ORIGINAL PAPER

Regulating "Good" People in Subtle Conflicts of Interest Situations

Yuval Feldman[1] · Eliran Halali[2]

Received: 25 January 2016 / Accepted: 9 February 2017 / Published online: 4 March 2017

Abstract Growing recognition in both the psychological and management literature of the concept of "good people" has caused a paradigm shift in our understanding of wrongful behavior: Wrongdoings that were previously assumed to be based on conscious choice—that is, deliberate decisions—are often the product of intuitive processes that prevent people from recognizing the wrongfulness of their behavior. Several leading scholars have dubbed this process as an ethical "blind spot." This study explores the main implications of the good people paradigm on the regulation of employees' conflicts of interest. In two experiments, we examined the efficacy of traditional deterrence- and morality-based interventions in encouraging people to maintain their professional integrity and objectivity at the cost of their own self-interest. Results demonstrate that while the manipulated conflict was likely to "corrupt" people under intuitive/automatic mindset (Experiment 1), explicit/deliberative mechanisms (both deterrence- and morality-based) had a much larger constraining effect overall on participants' judgment than did implicit measures, with no differences between deterrence and morality (Experiment 2). The findings demonstrate how little is needed to compromise the employees' ethical integrity, but they also suggest that a modest explicit/deliberative intervention can easily prevent much of the wrongdoing that may otherwise result.

✉ Yuval Feldman
Yuval.Feldman@biu.ac.il

[1] Kaplan Professor of Legal Research, Bar-Ilan University Law School, Ramat Gan, Israel

[2] Department of Psychology, Bar-Ilan University, Ramat Gan, Israel

Keywords Regulating ethical behavior · Conflict of interest · Deterrence and legitimacy · Behavioral ethics · Unethical behavior

Introduction

Typically, states and organizations have used their sanctioning powers to prevent people who were somehow knowingly engaging in wrongful conduct, such as breaching contracts or otherwise eschewing their duties. By and large[1] the usage of enforcement tools in both law and management assume that in most cases, people consciously choose to engage in unethical behaviors, and that certain techniques, such as communicating through state laws, ethical codes that focus on corporate values (Somers 2001), and incentives (Gneezy et al. 2011; Camerer and Hogarth 1999; Weaver 1995), can be used to change such decisions. A large portion of the intervention mechanisms used by various government agencies, courts, and organizations have relied on these assumptions and created incentive mechanisms, increased enforcement efforts, and added new regulations to enhance transparency (Stapenhurst and Kpundeh 1999). In contrast, many theories of the behavioral approach to human judgment and decision making have challenged the basic assumptions of the neoclassical economic doctrine of rational choice (Feldman 2011). Among these, the literature related to the rising role of non-deliberative choice in people's behavior stands as a central and dominant alternative. A common theme in these paradigms is the view that many of the undesirable behaviors, traditionally the focus of prevention using

[1] We ignore here concepts such as mistakes which are treated by the negligence doctrine.

rational choice mechanisms, have to do with "good people" who do not necessarily engage in a fully deliberative process before performing a "bad" action (Banaji and Greenwald 2013). Therefore, the ability of current explicit mechanisms to curb unethical behavior may be limited and need to be reexamined. Thus, for example, organizational sanctions, which are prevalent in many codes of conduct (Adams et al. 2001; Schwartz 2002), might not stop people from behaving unethically if they do not perceive their action as one which raises an ethical problem.

The Rise of "Good People"

The focus in recent literature on "good people" represents a growing recognition that many ethical decisions are the result of implicit choices, rather than explicit ones, which are made by normative citizens. Simply reviewing the titles of current papers shows how central the theme has become (e.g., Banaji and Greenwald 2013; Bereby-Meyer and Shalvi 2015; Bersoff 1999; Hollis 2008; Mazar et al. 2008; Pillutla 2011; Shalvi et al. 2015). This theme of good people[2] suggests a growing recognition that many ethically relevant behaviors that were previously assumed to be choice-based, conscious, and deliberate decisions are in many cases the product of automatic processes that prevent people from recognizing the wrongfulness of their behavior, an idea dubbed by several leading scholars as an ethical blind spot (e.g., Chugh et al. 2005; Bazerman and Tenbrunsel 2011).

A deeper understanding of good people can be achieved based on the concept of dual reasoning, the assumption of two distinct systems of reasoning. This concept gained popular recognition in Kahneman's book, thinking fast and slow (2011). Generally speaking, the concept, which stands at the core of much of the research in behavioral law and economics, differentiates between an automatic, intuitive, and mostly unconscious process (labeled System 1) and a controlled and deliberative process (labeled System 2) (see also Stanovich and West 2000; Evans 2003; and for review: Evans 2008). Although this paradigm has been criticized by many scholars (e.g., Kruglanski and Gigerenzer 2011), the recognition of the role of automaticity in decision making has played an important part in the emergence of behavioral economics (e.g., Halali et al. 2013; Halali et al. 2014; Sanfey et al. 2003), and behavioral law and economics (e.g., Jolls et al. 1998), and it is the foundation for a new understanding and approach to self-interest (see Gigerenzer and Goldstein 1996).

Good People and Conflict of Interest

Given the growing focus on good people in psychology and management, and the rise of the dual reasoning research, the limited attention to its implications on the enforcement of ethics is puzzling (see Feldman 2014 for a review). Motivated reasoning is the main theoretical paradigm that supports the view that individuals' self-interest changes their understanding of reality (Kunda 1990). The addition of the behavioral ethics and dual reasoning line of research to motivated reasoning is connected to the perspective that moral judgments and decisions are the results of reasoning and deliberation, while self-interest was argued to be an automatic primary motive that needs to be constrained by appropriate inhibitory mechanisms (e.g., Moore and Loewenstein 2004). Along those lines, it was found that honesty requires the availability of cognitive-control resources (Gino et al. 2011; Mead et al. 2009) and time (Shalvi et al. 2012), demonstrating that behaving ethically requires more deliberative resources than behaving unethically. Similarly, Halali et al. (2013) have shown a similar effect with regard to fairness considerations in dictator games settings. These fairness considerations seem to require much more deliberation than self-interest considerations, while the latter are more intuitive (see also Achtziger et al. 2015; Uziel & Hefetz 2014; Xu et al. 2012, for the same pattern of results). Furthermore, Moore et al. (2010) showed that people truly believe their own biased judgments, with only a limited ability to recognize that their behavior was affected by self-interest. Thus, not only has motivated reasoning literature shown that self-interest affects people's understanding of the world around them, but behavioral ethics tells us that self-interest has this effect despite limited awareness in the individual of the existence of this effect.

This study will attempt to explore the question of how the implications of this literature can be applied to enforce ethics in an organization. While the classical debate in enforcement intervention, among both states and organizations, has typically been related to comparing the efficacy of deterrence and morality (for a review see Feldman 2011), the good people paradigm suggests (1) that even garden variety, normative people might engage in unethical behavior without fully recognizing that their behavior is unethical and (2) that they might be less likely to react to these traditional interventions focused on curbing unethicality. Consequently, organizations might require a new way of dealing with wrongdoing. In other words, what seems to be missing from current models of enforcement intervention, according to the good people paradigm, is a consideration of their own relevance in situations that fall within employees' moral blind spots. If indeed most unethical behavior in organizations occurs within such

[2] The "good people" argument does not use the term "good" to mean "moral" or "virtuous." Rather, the focus is on garden variety individuals who might, in various organizational settings, end up behaving unethically without fully recognizing that what they do is unethical.

moral blind spots, it may be necessary to find new types of top-down interventions to change individuals' ethical judgment and behavior.

