonesty; (3) allowing more flexible categorization increases the tendency for dishonesty; (4) the magnitude of dishonesty is largely insensitive to either the expected external benefits or costs associated with dishonest acts; and (5) people know that their actions are dishonest but do not update their self-concepts. We suggest that dishonesty governed by self-concept maintenance is likely to be prevalent in the economy, and understanding it has important implications for designing effective methods for curbing dishonesty.

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. Names such as Enron and WorldCom illustrate the eroding ethics in the accounting and auditing professions, the costs of which have been estimated at $37–$42 billion of the U.S. gross domestic product in the first year alone (Graham, Litan, and Sukhtankar 2002). In addition, it would be naïve to assume that dishonest behavior is limited to corporations and that it is not widely practiced by individual consumers. To give but 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). And if this evidence is not disturbing enough, perhaps the largest contribution to consumer 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 (Joyner 2002). In addition to these examples of dishonesty in the marketplace, the recent events surrounding Dr. Woo Suk Hwang and his fraudulent reports of cloning human embryos (Cyranoski 2006) reminds us of the possible range and magnitude of dishonest behaviors in the scientific community (Martinson, Anderson, and de Vries 2005).

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 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. On the basis of these inputs, people engage in a cost–benefit analysis in which they carefully weigh the advantages and disadvantages and 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.

Based on this standard External cost-benefit perspective, and as depicted in the next three hypotheses, there are three main forces that are expected to influence the frequency and magnitude of dishonesty:

EH1: Dishonesty will increase as the expected magnitude of reward from the dishonest act increases.

EH2: Dishonesty will increase as the expected probability of being caught in the dishonest act is reduced.

EH3: Dishonesty will increase as the expected magnitude of punishment for performing the dishonest act is decreased.

From a psychological perspective, and in addition to financial considerations, another set of important inputs to the decision whether to be honest 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 his or her behavior. Compliance with the internal reward 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 ability to modify or not modifying 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). 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 **(**Bénabou and Tirole 2006**)**. 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).

If we return to our gas station example, this perspective suggests that people who pass by a gas station will be influenced by not only 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 conceivably could 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 symbolism, its context, and its plasticity. In the next section we characterize these elements in a theory of self-concept maintenance (for related perspectives on self-signaling and identity see Bodner and Prelec 2001 and Bénabou and Tirole 2004, 2006), and test the implications of this theory in a set of experiments. In general,

we find that given the opportunity, people will engage in dishonest behaviors, that the magnitude of dishonesty is limited, and largely insensitive to the expected external benefits and costs associated with dishonest acts, and most interestingly, that the magnitude of dishonesty is very sensitive to manipulations related to the self-concept – helping individuals to be dishonest without updating their self-concept.

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 gain, for example, 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 a sacrifice of the other.

In this article, 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 range of dishonest actions 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, in the current work we focus on two particular means, categorization and attention to standards. Using these mechanisms individuals are able to record their actions (e.g., “I am overclaiming 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 successfully avoid the moral implications of their own behaviors (Gur and Sackeim 1979). When this mechanism is activated, people can categorize their actions in more compatible terms, find rationalizations for their actions, and ultimately avoid triggering any negative self-signals that might affect their self-concept, which will therefore not get updated.

Two important aspects of categorization are its relative ease and its limit. Easily categorized behaviors are ones that allow people to reinterpret them in a self-serving manner, and the exact ease or difficulty 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., he took a pencil from me once; this is what friends do). This thought experiment suggest that categorization can facilitate dishonesty (taking a pencil), but also that some actions are inherently more difficult to categorize successfully in compatible terms, in which contexts honesty is likely to prevail (Dana, Weber, and Kuang 2005; for a discussion of the idea that a medium can disguise the final outcome of an action, see Hsee et al. 2003). In other words, as the "degrees of freedom" in the categorization increase, so does the magnitude of dishonesty a person can commit without influencing his or her

self-concept (Baumeister 1998), because it becomes easier to categorize these behaviors in compatible terms (Schweitzer and Hsee 2002; see also Bénabou and Tirole 2004; Pina e Cunha and Cabral-Cardoso 2006).

The second important 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 measured relative to their standards, and therefore, their self-concept is less likely to be updated, and 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 band for actions that do not trigger updating of the honesty 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., level of plasticity and 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 the later on the inside world. However, they both involve attention, are sensitive to manipulations, and relate to the dynamics of acceptable boundaries of behavior. From this perspective, they are very related.

Thus, while the fuzziness that both 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 distinguish clearly whether the source of the fuzziness comes from the environment (categorization) or the individual (attention to standards). Regardless, the flexibility of the individual and the environment both influence the threshold of beneficial dishonesty, up to which individuals may not need to pay the price of updating their self-concept as less honest individuals. In sum, the theory of self-concept maintenance suggests the following hypotheses:

PH1: Dishonesty will increase as individuals pay less attention to their own standards for honesty.

PH2: Dishonesty will increase when individuals face situations that are more easily categorized in honesty-compatible terms.

PH3: Given the opportunity to be dishonest, individuals will be dishonest up to a level that does not force them to update their self-concept.

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

With Experiment 1, we tested our prediction that increasing people's attention to their standards for honesty will make them more honest. The general setup of all our experiments involved a multiple-question task, for 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 the "cheating" conditions, in which they had such an opportunity. Specifically, in Experiment 1, we contrasted 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 rather whether he or she thinks of those standards and compares his or her behavior to the standards at the moment of temptation to behave dishonestly. In other

words, if a mere reminder of the standards for honesty has an effect we may assert that one didn't naturally attend to those standards otherwise. If awareness at the moment of temptation determines dishonesty, reminders should be highly effective in reducing dishonesty. In Experiment 1we 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). In the second, ostensibly separate task, each student received 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. 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), at the end of the four minutes, participants handed both the test and the 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 (after the 10 books and Ten Commandments recall task), participants indicated the 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 putting the test sheet away did they hand the answer sheet to the experimenter. The entire experiment represents a 2 (type of reminder) × 2 (ability to cheat) between-subjects design.

Results and Discussion

The results of Experiment 1 confirm our predictions. The type of reminder had no effect on participants' performance in the two control conditions ($\text{mean}_{\text{10 books}} = 3.1$ vs. $\text{mean}_{\text{Ten Commandments}} = 3.1$; $F(1,225) = .012$, $p = .91$), which suggest that the type of reminder did not influence ability or motivation. Following the book recall task, respondents cheated when they were given the opportunity to do so ($\text{mean}_{\text{10 books}} = 4.2$) but not so after the Ten Commandments recall ($\text{mean}_{\text{Ten Commandments}} = 2.8$; $F(1,225) = 5.24$, $p = .023$), and the interaction between the two factors (type of reminder and ability to cheat) was significant ($F(3, 225) = 4.52$, $p = .036$).

Interestingly, the level of cheating remained much below the maximum (i.e., 20). In fact, participants cheated on average "only" 6.7% of the possible magnitude. Most important, and in line with the self-concept maintenance idea, reminding participants of standards for morality eliminated cheating completely in the Ten Commandments/recycle condition: participants "solved" as many matrices as in the control condition ($F(1,225) = .49$, $p = .48$).

We designed Experiment 1 to focus on the attention to standards mechanism (PH1), but one aspect of the results—the finding that the magnitude of dishonesty was limited and well below the maximum possible level—suggested that the categorization mechanism (PH2) might have been at work as well.

One possible alternative interpretation of the book recall with opportunity to cheat (books/recycle) condition is that participants already had developed over their lifetime standards for moral behavior, according to which over-claiming correct answers 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 over-claims on taxes, the very act of over-claiming is generally accepted and therefore not necessarily considered immoral. However, if this account were the reason for our findings, increasing people's attention to morality would not have decreased the magnitude of dishonest behavior. 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 ($s_n = .31$) 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 Ten Commandments remembered as a proxy for religiosity, the lack of relationship between religiosity and amount of dishonesty suggests that the efficacy of the Ten Commandments lied in increased attention to 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 situations like these. A separate set of students predicted we would find; they 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 (mean = 9.5 vs. 3.1), and they anticipated that reminding them of the Ten Commandments would not decrease cheating (mean = 7.8, $t(73)=1.61$, $p = .11$). The contrast of the predicted results with the actual behavior suggests that participants understand the economic motivation for over-claiming, that they over-estimate 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 we tested, an honor code, refers to a mechanism 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 this method used by many academic institutions (McCabe, Trevino, and Butterfield 2002), our self-concept maintenance account may