From an applied perspective, the unaware, unethical effect of self-interest on the behavior of a broad and normative segment of the population—described above as good people—might impose an important challenge in creating new tools to discourage such types of unethical behavior (see Feldman et al. 2013). The ability of incentives and deterrents to affect non-deliberate behavior has been discussed by scholars such as Bazerman and Tenbrunsel (2011), who suggest that "such measures simply bypass the vast majority of unethical behaviors that occur without the conscious awareness of the actors, who engage in them" (p. 111). Following the same line of thought, Banaji and Greenwald (2013) have challenged the classic enforcement approach that focuses on external measures and incentives to control unethical behavior, as such an approach relies on an unjustified central role to self-control, autonomy, and responsibility for one's actions.

Yet, although in various situations "good people" may not be affected by incentives, it is not clear to what extent employees will entirely ignore the existence of incentives. For example, even if the process of self-deception might block one's full awareness of the unethicality of their own behavior, it is possible to predict that introducing sanctions will cause people to be more aware of their behavior and lessen the effect of automaticity. Therefore, traditional intervention techniques that target awareness should not be disregarded, but rather reexamined in light of behavioral ethics literature.

This challenge of reexamining traditional regulatory approach, and possibly creating new ones, is especially important in dealing with subtle conflict of interest in organizations and in government. This specific type of unethical challenge is attributed according to various scholars by people whose behavior breaks no specific law (Lessig 2011).

Subtle Conflict of Interest

In attempting to understand how to intervene and cause people to engage in an ethical way, the current study focuses on the concept of subtle conflict of interest. Conflict of interest is the basic paradigm that lies at the heart of most organizational misconducts and receives a special attention in almost every type of ethical code (Stevens 1994). It also contributes to the spread of corruption across numerous government and business contexts. For example, in the medical field, there is a fertile ground for conflicts of interest, where even scientists and doctors, who believe they are doing what is best for the public health, might ultimately engage in behaviors that favor the entities who compensate them (Feldman et al. 2013). Another common example can be found in clinical studies that are financed by pharmaceutical companies, which provides an incentive for physician-researchers to reach certain results that would benefit those companies (e.g., Friedberg et al. 1999; Hillman 1987; Rodwin 1989, 2012). An additional example is the transition of professionals from the public to the private sector. This is a common phenomenon, which is highly problematic for the ability of public sector employees to focus only on the interest of the public (Che 1995). This process, referred to as the "revolving doors," presents a possible conflict of interest. However, the problem is much greater than these examples suggest. One hypothesis suggests that anticipation of future opportunities in regulated firms may cause regulators to be less aggressive when administering regulatory policy, even without full awareness that this anticipation may alter their behavior (Che 1995). In many other areas, including those of lawyers vis-à-vis their clients, executives vis-à-vis shareholders, prosecutors in plea bargains, and academics involved in the promotion of their colleagues, most good people may believe that the option that promotes their self-interest is also the correct one. Thus, in such situations, there might be only limited wisdom in threatening people with punishment for corruption.

Understanding the behavior of people in all of these situations cannot be gleaned from current behavioral ethics studies of dishonesty where subjects sit in the lab and are asked, for example, to report how many assignments they have solved (e.g., Mazar et al. 2008). In these situations, the decision to cheat is clear and unambiguous, which is not the case when employees behave unethically in conflict of interest situations. This important difference is especially salient in contexts where people's motivated reasoning might lead them to feel that their choice to prioritize their self-interest is, in fact, also the right solution for the organization they work for (see also Zamir and Sulitzeanu-Kenan 2016). Thus, conflicts of interest—especially subtle ones—are different from the focus of most dishonesty studies: In such situations, both the interest and the behavior are subtle and could be seen as almost legitimate. In contrast, in most of the classical dishonesty studies, the lie is clear to the participant. For example, participants know that for every matrix that they over-report, they get more money. Thus, those who behave dishonestly know that what they do is wrong; they do it for a profit and find various excuses to justify their behavior (Ayal and Gino 2011). However, in many of the real-life organizational studies, people constantly evaluate and judge whether a certain employee is good, or a certain program is good, or a certain company is worth buying.

Indeed, there are some existing studies that have taken this approach, such as Cain et al. (2005) study on conflict of interest in counting jellybeans; but even in those situations, participants knew that the numbers they provided (in their advice) were false. A different study by Pittarello et al. (2015) focused on more ambiguous situations, where participants had to identify the location of blurred stimuli. In this instance, subjects could easily justify dishonesty in reporting the location of the stimulus simply because there is no one correct answer. As a result, subjects may not have been consciously aware of their dishonesty.

We believe that the experimental approach we present in the current work contributes to the literature in a number of ways. First, we use Amazon Mechanical Turk (MTurk) employees who were hired to evaluate a research center. This example is relatively similar to tasks that many employees engage in because they do not make an ultimate up or down decision about the research center; rather, they evaluate a certain proposal, which is much more similar to the type of evaluation assignments people perform in their regular jobs. Second, we focus on an ambiguous context, which we believe to be the most common area where good people could engage in motivated reasoning and self-deception regarding their unethicality (see Pittarello et al. 2015). Third, our study could contribute to the hundreds of enforcement studies, which documented numerous factors that change the efficacy of deterrence and morality on changing people's behavior. We believe that learning how different interventions operate when dealing with subtle conflicts of interest will contribute also to the nudge literature, which was used to curb the unethicality of good people (Shu et al. 2012), and which is now a leading theory in the domain of non-deliberative choices in general and in psychology in particular. Our project's assumption is that, in the rush to adopt nudges and implicit intervention to regulate the implicit behavior of people, we need to explore the efficacy of more traditional interventions in order to test whether or not we should abandon these methods, even when dealing with good people who supposedly do not fully recognize the wrongdoing in their behavior. In other words, the aim of our study is to test the complexity of the good people argument in ethical decision making where, on the one end of the spectrum, people fail to recognize the wrongness of their own behavior, and on the other end, people create a self-imposed ethical line which they don't cross.

The Current Work

In the current work, we examined how material interests imperceptibly affect decision making, and the effectiveness of negating self-interests in conflict with organizational duty through classical and new intervention approaches. Many questions in this field remain open, as most of the new research on decision making and behavioral ethics does not appear to gain expression in the research and practice of conflict of interest. Little is known about what should be done to effectively change the putative influences outlined above. Understanding the process by which self-interest operates is naturally a key to understanding how to curb these influences and to determine which intervention method legal policy makers should focus on in various contexts of interest.

To answer these questions, we designed two studies that focus on people's behavior in conflict of interest situations. In the first study, we focused on understanding the process through which conflicts of interest might affect people more—intuitive/automatic or analytical/deliberative mindsets. Following the vast majority of the research on bounded ethicality (see Bereby-Meyer and Shalvi 2015 for review), we hypothesized that, indeed, compared to analytical/deliberative mindset, intuitive/automatic mindset increases unethical decision in response to subtle conflicts of interest. In the second study, we focused on pinpointing the best intervention methods to curb such behaviors—deterrence or morality—and the best way to implement them—explicitly or implicitly. Deterrence and morality are very common intervention practices used by the authorities: *Deterrence* serves as a traditional function of the law and relies on extrinsic motivation to shape behavior; in contrast, *morality* focuses on changing an individual's intrinsic motivation. Extrinsic motivation refers to actions driven by external commands or incentives and can be achieved by targeting the potential offender's financial status through rewards and fines. Conversely, intrinsic motivation is linked to behavior driven from within the individual, usually out of a sense of moral or civic duty (Deci et al. 1999; Kasser and Ryan 1996),[3] and it is affected by targeting the sense of morality (see Feldman 2009, 2011).

As far as we know, the current research provides a first look into the ability of organizations to curb the unethical behavior of "good" people in subtle conflicts of interest. The focus on people in subtle conflicts of interest attempts to replicate situations in which it would be very easy for people to behave in self-interested way without fully acknowledging that their behavior is unethical. In obvious situations, where people are bribed to act against their duty of loyalty to the state or to the corporation, the now

[3] Originally, most discussions of intrinsic motivation have been within the context of interest in the task. *See generally* Deci, Koestner, and Ryan (1999), describing the research approach and results of a number of studies on intrinsic motivation; *see also* Kasser and Ryan (1996), examining the differences in individual well-being associated with focusing on extrinsic and intrinsic goals.

common "blind spot" argument is tenuous and is less likely to occur.

In our subtle conflict of interest scenario, we used a case study where the gap between ethical demands and self-interest is minimal. Specifically, we placed our participants in conflict of interest between what they were hired to do (to evaluate a specific research center in an objective way) and their personal interest (to write good things about the research center so they might be invited to participate in an additional study for additional compensation). In addition, we included other components to make this experimental setting suitable for examining the behavior of good people from two different angles.

First, instead of actual bribes or anything else overtly unethical, we introduced subtle conflicts of interest, which leave more room for implicit corruption. This distinction appears in the conflict of interest literature (Moore and Tanlu 2010). For the most part, it has been argued that it is more beneficial to investigate the behavior of good people when they are facing a subtle conflict of interest because they are more likely to be unaware of the influence that such conflict has on them (Chugh et al. 2005). Hence, consistent with research on the contribution of ambiguity to dishonesty discussed above, we created a situation where the incentive to shirk was presented in an ambiguous way, where there was no direct link between the behavior of the MTurk employee and reaping the rewards. Rather, the employee's evaluation of the research center only slightly increases the odds of beneficial rewards in the future, but does not guarantee them. This ambiguity is not only interesting theoretically, but also practically: Many of the revolving doors conflicts occur in areas where there is no clear link between one's level of favoritism in the public sector and her likelihood of being hired by a relevant regulated party upon exiting to the public sector (Cornaggia et al. 2016; Gormley 1979).

Second, we created two levels of bias participants could express to increase their chances of getting a future reward (i.e., issues regarding the research conducted in the research center and issues regarding the researchers working in the research center). By creating these two levels, we allow for another examination of good peoples' behavior in conflict of interest situations. According existing studies (Mazar et al. 2008), part of the concept of good people is related to the perspective that good people choose not to lie as much as they could by rational choice accounts. By creating two possible levels of biased evaluation in conflict of interest situations, we allow for extending those insights to understanding the behavior of people in conflict of interest situations and their self-imposed limits.

In both experiments, we recruited participants from the online labor market, Amazon Mechanical Turk (MTurk). MTurk is an online labor market in which employers can employ workers to complete short tasks (generally in fewer than 10 min) for relatively small amounts of money (generally up to $1). Workers receive a baseline payment (i.e., show-up fee) and can be paid an additional bonus depending on their performance. Importantly, while the reputation of most in-lab participants is usually unknown, as most behavioral labs do not (or cannot) use a reputation system to create lasting, publicly available reputations, MTurk does know of participants' reputations. Therefore, reputations of both the employers (i.e., requestors) and the workers can be sacrificed if either one of them behaves unfairly (for a further descriptions of Mechanical Turk sampling, see Buhrmester et al. 2011; Paolacci and Chandler 2014; Rand et al. 2012). Considering this unique characteristic, we reasoned that a sample of MTurk participants would be a better representation of the relationship between employers and employees in the real world.

In both studies, we measured participants' objectivity in evaluating the research institute and its scientists as described in the questionnaire. We presented participants with various conditions, which created opportunities for them to advance their manipulated self-interest by shifting their judgments in favor of the described institution. If participants in the experimental group provided a biased evaluation of the research institution relative to a control group who had no financial interest in the evaluation, we could identify deviation in their evaluations from the control group's objective evaluations. Such deviations could not be defined as corrupt or unethical. However, as suggested in the introduction, we have intentionally focused on these ambiguous contexts as a way to account for the Blind Spot argument (e.g., Banaji and Greenwald 2013; Chugh et al. 2005; Sezer et al. 2015). Participants were then assigned to a few randomized group in which they learned, either implicitly or explicitly, of being either under a regime of penalty or of appeal to morality. Next, participants answered two questionnaires regarding the research institute. The first one included items focusing on the research conducted at the institute and on the scientists working there; the second focused strictly on the research and was aimed to assess the participants' agreement with different statements and their willingness to actively engage along the lines delineated in the statements. Lastly, we tried to assess participants' sense of objectivity regarding the research institute when answering the previous questionnaires.

Experiment 1

Participants

Ninety-nine participants (52.5% males, 47.5% females) completed the experiment online through MTurk in exchange for $1. Additional collected demographics included Race (74.7% White, 7.1% Black, 4.0% Hispanic, 10.1% Asian, 4.0% Other), Age (18.2% 18–24 years old, 44.4% 25–34 years old, 23.2% 35–44 years old, 9.1% 45–54 years old, 5.0% 55 years old and over), and level of education (1.0% less than high school, 10.1% high school/GED, 26.3% some college, 8.1% 2-year college degree, 43.4% 4-year college degree, 8.1% master's degree, 3.0% PhD/MD/JD). Participants were all US residents with a previous HIT Approval Rate of 80% or better. We excluded responses from one participant who attempted to complete the study multiple times.[4] All participants signed an informed consent form before participating in the study. We randomly assigned participants to one of two experimental mindset conditions (intuitive/analytical).

Procedure

After signing the consent form, participants went through a mindset manipulation. Next, participants read a paragraph describing the Edmond J. Safra Research Center and were introduced to the conflict of interest. Subsequently, participants received three different questionnaires: (1) an 18-item questionnaire, (2) a binominal questionnaire, and (3) an objectivity questionnaire. These questionnaires aimed to evaluate (1) their opinion, (2) their support for the research institute, and (3) their sense of objectivity with regard to the research institute when answering the questionnaires. Finally, participants answered a demographic questionnaire.

Materials

Mindset Manipulation

Relying on Shenhav et al. (2012; also used by Rand et al. 2012), we manipulated mindset by asking participants to write a paragraph of 8–10 sentences recalling an episode from their life in which their *intuition/first instinct* (i.e., intuitive/automatic mindset) or *carefully reasoning through a situation* (i.e., analytical/deliberative mindset) led them in the right direction and resulted in a good outcome. Then, prior to the 18-item questionnaire, participants were asked to rely on intuition/reasoning when making their responses. Lastly, in order to further strengthen participants' reliance on reasoning in the analytical condition, we told the participants in this condition that, following their responses to the 18-item questionnaire, they will be required to describe in writing the reasons for their evaluations (Wilson and Schooler 1991).

The Conflict of Interest

We created a potential for conflict of interest (COI) by telling participants that "currently, the Edmond J. Safra Center is running an important additional online experiment (with additional, higher relative pay) to examine modes of…," and that "participants for this experiment will be selected based on their answers to the current survey." Following this statement, we asked participants to indicate whether they would like to be considered for this additional experiment. Eleven participants indicated that they did not want to be considered for the additional experiment (five in the intuitive mindset condition and six in the analytic mindset condition). The goal of the conflict of interest manipulation was to place participants in a conflict of interest between what they were hired to do (i.e., to evaluate the research center in an objective way) and their personal interest (i.e., to write good things about the research center so they might be invited to participate in additional studies for additional compensation). Because participants who did not want to be considered for the additional experiment could not get the opportunity to earn more money on their evaluations, their evaluations were not subject to a conflict of interest. In other words, they did not have any material temptation to write good evaluations. Since the current work focuses on conflict of interest, we excluded these participants from further analyses. Every participant who wanted to be invited for the additional study was given a link to that study at the end of the experiment and received additional bonus for that additional study.