be a concise candidate sufficient to explain its effectiveness. If an honor code statement indeed produces results similar to the Ten Commandments manipulation we may shed light on the internal process underlying its success (see http://www.academicintegrity.org). In addition to manipulating the awareness of honesty standards at the point of temptation, Experiment 2 extended the results of the previous experiment by manipulating the financial incentives for performance (i.e., external benefits), and by doing so also tested EH1.

Method

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

The control and recycle conditions were identical to those in the previous experiment, except this time, the experimenter paid each participant, and the task lasted five minutes. The recycle+honor code condition was similar to the recycle condition except that we asked respondents to sign a statement appearing at the top of the test sheet that read: “I understand that this short survey falls under MIT’s [Yale’s] honor system”; below the statement, participants printed and signed their names. 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 another test for EH1, we manipulated the payment per correct matrix solved ($.5 vs $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 = 5.5$) relative to those in the two control conditions (50¢ and $2: $M = 3.3$; $F(1,201) = 15.99$, $p < .001$), 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 EH1. One possible interpretation of this decrease in dishonesty with increased incentives is that the level 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-flexibility among participants was affected by a mix of these two factors, we would have expected the number of questions they 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: $M = 3.0$), eliminated cheating to the extent that the performance in these conditions was undistinguishable from the two control conditions (50¢ and $2: $M = 3.3$; $F(1,201) = .19$, $p = .66$) but it was significantly different from the two recycle conditions (50¢ and $2: $M = 5.5$;

$F(1,201) = 19.69$, $p < .001$), even though the two recycle+honor code conditions were procedurally very similar to the two recycle conditions, and objectively presented no implications of external punishment. 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. 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 but the reminder of morality that was at play.

••• Figure 2 •••

Again we asked a separate set of students to predict these results (n = 82), and while they predicted that the increased payment would marginally increase deception (means = 6.8 vs. 6.4, $F(1,80) = 3.3$, p=.07) (in essence predicting EH1), they did not anticipate that the honor code would decrease dishonesty (means = 6.9 vs. 6.2, $F(1,80) = .74$, p=.4). The contrast of the predicted results with the actual behavior suggests that participants understand the economic motivation for over-claiming, that they over-estimate its influence on behavior, and that again they underestimate the effect of the self-concept in regulating honesty.

EXPERIMENT 3: INCREASING CATEGORIZATION-FLEXIBILITY

Making people mindful by increasing their attention to standards for honesty can curb dishonest behavior, but the theory of self-concept maintenance also implies that increasing the flexibility with which people can interpret their actions should increase the magnitude of dishonest behavior (see also Schweitzer and Hsee 2002). To test this hypothesis, in Experiment 3, we manipulated whether the opportunity for deception 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 a more flexible interpretation their actions, make the moral implications of cheating less accessible, and hence make it easier for participants to cheat at higher magnitudes.

Method

Four hundred fifty students participated in this matrix task experiment. Participants had five minutes to complete this task and were promised 50¢ for each correctly solved matrix. We used three between-subjects conditions: the control and recycle conditions that we used 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. When the five minutes for the matrix task ended, participants in the recycle+token condition recycled their test sheet and submitted only their answer sheet to an experimenter, who gave them the corresponding number of tokens. Participants then went to a second experimenter, who exchanged the tokens for money (this experimenter also paid the participants in the control and recycle conditions). We counterbalanced the roles of the two experimenters.

Results and Discussion

Participants in the recycle condition "solved" significantly more questions than participants in the control condition (means = 6.2 vs. 3.5, respectively; $F(1,447) = 34.26$, $p < .001$), which suggested they cheated. As we previously found, 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 (PH2), introducing tokens as the medium for

immediate benefit increased the magnitude of dishonesty relative to the recycle condition (mean = 9.4; $F(1,447) = 47.62$, $p < .001$), presumably without any changes in the probability of being caught or the severity of the punishment.