The 18-Item Questionnaire

The 18-item questionnaire included nine items that focused on the research conducted by the institute, such as "Research conducted by this center is more important than most other research I'm familiar with in the social sciences," and eight items that focused on the scientists working at the institute, such as "The salaries of scientists at this center should be higher than other scientists' salaries" (see "Appendix" for the full questionnaire). Participants were asked to indicate their answers "as objectively as possible" on a scale of 1 (strongly disagree) to 6 (strongly agree). Agreement on all items indicated an evaluation favorable to the research institute. To make sure participants read each item before answering, one of the items (number 10) required participants to provide a specific rating ("2") and did not include any question regarding the research or the scientists of the institute. Except for one participant who missed the answer, all of the participants responded correctly to this item.

[4] These responses were identified based on duplicated IP addresses and GPS locations.

The Binominal Questionnaire

The binominal questionnaire included four binominal questions regarding each of the following three statements about the research institute: (a) "Research conducted by the Safra Center is crucial for the well-being of society," (b) "The Safra Center's research will change the way we look at public institutions," and (c) "The Safra Center's mission is the first attempt ever to deal with one of our most important problems." In reference to each statement, participants had to indicate whether (a) it is accurate/inaccurate, (b) they agree/disagree with it, (c) would/would not make it to a potential donor, and (d) would/would not be willing to sign a petition.

The Objectivity Questionnaire

The objectivity questionnaire included the following yes/no questions assessing their sense of objectivity with regard to the research institute when answering the questionnaires: (a) "Do you think you were influenced by anything while you were answering the questions?" (b) "Were you completely objective during this study?" and (c) "Did you consider anything besides your best judgment while answering these questions?"

Results and Discussion

The average time of survey completion in the intuitive and the analytic mindset conditions was 11:14 (SD = 8.7) and 12:68 (SD = 7.2) minutes, respectively. We excluded three outlier participants (two in the intuitive mindset condition and one in the analytic mindset condition) from all analysis because their completion time (41 and 46 min in the intuitive condition and 44 in the analytic condition) was more than three standard deviations away from the mean in their condition (*Z* values were 3.34, 3.90, and 4.45, respectively). Therefore, including all aforementioned excluded participants, we excluded a total of 15 participants from all the following analyses. The pattern of the following reported results was similar when all these participants were included in the analysis.

The 18-Item Questionnaire

For each participant, we calculated a separate mean score for the research-related items (mean = 3.9, median = 3.9, SD = .77) and for the scientist-related items (mean = 3.2, median = 3.3, SD = 1.0). Cronbach's alpha reliability of these items was .88 and .90, respectively. Next, we entered the mean scores into a mixed-model analysis of variance (ANOVA), with mindset condition as a between-participants variable and the item issue (research, scientists) as a within-participants variable.

A significant main effect for item issue, $F_{(1,82)} = 86.88$, $p < .001$, $\eta_p^2 = .51$, indicates that participants' mean evaluation regarding the research conducted at the institute ($M = 3.98$, SD = .79) was more positive than their mean evaluation of the scientists working at the institute ($M = 3.22$, SD = 1.07) across all conditions. Importantly, and as expected, participants' mean evaluation in the intuitive mindset ($M = 3.81$, SD = .83) compared to the analytic ($M = 3.44$, SD = .86) mindset condition was significantly more positive, $F_{(1,82)} = 3.99$, $p < .05$, $\eta_p^2 = .05$, with no condition × item issue interaction ($F < 1$, n.s.).

The Binominal Questionnaire

The Kuder–Richardson (Kuder and Richardson 1937) formula 20 reliability of the 12 items in the binominal questionnaire was .84. For each participant, we calculated the proportion of answers in favor of the research institute. We entered this proportion into a one-way ANOVA with mindset condition as a between-participants variable. As in the 18-item questionnaire, participants' favoritism toward the research institute in the intuitive ($M = .72$, SD = .20) compared to the analytic ($M = .66$, SD = .30) mindset condition was more positive. This difference, however, did not reach statistical significance $F_{(1,85)} = 1.23$, $p = .27$, $\eta_p^2 = .02$.

The Objectivity Questionnaire

Because the objectivity questionnaire included only three binominal items, we entered the meaning of the participants' answer (0: objective, 1: non-objective) into a generalized probit estimation equation for binominal data, with mindset condition (intuitive, analytic) as a between-participants independent variable, question (1, 2, 3) as within-participants independent variables, and the participants as a random factor. A significant main effect for question, *Wald* $\chi^2_{(2)} = 13.60$, $p = .001$, revealed that more participants indicated non-objective behavior on question 1 (20.2%) than on question 2 (13.1%), *Wald* $\chi^2_{(1)} = 3.36$, $p = .067$, and more on question 2 than on question 3 (4.8%), *Wald* $\chi^2_{(1)} = 7.29$, $p = .007$. Yet, the main effect for condition and the question * condition interaction was not significant: All *Wald* $\chi^2 < 1$, *n…s.*, indicating that while participants in the intuitive, compared to the analytic, mindset condition favored the research institute they, did not feel less objective, or at least did not report they were less objective.

Thus, the results of the first study show that intuitive/automatic mindset—i.e., lack of deliberation—is related not only to dishonesty, on which much of the literature has focused, but is also related to unethical behavior in subtle conflict of interest situations. Hence, they broaden the horizons of current behavioral ethics literature by shifting the focus away from pure dishonesty. Under subtle conflict of interest between what participants were hired to do (i.e., to evaluate the research center in an objective way) and their personal interest (i.e., to write good things about the research center so they might be invited to participate in additional studies for additional compensation), intuitive/automatic mindset increases the likelihood that participants will provide favorable reviews toward the Safra Center relative to the participants who were put in the analytical/deliberative mindset. However, this study provides no suggestions as to *how* can we regulate people's behavior, given that we are more likely to behave in a non-objective way in an intuitive mindset, an explicit intervention or an implicit one?

Experiment 2

Participants

Three-hundred and twenty participants (51.6% males, 48.4% females) completed the experiment online through MTurk in exchange for $1. Additional collected demographics included Race (76.2% White, 7.5% Black, 4.7% Hispanic, 7.5% Asian, 4.1% Other), Age (21.3% 18–24 years old, 42.5% 25–34 years old, 18.4% 35–44 years old, 8.1% 45–54 years old, 9.7% 55 years old and over), and level of education (.9% less than high school, 9.4% high school/GED, 27.5% some college, 8.8% 2-year college degree, 40.0% 4-year college degree, 10.9% master's degree, 2.5% Ph.D./MD/JD). Participants were all US residents with a previous HIT approval rate of 80% or better. All participants signed an informed consent form before participating in the study. Participants were randomly assigned to six experimental conditions: four conditions with conflict of interest and different forms of deterrence or morality interventions, one control condition with conflict of interest without any intervention, and one control condition with no conflict of interest and no intervention. Below are the six groups[5]:

1. Conflict of interest with explicit deterrence ($n = 52$).
2. Conflict of interest with explicit morality ($n = 55$).
3. Conflict of interest with implicit deterrence ($n = 54$).
4. Conflict of interest with implicit morality ($n = 54$).
5. Control group: conflict of interest with no intervention ($n = 56$).
6. Control group: no conflict of interest and no intervention ($n = 49$).

Procedure

After signing the consent form, participants read the same paragraph as in Experiment 1, describing the Edmond J. Safra Research Center. Next, they were introduced to the conflict of interest, followed by various forms of morality or deterrence interventions. Subsequently, participants received the same 18-item questionnaire, binominal questionnaire, objectivity questionnaire, and demographic questionnaire as in Experiment 1. In the control with no conflict of interest condition, participants answered the set of questionnaires immediately after they read the paragraph describing the research institute (i.e., participants were not exposed to the conflict of interest and did not go through any intervention). In the control with conflict of interest condition, participants answered the questionnaires after being exposed to the description of the conflict (i.e., participants did not go through any intervention).

Materials

The Conflict of Interest

We have created a potential for conflict of interest (COI) using the same manipulation as in Experiment 1. Fifteen participants in the COI conditions indicated that they did not want to be considered for the additional experiment (three in explicit deterrence, four in explicit morality, three in implicit deterrence, two in implicit morality, and three in no intervention). As in Experiment 1, we excluded these participants from further analyses as if they did not want to be considered for the additional experiment they could not have experience the conflict of interest during the current experiment. Every participant who wanted to be invited for the additional study was given a link to that study at the end of the experiment and received additional bonus for that additional study.

Explicit Deterrence

We manipulated explicit deterrence by asking participants to read a paragraph on government crackdown and stating that "In accordance with this worldwide trend, we believe that people who let their conflict of interest affect their

[5] In the original design, along with the material-based conflict of interest used in the current paper, we had five more group of participants that went through an identity-based conflict of interest manipulation as an additional type of conflict of interest. One group with no intervention manipulation and four groups with the same intervention manipulations we used for the material-based conflict of interest. We did not find any effect for the identity-based conflict of interest on participants (as compared to the control group of no COI condition). Since the focus of the current experiment was to examine how can we regulate people's behavior in a conflict of interest situation, there was no point in including these conditions in the paper, so we focused only on the material-based conflict of interest conditions.

objectivity and integrity when completing this survey should be penalized. Hence, participants who let their conflicting interests affect their judgment might lose some of their compensation for the work they do for us." Next, we asked participants to answer a three-item questionnaire to verify their understanding of the explicit deterrence intervention (see "Appendix" for a full display of the manipulation and the followup questions). Three participants in the explicit deterrence intervention failed to answer the three deterrence comprehension items correctly. We excluded these participants from further analysis.

Explicit Morality

We manipulated explicit morality by asking participants to read a paragraph explaining why, in a situation of conflicting interests, acting based on one's self-interest is immoral, and stating that "In accordance with this worldwide trend, we believe that people who let their conflict of interest affect their objectivity and integrity when completing this survey are not acting in a moral and ethical way. Hence, participants who will let their conflicting interests affect their judgment might harm the public good." Next, we asked participants to answer a three-item questionnaire to verify their understanding of the explicit morality intervention (see "Appendix" for a full display of the manipulation and the followup questions). Nine participants in the explicit morality intervention failed to answer the three morality comprehension items correctly. We excluded these participants from further analysis.

Implicit Deterrence

We manipulated implicit deterrence using a 35-item word completion test in which 11 of the items were words related to deterrence (e.g., punishment, subpoena, indictment) in order to prime participants with concepts of deterrence. Each of these 11 prime words was tested and found to have several hundred-thousand Google results shared with the word deterrence (see "Appendix" for a full list of these words). Methodologically, in priming the targets using lists of words, we have followed a rationale similar to those used in other studies in which priming words are used to induce a state of mind (e.g., Srull and Wyer 1979; Norenzayan and Shariff 2008). In contrast to scrambled sentences used in those papers, we have used a word completion task to get people to think about the two modes of compliance motivation—deterrence and morality. Further, this method was recently used as a dependent measure in a very influential paper on bounded ethicality (Shu et al. 2012). We excluded two participants from further analysis because they failed to identify five or more of the 11 deterrence prime words.

Implicit Morality

The implicit morality intervention was the same as the implicit deterrence intervention except that the 11 prime words were related to morality (e.g., integrity, morality, honesty) rather than to deterrence. Similar to the implicit deterrence intervention, each of these 11 prime words was tested and found to have several hundred-thousand Google results shared with the word morality (see "Appendix" for a full list of these words). The remaining 24 neutral items were the same as in the implicit deterrence intervention. We excluded from any further analysis two participants who failed to identify six or more of the 11 morality prime words.

The 18-Item Questionnaire

The 18-item questionnaire was the same as in Experiment 1.[6]

The Binominal Questionnaires

The three statements and the binominal questionnaires were the same as in Experiment 1.

The Objectivity Questionnaire

The objectivity questionnaire was the same as in Experiment 1.

Results and Discussion

Table 1 displays the average time and standard deviations for survey completion in all experimental conditions. We excluded three outlier participants (one in the explicit deterrence condition, one in the explicit morality condition, and one in the control-COI condition) from all analysis since their completion time (48, 45, and 188 min, respectively) was more than three standard deviations away from the mean in their condition (*Z* values were 3.63, 4.04, and 6.85, respectively). Therefore, including all aforementioned excluded participants, we excluded a total of 34 participants from all the following analyses. The pattern of

[6] By mistake, one of the items referring to the researchers in the center was worded in the opposite way to all other items (i.e., disagreement indicated a favorable evaluation of the research institute). Because this was the only item formulated in such a way, and because Cronbach's alpha reliability of the researcher items, with the inclusion of the reversed responses to this item resulted in a drop from .83 (without this item) to .75, we excluded this item from further analysis. The pattern of the following reported results was similar when this item was included in the analysis.

Table 1 Average time and standard deviations for survey completion in all experimental conditions in Experiment 2

	Average survey completion time (and SDs)
1. Explicit deterrence	20.2 min (7.7)
2. Explicit morality	18.7 min (6.5)
3. Implicit deterrence	21.0 min (8.8)
4. Implicit morality	23.1 min (8.1)
5. Control-COI	22.7 min (24.1)
6. Control-no COI	20.5 min (9.4)

the following reported results was similar when all of these participants were included in the analyses.

The 18-Item Questionnaire

For each participant, we calculated a separate mean score for the research-related items (mean = 4.1, median = 4.1, SD = .85) and for the scientist-related items (mean = 3.1, median = 3.1, SD = .95). Cronbach's alpha reliability of these items was .89 and .83, respectively. Next, we entered the mean scores into a mixed-model analysis of variance (ANOVA) with condition as a between-participants variable and the item issue (research, scientists) as a within-participants variable.

A significant main effect for item issue, $F_{(1,280)} = 443.41$, $p < .001$, $\eta_p^2 = .61$, indicates that participants' mean evaluation regarding the research conducted at the institute ($M = 4.07$, SD = .86) was more positive than their mean evaluation of the scientists working at the institute ($M = 3.07$, SD = .95) across all conditions. The main effect of condition was significant, $F_{(5,280)} = 5.22$, $p < .001$, $\eta_p^2 = .09$, as was the condition × item issue interaction, $F_{(5,280)} = 3.15$, $p < .01$, $\eta_p^2 = .05$.

As shown in Fig. 1, subsequent analyses of the condition × item issue interaction revealed that participants' evaluations of the research conducted at the institute in the control-COI condition and the two implicit intervention conditions (deterrence, morality) were significantly higher than those of the control-no COI condition and the two explicit intervention conditions (deterrence, morality), $F_{(1,462)} = 20.52$, $p < .001$, $\eta_p^2 = .07$. These results indicate that the opportunity to earn extra money by participating in another experiment of the research institute caused participants to behave less ethically: They demonstrated favorable disposition toward the research conducted at the institute in the control-COI condition compared to the control-no COI condition. As for the different forms of interventions, while both the explicit deterrence and explicit morality interventions were effective, that is, they resulted in evaluations similar to those in the control-no COI condition, the implicit deterrence and implicit morality interventions were not effective. Participants' evaluations in these groups did not differ from those in the control-COI condition.