In line with PH2, making it easier for participants to categorize their behavior in a more self-serving manner by "claiming" more tokens instead of "stealing money" likely reduced the negative self-signal that they otherwise would have received from their dishonest behavior. In terms of our current account, the recycle+token condition increased the threshold for acceptable 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 these results differ from what a separate set of students (n = 59) predicted we would find; they correctly anticipated that participants would cheat when given the opportunity to do so (mean = 6.6), but they anticipated that being able to cheat in terms of tokens would not increase the tendency for cheating (mean = 7, $t(57) = 4.5$, $p = .65$). This again suggests that individuals underestimate the effect of the self-concept in regulating honesty.

EXPERIMENT 4: RECOGNIXZING DISHINESTY BUT NOT UPDATING THE SELF-CONCEPT

Our account of self-concept maintenance suggests that by engaging only in relatively minimal cheating our participants stayed within the threshold of acceptable 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 or attention to standards mechanisms prevented this factual knowledge from being morally evaluated. Thus, people did not necessarily confront the true meaning of their actions (e.g., "I am dishonest"). This suggests that while people know they are overclaiming, they are not attending to the implications of this knowledge. We test these predictions (PH3) 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. These predictions would be higher than 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 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 own honesty right after they completed the first matrix task (in which they either had or did not have a chance to cheat). If participants who cheated had lower opinions about themselves in terms of honesty than those in the control condition, 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 behavior, and consequently, that they have 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—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 of the matrix task 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 how many matrices they had solved correctly (participants in the control condition submitted both the test and the answer sheets to the experimenter, who verified their answers).

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

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 to them that this second matrix task would be identical to that of the control condition. In other words, it was clear to participants that the next matrix task left no room to overclaim as the experimenter would check the results. 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 interests to be as accurate as possible in their predictions 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 (as in the control condition) with a different set of numbers. The entire experiment thus represented a two-condition, between-subjects design, differring only in the first matrix task (possibility to cheat). The three remaining tasks (a personality test, a prediction task, and a second matrix task) were the same for all participants.

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 did 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 = 4.5) did not differ from the control condition's performance on the first task ($t(43) = .65$, $p = .519$). These findings implied that, as in the previous experiments, participants cheated when they had the chance to do so. Moreover, 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 (control: $M = 6.32$, $s_n = 3.29$, recycle: $M = 6.32$, $s_n = 2.63$). Moreover, participants in the control ($M = 72.5$, $s_n = 4.4$) and recycle ($M = 68.8$, $s_n = 4.64$) conditions were equally confident about their predictions ($t(42) = .56$, $p = .57$). 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.

The third finding is presented in Figure 6: 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 level 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-honesty questions (how honest do you consider yourself and how moral are you today compared with yesterday). In the control condition, we asked them to imagine what an average student

who solved four matrices would answer to these two questions. In the recycle condition, we asked them to imagine what an average student who solved four matrices, but claimed to have solved six, would answer to these two questions. As can be seen in Table 1, they predicted that cheating would decrease both a person's general view of him- or herself as an honest person ($t(37) = 3.77$, $p < .001$) and his or her morality compared with the day before the test ($t(37) = 3.88$, $p < .001$). This finding provides further support to 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 of an acceptable band of dishonesty, the size of which depends on the ability to categorize or label actions as something other than dishonest. Moreover, the results of Experiment 4 provide some evidence that deception can take place without an associated change in the self-concept. Overall, these findings are in line with our theory of self-concept maintenance: People are torn between the temptation to benefit from cheating and the benefits of maintaining a positive view about themselves. To solve this dilemma individuals find a balance between these two motivating forces such that they can engage to some level in dishonest behavior but they do so without updating their self-concept.

Although these findings are consistent with a theory of self-concept maintenance, there are also multiple alternative accounts for these results. In the final two experiment, we would like to try and deal with a few of these.

One possible alternative account is that the different manipulations (moral reminders for example) influence 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).

A second alternative account that comes to mind would posit that participants were driven only by self-esteem (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 feel stupid in comparison with everybody else (we used the matrix task partially because it is not a task that our participants relate to IQ, but this account might still be possible).