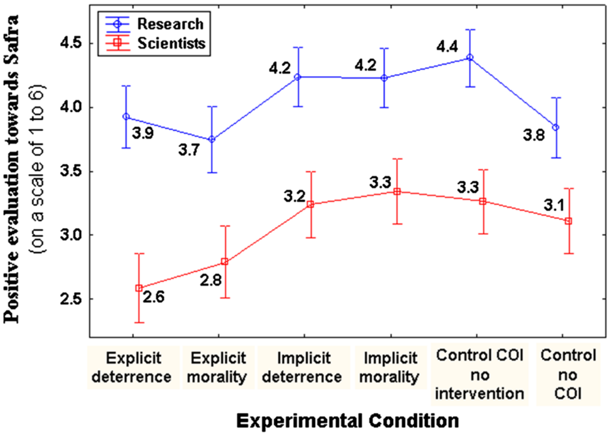

Fig. 1 Participants' evaluations toward Safra as a function of condition and item issue. *Error bars* represent .95 confidence intervals

In contrast to the effect of the COI on participants' evaluations of the research (as illustrated by the difference between the control-COI and the control-no COI conditions), the COI appears not to have affected participants' evaluations of the scientists working at the institute, as the evaluations in the control-COI condition were not significantly different from the control-no COI condition ($F < 1$, n.s.). Interestingly, as shown in Fig. 1, while participants' evaluations of the scientists in the COI conditions with implicit interventions (deterrence, morality) did not differ from the two control groups (COI, no COI), in the COI conditions with explicit interventions (deterrence, morality) participants' evaluations were significantly lower than those of participants in the four former conditions (control-no COI, control-COI, implicit deterrence, implicit morality), $F_{(1,280)} = 21.81$, $p < .001$, $\eta_p^2 = .07$. This pattern of results suggests a chilling effect (Craswell and Calfee 1986), that is, even more "objective" evaluations than in the control group, where there was no conflict.

The Binominal Questionnaires

The Kuder–Richardson (Kuder and Richardson 1937) formula 20 reliability of the 12 items in the binominal questionnaires was .85. For each participant, we calculated the proportion of answers in favor of the research institute. We entered this proportion into a one-way ANOVA with condition as a between-participants variable. The main effect of condition was significant, $F_{(5,280)} = 2.31$, $p < .05$, $\eta_p^2 = .04$. Subsequent analysis revealed similar results to those observed in the analysis of the participants'

Table 2 Proportion of answers in favor of the research institute in all experimental conditions in Experiment 2

	Proportion of answers in favor of the research institute (%)
1. Explicit deterrence	63.0
2. Explicit morality	58.9
3. Implicit deterrence	71.9
4. Implicit morality	68.0
5. Control-COI	73.6
6. Control-no COI	62.4

Table 3 Proportion of non-objective answers in all experimental conditions in Experiment 2

	Proportion of non-objective answers (%)
1. Explicit deterrence	9.6
2. Explicit morality	21.1
3. Implicit deterrence	10.9
4. Implicit morality	17.3
5. Control-COI	19.9
6. Control-no COI	9.5

evaluations of the research conducted at the institute, as measured by the 18-item questionnaire. Specifically, participants' favoritism toward the research institute in the control-COI condition and in the two implicit intervention conditions (deterrence, morality) was significantly higher than in the control-no COI condition and the two explicit interventions conditions (deterrence, morality), $F_{(1,280)} = 9.86$, $p < .002$, $\eta_p^2 = .03$ (see Table 2). These results replicate the results regarding the research conducted in the institute, as measured by the 18-item questionnaire, indicating that the opportunity to earn extra money by participating in another experiment caused participants to be more favorably inclined toward the research conducted at the institute, and that only the explicit forms of deterrence and morality interventions were effective.

The Objectivity Questionnaire

Because the objectivity questionnaire included only three binominal items, we entered the meaning of the participants' answer (0: objective, 1: non-objective) into a generalized probit estimation equation for binominal data, with condition (depletion, no-depletion) as a between-participants independent variable, question (1, 2, 3) as within-participants independent variables, and the participants as a random factor. A significant main effect for question, *Wald* $\chi^2_{(2)} = 55.61$, $p < .001$, revealed that more participants indicated non-objective behavior on question 1 (24.1%) than on question 2 (13.6%), *Wald* $\chi^2_{(1)} = 20.96$, $p < .001$, and more on question 2 than on question 3 (6.3%), *Wald* $\chi^2_{(1)} = 12.27$, $p = .001$. Moreover, the main effect for condition was marginally significant: *Wald* $\chi^2_{(5)} = 10.40$, $p = .065$ (see Table 3), while the question * condition interaction was not significant, *Wald* $\chi^2_{(10)} = 5.48$, $p = .857$).

Subsequent analysis of the condition effect revealed that, in the control-COI condition and the two morality intervention conditions (explicit, implicit), more participants indicated non-objective behavior compared to participants in the control-no COI condition and in the two deterrence interventions conditions (explicit, implicit), *Wald* $\chi^2_{(1)} = 9.17$, $p = .002$. Note that the two morality intervention conditions (explicit, implicit) behaved differently in the main analysis of the research conducted at the institute, as measured by the 18-item and by the binominal questionnaires. Specifically, under the implicit morality condition, participants were less objective as compared to the control-no COI (more in favor of the Safra Center), whereas under the explicit morality condition, they were not. It is possible, therefore, that under these two morality conditions, participants' higher proportion of self-reported non-objective behavior is the result of self-justification following the morality intervention, and not a result of a true sense of non-objectivity *during* the study. In contrast, participants' higher proportion of self-reported non-objective behavior in the control-COI condition, compared to the control-no COI condition, cannot be explained by self-justification, as these participants did not go through any intervention manipulation. Further, participants' higher proportion of self-reported non-objective behavior in the control-COI condition is consistent with these participants' exaggerated favorable evaluations of the research conducted at the institute (i.e., their actual unethical behavior). It appears, therefore, that these participants not only behaved unethically but also were aware of it.

Summary of Findings and Discussion

In the current study, we examined how different intervention techniques affected people when they were faced with a subtle conflict of interest (i.e., opportunity to earn additional money if invited to participate in a future study) and an opportunity to engage in a subtle behavior whose ethicality is ambiguous (i.e., rating the Safra Center in somewhat more favorable way could be attributed to other motives). First, by manipulating the self-interest of participants relative to a control group of participants who could not earn additional fee for an additional experiment, we found a substantial "corrupting" effect when we observed that the control-COI group reported a more favorable view of the target stimuli than did the control-no COI group. Although the potential effect of money on behavior is neither new nor surprising, the fact that an opportunity to earn such small amounts of extra money in future research, subtly mentioned to the participants, increased their evaluation of the research institute, even though they were explicitly asked to conduct their evaluations objectively, reveals the corrupting potential of subtle conflict of interests. Further, compared to control-COI, explicit interventions (both deterrence and morality) had a significant constraining effect on participants' judgments. In contrast, implicit interventions (again, with a similar effect of deterrence and morality) did not affect participant's judgments. These patterns were obtained with two different dependent variables—an 18-item Likert-type scale questionnaire and a binominal questionnaire. Thus, raising the participants' awareness of the possibility of their unethical behavior, using rather simple explicit interventions, was sufficient even in the context of subtle conflict of interests to prevent some of those behaviors.

Taken together with the results of Experiment 1, according which unethical behavior in subtle conflict of interest situations is more pronounced under intuitive/automatic compared to analytical/deliberative mindsets, Experiment 2's results support the claim that unethical behavior is associated with automatic, System 1 processing (e.g., Mead et al. 2009; Shalvi et al. 2012). Furthermore, consistent with the suggestion of the dual-model perspective that System 2 has the ability to override or inhibit default responses emanating from System 1 (Stanovich 1999), it appears that the type of the intervention being used (deterrence/morality) is less important, as long as it is conducted explicitly so that it triggers deliberate System 2 processing. By contrast, at least in the context studied here, implicit interventions did not have any effect on participants' unethical behavior, suggesting that interventions based on System 1 processing might be less effective in overriding System 1 unethical behavior.