A third alternative explanation for these 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 if they only cheat by a few questions would be zero (or negligible), and so they cheated in a calculated way up to this threshold – in essence estimating what they could get away with and cheating up to that level.

What these three accounts shave in common is that all of them are sensitive to the behavior (or expected behavior) of others. In contrast, the self-concept maintenance theory we posed suggests 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 level. If the level of cheating is driven by norm compliance, drive toward achievement, or threshed in detecting deception, the number

of matrices that participants claimed to have solved should increase when they believe that the average performance of others is higher. If, however, the level of cheating is driven only by self-concept maintenance considerations, being informed that others solve many more matrices should have no effect on the level of dishonesty.

Method

One hundred eight students participated in the matrix task experiment. We manipulated two factors between participants: the ability to cheat (control and recycle, as in Experiments 2 and 3) and beliefs about the number of matrices that the average student solves in the time allotted (four matrices, which is the accurate number, or eight matrices which was an exaggeration). As before, the DV was the number of matrices reported solved.

Results and Discussion

Participants in the control conditions solved 3.3 and 3.4 matrices on average in the 4 vs. 8 believed standard performance conditions, respectively, while those in the corresponding recycle conditions "solved" 4.5 and 4.8 matrices. A two-factorial ANOVA of the number of matrices solved as a function of the ability to cheat and the belief about other's performance showed a main effect of the ability to cheat ($F(1,104) = 6.89$, $p = .01$) but no main effect of the average levels of performance manipulation ($F(1,104) = .15$, $p = .7$), as well as 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 performance of others. This finding argues against norm compliance, drive toward achievement, or threshed in detecting deception as alternative explanations for our findings.

The fact that our results do not support these social influence factors does not mean that these effects are not prevalent and perhaps even very powerful in the marketplace. What it does suggest is that these social factors might work slowly to change the overall standards for what people consider as honest and not honest in society, but that they do not have a large influence on the more temporary fluctuations in honesty – such as the ones that took place within our experiments.

EXPERIMENT 6: SENSITIVITY TO EXTERNAL REWARDS

Since detection of deceptive acts is central to the standard cost-benefit view of dishonesty, we wanted to test its influence more directly, and do so by varying the probability of being caught on multiple levels. In addition, 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 likelihood of getting caught on three levels and measure the amount of dishonesty across these cheating conditions. If the pure external cost-benefit incentives are at work in our setup, we should find that the level of dishonesty increases as the probability of being caught decreased (EH2). On the other hand, if self-concept maintenance controls 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, small sessions, in which each participant sat in a private booth (N = 326). At the start of each session, the experimenter explained the

instructions for the entire session. The first part of the procedure remained the same for all four conditions, but the second part varied. 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 solve correctly. After the 15 minutes, participants received a "bubble sheet" onto which to transfer their answers. Similar to many multiple-choice tests, for each question, the bubble sheet provided the question number with three circles labled a, b, and c, and participants were asked to mark the corresponding circle. The manipulation of the four conditions pertained to the type of bubble sheet and to what participants had to do 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 the number of correct answers, and paid them 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 mark the corresponding incorrect circle on the bubble sheet or be dishonest and mark the correct circle. 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 test to the bubble sheet. The recycle condition (second cheating condition) was similar to the no-recycle condition, with the difference that participants were instructed, after transferring their answers to the premarked bubble sheet, to 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) decreased the probability of being caught even further 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 sight of students), we collected the jar and measured the amount of money in it.[4]

Results & Discussion

Participants in the control condition solved 32.6 questions correctly, while those in the No-recycle, Recycle, and Recycle+ conditions "solved" 36.2, 35.9, and 26.1 questions respectively. An overall ANOVA of the number of questions reported solved revealed a highly significant effect of the conditions ($F(3,322) = 19.99$, $p < .001$). The reported performance in the conditions where cheating was possible was significantly higher than that 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 only approximately 20% of the possible average magnitude, which was far from the maximal

dishonesty possible 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 recycle conditions was due to a "few bad apples" (a few people who cheated a lot) rather than to a general shift in the number of answers reported as correct across many people. As can be seen in Figure 3, the increased dishonesty, when possible, was not due to a "few bad apples" but rather to a general increase in the number of "correct responses," which resulted in a rightward shift of the response distribution. 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 performance distributions across the three cheating conditions did not differ from each other (no ps $< .35$).[5]

While experiment 5 was particularly useful for this analysis (because it included multiple conditions, and because of the normality of the distribution of questions "solved), it is also useful to test whether this conclusion also holds across all the experiments. To do so, we converted the performance from all experiments to proportional, that is, the number of questions "solved" relative to the maximum possible. Analyzing all tasks across our experiments (n=1408), we find

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 persons 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 of both external and internal rewards (Gneezy 2005). According to this theory, people who think very highly of themselves in terms of honesty make use of various mechanisms that allow them to engage in some level of dishonest behavior while retaining their positive view of themselves. We focus in particular on two related but psychologically distinct mechanisms, categorization and attention to standards, which we argue have a wide set of important applications in the marketplace.

The experiments we presented demonstrated that when people had the ability to cheat, they cheated, but the per person magnitude of their dishonesty was relatively low (at least relative to the possible maximal dishonesty). We also found that people were generally insensitive to the expected external costs and benefits, but they were very sensitive to contextual manipulations. In particular, the level of dishonesty droped when people paid more attention to honesty-standards but climbed when they had more wiggle room for a categorization that was more compatible with an honest self-concept (Dana, Weber, and Kuang 2005).

Some of the results supported the self-concept mechanism more directly (Experiment 5) 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 the lack of updating of the self-concept was not something that our students predicted. Ex ante, they expected dishonest actions to have a negative affect on the self-concept—but that was not the case. This misunderstanding of the workings of the self-concept also manifested in respondents' inability to predict the effects of moral reminders (the Ten Commandments and honor code) and medium effects (tokens) – suggesting that people generally expect others to behave in line with the standard external cost-benefit perspective and are unappreciative of the regulative effectiveness of the self-concept.

Together, these findings support our theory of self-concept maintenance. First, they imply that the band of acceptable dishonesty is limited by internal reward considerations. Second, they suggest that the size of this band depends on the ability to categorize actions as something other than dishonest, as well as the attention that people pay to their standards for honesty at the time of the dishonest act.

The theory we propose can in principle be incorporated into economic models. Some formalization related to our theory appears in recent economic theories of utility maximization based on models of self-signaling (Bodener and Prelec 2001) and identity (Bénabou and Tirole 2004, 2006). These recent 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. These models can be adopted to account for self-concept maintenance by incorporating categorization and attention: increasing attention to personal standards for honesty (meta-utility function and salience parameter s1, respectively) and flexibility for categorization (interpretation function and probability 1-λ, respectively). 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 also could provide useful paths for further investigations in both economics and psychology. For example, the results in Experiment 6 showed that increasing the level of external costs (probability of being caught) did not decrease the level of dishonesty. This finding raises the possibility of a relationship that appears like a step function in which dishonesty up to a certain level is trivial, but beyond that threshold, it takes on a more serious, and costly, meaning.

Another aspect related to the combination of external and internal rewards pertains to the direction of the influence of increased payments in Experiment 2. Increasing the level of external

benefits (monetary incentive) decreased the level of dishonesty (though insignificantly), which implies that, similar to the magnitude of dishonesty, the money involved in dishonest actions could be self-limiting in the sense of decreasing the flexibility that people have to categorize their actions. If so, then the level of this type of dishonesty can be decreasing with external rewards. This hypothesis 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 find cheating in the 10¢ and 50¢ conditions but no cheating for $2.50 and $5 conditions. Furthermore, the amount of cheating is approximately the same for 10¢ and 50¢.

Finally, it is worthwhile pointing to some of the limitations of our results with regard to the relationships 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 reward is much larger than the internal costs), such that ultimately behavior would be largely influenced by external rewards.