Yet, in the current work we used a word completion task to implicitly manipulate morality and deterrence interventions. While this method was recently used as a dependent measure in a very influential paper on bounded ethicality (Shu et al. 2012), it might be less effective as an independent implicit manipulation; however, the current null effects for this method should not be interpreted as an inability of implicit interventions in general to prevent unethical behavior in subtle conflict of interest. Specifically, it is possible that different implicit methods, or even the same one with a higher proportion of prime words (in the current work we had only 11 primes out of 35 words), might be more affective in constraining unethical tendencies. In line with this, Gino and Desai (2012) have recently found that exposure to childhood cues, as a different form of implicit intervention, reduced unethical behavior.

Nevertheless, taken together, the current findings strengthen our claim that traditional intervention techniques, which assume awareness should not be completely washed away by a more innovative nudge-like techniques advocated in recent behavioral ethics literature. Instead, the classic deterrence literature should be modified in light of the literature on behavioral ethics. It should not be abandoned all together even when dealing with "good" people. It might be the case that as Chugh et al. (2005) argue, incentives will not change the behavior of those who engage in implicit corruption. However, since people's level of awareness might not be anticipated ex-ante, incentives' importance should not be undermined without further empirical examination, which should be done in specific organizational and legal contexts.

These results improve our understanding of the approach organizations should adopt to deal with the unethicality of good people. Understanding the combination of these two possible evaluations—substantial (evaluating the center) and personal (evaluating the scientists working in the center)—contributes to our understanding of the good people's unethicality. Participants were biased in their responses about the research center due to the mere possibility of participating in additional, more profitable, research. However, they were not providing biased estimates of the scientists. This can be explained by the fact that the latter judgment was based on far less information than the former and gave participants.

The fact that it was more difficult to "corrupt" participants as to the scientist measure is consistent with previous findings that people behave unethically to the extent that they can justify their actions (Schweitzer and Hsee 2002; Shalvi et al. 2011). Participants in our study have shown self-restraint against corrupting influences in situations in which they could not have produced a justifiable consideration for changing their judgment. People could feel good about themselves for expressing favorable views about research on ethics, but it may have been more difficult to find justifiable reasons for expressing favorable views about the scientists themselves when they were not given any information that would help them make this judgment. Thus, our findings support the focus on "good" people, as from a rational choice perspective there were fewer justifications to express views that were more favorable to the scientists than to Safra Center generally. Interestingly, however, in contrast to Experiment 2's results, in Experiment 1, in response to a subtle conflict of interest intuitive/automatic mindset, compared to analytical/deliberative mindset, increased unethical favorable judgments regarding both the research conducted in the institute and regarding the scientists working there. These findings are consistent with the results reported by Shalvi et al. (2012) and highlight how potentially calamitous people's intuitive/automatic self-serving tendency can be as it seems to evoke unethical behavior even in settings in which people typically refrain from behaving unethically (when self-justifications are lacking).

Another important finding that emerges from the study has to do with the self-awareness of participants as to whether they were objective in their evaluation. As suggested in "Introduction," this is an important area of research for the interaction between behavioral ethics and the law because of the centrality of awareness to legal theory and practice. In this regard, both explicit and implicit priming of morality appears to have led people in our manipulated conflict of interest to rate their objectivity lower than did members of the control-no COI group, regardless of their actual level of objectivity, as measured by their evaluations of the research institute. More importantly, however, participants in the control-COI group, who indeed were less objective in their evaluations than their peers, reported being less objective compared to participants in the control-no COI group. This finding suggests that members of the control-COI group were aware of their unethical behavior, raising the question of how "good" people who do these "bad things" really are.

Limitations

The findings of this exploratory study suggest several potential implications for theory regarding the interplay between behavioral ethics and law. However, before proceeding with the substantive discussion, we must draw attention to some of the limitations of the study. Naturally, given the exploratory nature of this research, its results should be treated with some skepticism. Nevertheless, our main findings of the effect of the manipulated conflict of interest on participants' evaluations of the research conducted at the institute, and the effectiveness of the different forms (deterrence vs. morality) and methods (explicit vs. implicit) of intervention were replicated both with the 18-item and the binominal questionnaires.

The main limitation arises with regard to the comparison between the intervention through morality and through the likelihood of sanctioning. There is a limit to what can be learned by comparing deterrence and morality when the concepts are manipulated in an online context. The true effect of deterrence is usually measured in the field or in a lab, where the overall bonuses are at stake. In this experiment, because of various IRB restrictions, our threat was relatively mild. This limitation did not apply to the morality-based intervention, which was naturally less problematic from an IRB perspective. Nevertheless, and despite these limitations, the overall greater efficacy of deterrence than that of morality strengthens the results we obtained, which seemed robust across all the experimental conditions. More importantly, the purpose of this study, as stated in "Introduction," was not to compare which intervention is stronger. As a result, findings are limited to the particulars of the designs we have used. Hence, the current study mostly attempts to draw the attention of legal scholars to the need to revisit regulation and enforcement mechanisms in light of the research on behavioral ethics. That being the case, the mere fact that we have found effects and shown some consistency in the effects of certain intervention should lay ground for further research across different contexts as to what types of intervention work better in each context.

Another important limitation that we need to take into account is that this research has only offered a way to look at conflict of interest, but is far from exploring all relevant contexts and factors. Our findings are likely to change given the number of players, social and organizational norms, and many other factors. However, given the difference between this experimental approach to business ethics relative to the traditional dishonesty studies, we hope that more studies will follow this research and explore these and many more factors. Those future studies will help build the needed body of literature to help organizations plan their ethical strategy.

Further, an additional limitation is that our current paradigm did not allow us to measure what effects may have resulted had our participants not been paid: This is because all MTurk studies are paid. Nevertheless, this effect could easily be measured in another study. In any case, this is still a conflict of interest paradigm between people's job requirement and the possibility of participating in a future study. Furthermore, the situation that we attempt to replicate—the revolving door where people perform one task thinking about how it would improve their ability to gain another job—could also be seen as involving not just money but also serving their self-actualization.

Finally, the current work did not attempt to explore the potentially interesting role of individual differences in bounded ethicality, specifically those differences dealing with subtle conflict of interest. Future research should put some effort in this direction and examine demographic variables as gender, education, et cetera, which the current design was underpowered to explore, as well as potential personality variables that could expand our knowledge on the good people paradigm. In this respect, it is worth to mentioning that while the vast majority of our samples did ask to be considered for the additional experiment (and therefore was subject to the manipulated conflict of interest), between 5 and 10% of our participants in both studies did not (and therefore avoided the conflict of interest situation). Interestingly, neither one of the demographics we collected in the current work explained participants' choice on this matter. It would be interesting to identify personality characteristics that may help individuals to better resist conflict of interest situations.

Policy Implications

The design and findings presented above have several normative implications. First, the realization of how little is needed to change people's behavior should be both alarming and comforting for policy makers. We have seen that it is easy to cause people to abandon their objectivity: A subtle promise to hire participants for an additional experiment, which might benefit them with $1, had a substantial effect on their objectivity. Thus, where objectivity is valued, policy makers should think deeply about where promises of that kind should be allowed to occur. This has a great significance to situations such as revolving door conflicts and token gifts.