From a practical perspective, one of the two main questions about this “under-the-radar” dishonesty pertains to its role and 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 exceed the magnitude of rationally planned dishonesty committed by “standard run of the mill” criminals. Across the six experiments (excluding the 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 planned dishonesty), whereas a large majority cheated

only slightly (and thus presumably engaged in acceptable dishonesty that flies under the self-concept radar). Furthermore, the total costs we incurred due to the moderate 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 planned dishonesty—might be misplaced. That is, the rewards for minimizing small dishonest acts by a larger segment of the population may be greater than those related to attempts to reduce what we consider as "standard" crimes. By analogy, if our goal were to pay an amount that reflected the actual performance level of our participants, we should have tried to reduce low-magnitude dishonesty by the majority rather than focus on the few participants who cheated intensely.

Another important applied speculation involves the medium experiment. As society moves away from cash, and electronic exchanges become more prevalent, mediums are rapidly increasing 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 opportunities for under-the-radar dishonesty. Another interesting observation is 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 likely will play important roles in the advancement of this country and who seem a lot like

us and others we know. The prevalence of dishonesty among these people and the finding that on an individual level, the magnitude of dishonesty was typically somewhat honest rather than completely dishonest suggests that we have tapped into what common, everyday behavior is about. As Goldstone and Chin (1993) conclude, people seem to be moral relativists in their everyday lives.

The natural next question from a practical perspective thus relates to approaches to curbing under-the-radar dishonesty. The results of the honor code, Ten Commandments, and token manipulations are promising, in that they suggest that increasing people's attention to their standards and decreasing the degrees of freedom available to interpret their own actions could be effective remedies. However, the means by which to incorporate such manipulations into scenarios in which people might be tempted to be dishonest (e.g., returning clothes, filling out tax returns, or insurance claims), how abstract or concrete these manipulations must be in order to be effective (see Hayes and Dunning 1997), and how to fight 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.

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.

Cyranoski, David (2006), "Verdict: Hwang's Human Stem Cells Were all Fake," *News@Nature.com*, January 10 (accessed January 10, 2006), DOI: 10.1038/439122a.

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.

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," *The Brookings Institution*, July 22, (accessed July 1, 2006), [available at http://www.brookings.edu-/views/papers/graham/20020722.htm].

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.

Joyner, Tammy (2002), "Corporate Crime not Limited to Bigwigs," *Atlanta Journal-Constitution*, (August 6), A1.

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.

Martinson, Brian C., Melissa S. Anderson, and Raymond de Vries (2005), "Scientists Behaving Badly," *Nature*, 435 (June 9), 737-738.

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 theory of self-concept maintenance is based on the manner in which people define honesty and dishonesty for themselves, regardless of whether their definition matches the objective definition or not.
2. For some contexts, it might be more important to manipulate the severity of the punishment.
3. A separate study in which we asked participants to estimate the probability of being caught across the different conditions verified our expectations that these conditions were indeed perceived in the appropriate order of the likelihood of being caught (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. We replicated these findings in two other prediction tasks (within and between subjects) that referred not to another hypothetical student but to the person answering the survey. We found that students anticipated a significant deterioration in their own self-concept if they were to overclaim by two matrices.

TABLES

Table 1: Number of matrices reported as correctly solved in the first and second task, as well as predicted and actual self reported measures of honesty and relative morality, across the two 1st task conditions.

1st Task Condition	Matrices Solved		Honesty		Relative Morality	
	1st Task	2nd Task	Predicted	Actual	Predicted	Actual
Control	4.2	4.6	67.6	85.2	.4	.4
Recycle	6.7	4.3	32.4	79.3	-1.4	.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.

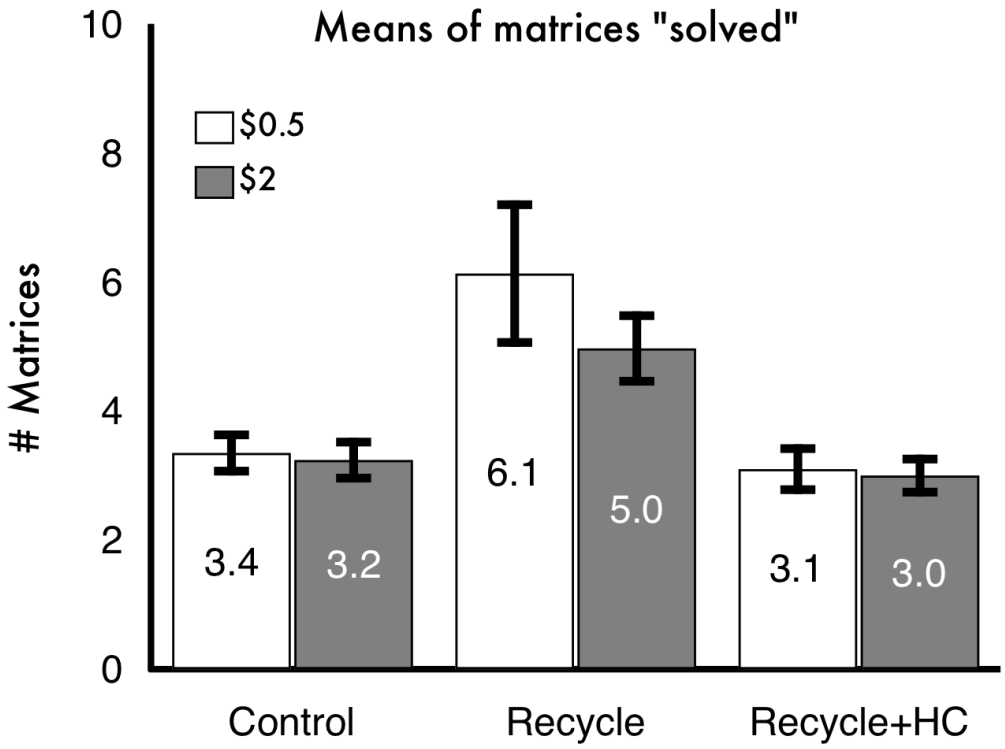
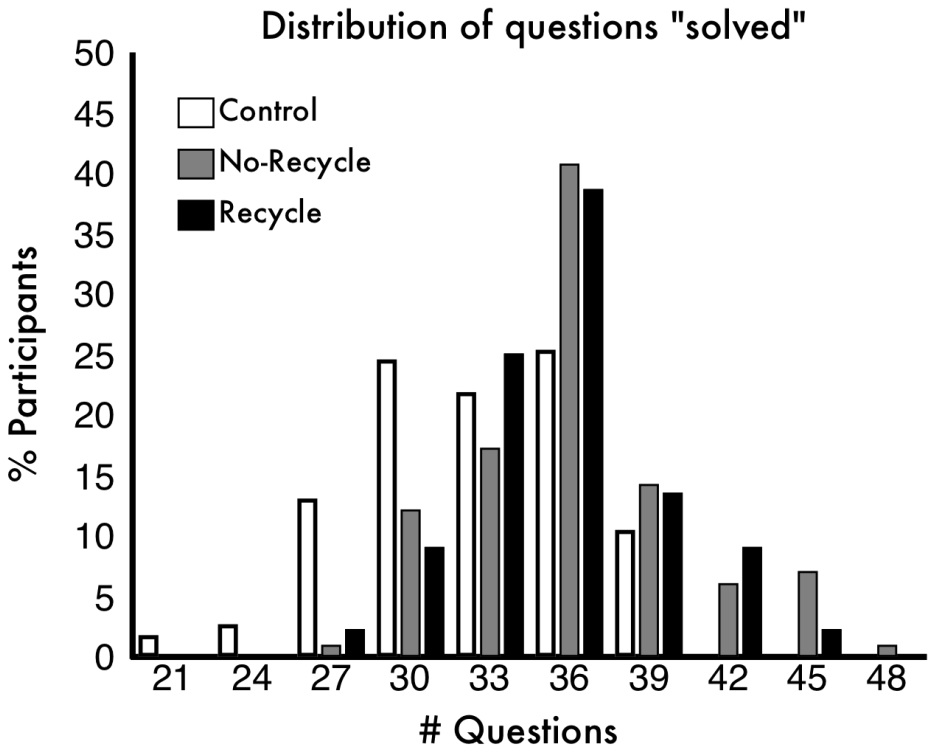
Figure 3: Experiment 6: Frequency distribution of number of "solved" questions in the control condition (no ability to cheat) and two cheating conditions: no-recycle and recycle. The values on the y-axis represent the percentage of participants having "solved" a particular number of questions; the values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