Second, another important component in the paradigm of "good people doing bad things" is that most people are unaware of their lack of objectivity in their own evaluations. Nevertheless, it seems that with sufficiently strong explicit communication regarding a potential wrongdoing, many people would immediately change their behavior, regardless of the nature of the intervention (e.g., deterrence, morality). This would not lead to a change in the behavior of "bad" people, who would engage in further cost–benefit calculations to assess the wisdom of engaging in bad behavior. In other words, much of the concern with the inability to deter the unaware individual (e.g., Chugh et al. 2005) might prove premature. People might indeed be not fully aware to the unethicality of their behavior, but traditional explicit reminders of both deterrence and morality might be sufficient to cause them to at least start correcting for it.

Third, recognizing the role of traditional regulatory tools in shaping implicit behavior is also relevant to one of the most important regulatory changes in recent years—the BIT (behavioral insight team) revolution, which is based on the influential nudge approach (Alemanno and Sibony 2015; Thaler and Sunstein 2008). BIT advises governments on how to use knowledge from psychology and behavioral economics in shaping people's behavior in socially desirable ways (Feldman and Lobel 2015). Generally speaking, while gaining increasing popularity when it comes to pensions and energy saving BIT has been less dominant in attempting to regulate ethical behaviors. The current BIT approach does not deal with the ability of traditional explicit intervention methods to decrease various automatic processes related to corruption or lack of tolerance and discrimination against minorities. Future research should examine how to combine traditional explicit interventions with implicit interventions when attempting to shape ethical behavior of people in organizations and beyond.

Acknowledgements We thank the Edmond J. Safra Center for the Study of Ethics, Harvard University (Grant No. 10), and the Jerusalem Crime Group for its financial support. We thank Dan Simon, Barak Ariel, Mazarin Banaji and Christoph Engle for their helpful comments.

Appendix

The 18-Item Questionnaire

Please state your agreement or disagreement with the following statements, as objectively as possible, on a scale of 1 (strongly disagree) to 6 (strongly agree).

1. Research conducted by this center is more important than most other research I'm familiar with in the Social Sciences [R].
2. The research done by the center is more valuable than research done by other similar centers [R].
3. Universities should divert funds for this center's research [R].
4. There should be less scrutiny into the actions of members of this center [S].
5. The center's research would provide useful information for the scientific community [R].
6. It would be a valuable use of my time to read research about institutional corruption and how to increase public trust in institutions [R].
7. Mistakes by researchers in this foundation should not be punished as harshly as other researchers [S].
8. Government subsidies for this center are better investments than subsidies for other centers [R].
9. Salaries of researchers in this center should be higher than other researchers' salaries [S].
10. For this question, please answer with the "2" button.
11. Research by this center is crucial for the future success of the international community [R].
12. Governments should divert research funds from other areas to this area [R].
13. International foundations should consider allocating funds to this center [R].
14. Researchers in this center should have greater freedom in how they use public grants [S].
15. Researchers at the Safra Center are more likely to donate to charity than other researchers [S].
16. Researchers at the Safra Center are more concerned with helping people than researchers at other institutions [S].

17. Researchers at the Safra Center are more likely to misuse funds than other researchers [S].
18. Researchers at the Safra Center are less likely to plagiarize work than other researchers [S].

[R] Items focusing on the research conducted by the institute.
[S] Items focusing on the scientists working at the institute.

The Binominal Questionnaire

We would like to ask for your help in rating various statements the Safra Center could potentially use in a future fund-raising campaign. Please indicate whether this statement is accurate/inaccurate, you agree/disagree, would say this statement to potential donors/would not say this statement to potential donors, and would sign a petition containing this statement/would not sign a petition containing this statement.

- Research conducted by the Safra Center is crucial for the well-being of society.

1.	Accurate	Inaccurate
2.	Agree	Disagree
3.	Would say to potential donors	Would not say to potential donors
4.	Would sign a petition	Would not sign a petition

- The Safra Center's research will change the way we look at public institutions.

5.	Accurate	Inaccurate
6.	Agree	Disagree
7.	Would say to potential donors	Would not say to potential donors
8.	Would sign a petition	Would not sign a petition

- The Safra Center's mission is the first attempt ever to deal with one of our most important problems

9.	Accurate	Inaccurate
10.	Agree	Disagree
11.	Would say to potential donors	Would not say to potential donors
12.	Would sign a petition	Would not sign a petition

The Objectivity Questionnaire

1. Do you think you had any sort of influence while you were answering the questions?
   - Yes (if so, please state what you were influenced by)
     ____
   - No
2. Were you completely objective during this study?
   - Yes
   - No (if so, please state why you were not completely objective)
     ____
3. Did you think of any factor besides your best judgment while answering the questions?
   - Yes (if so, please state what else you used)
     ____
   - No

The Explicit Deterrence Manipulation

Many countries have focused on cracking down on people and businesses who act unethically. Those who are involved in multiple interests, and let one of those interests corrupt their actions are especially important targets. Global leaders have decided that such conflict of interest situations are intolerable. Governments around the world took action against hundreds of unethical individuals last week. As a result, both individuals and organizations must be extra cautious when doing business with the government. Otherwise, if they let conflict of interest situations influence their decisions, they will be heavily prosecuted.

In accordance with this worldwide trend, we believe that people who let their conflict of interest affect their objectivity and integrity when completing this survey should be penalized. Hence, participants who let their conflicting interests affect their judgment might lose some of their compensation for the work they do for us.

Who have decided conflict of interest situations are intolerable?

- Everyday people
- Global leaders
- Big business companies

What will happen to people if they let their conflict of interest situations influence their decisions?

- They will receive a warning
- They will be rewarded
- They will be prosecuted

What will happen to participants in this survey if they are influenced by their conflict of interest when completing the survey?

- Their compensation might be affected
- Their reputation might be harmed
- The validity of their answers might be affected

The Explicit Morality Manipulation

Conflict of interest situations are among the greatest problems the world faces today. A conflict of interest occurs when an individual or organization is involved in multiple interests, one of which could possibly corrupt the motivation for an act in the other. Such situations harm the public good, as the correct decision in a national dilemma may be rejected due to these corrupt individuals or organizations. Conflicts of interest also threaten the merit-based system, as individuals are chosen based on who they know, not what they know. These actions are immoral, so conscientious individuals should do everything in their power to avoid conflict of interest situations.

In accordance with this worldwide trend, we believe that people who let their conflict of interest affect their objectivity and integrity when completing this survey are not acting in a moral and ethical way. Hence, participants who will let their conflicting interests affect their judgment might harm the public good.

What should conscientious individuals do in regard to conflict of interest situations?

- Avoid them
- Seek them out
- Take advantage of them

What do conflict of interest situations harm?

- A person's feelings
- The environment
- The public good

What will happen to participants in this survey if they are influenced by their conflict of interest when completing the survey?

- They might harm the public good
- Their integrity might be harmed
- The validity of their answers might be affected

The Implicit Deterrence Manipulation

c_ _ _uption	corruption
jai_	jail
poli_ _	police
punish_ _ _t	punishment
fin_	fine
_ubpoena	subpoena
jud_e	judge
in_ictm_nt	indictment
in_ _st_gat_on	investigation
br_be	bribe
uilt	guilty
cro_ _ing	crossing
rotat_ _ _	rotation
_ miling	smiling
s_ll	sill
fi_ _y	fiery
flou_	flour
b_ld	bald
r_ _t	root
fe_er	fever
w_ _ds	weeds
fema_ _	female
_ _ gineer	engineer
al_gn	align
d_sconn_ _ted	disconnected
catal_ _	catalog
_ orn	corn
mer_ e	merge
fantast_ _	fantastic
_uman	human
exc_ll_nt	excellent
cop_ _r	copier
tra_	trap
bl_e	blue
effic_ _nt	efficient

The Implicit Morality Manipulation

integri_ _	integrity
_rust	trust
mor_li_y	morality
hon_sty	honesty
objectivi_ _	objectivity
princi_ _es	principles
_irtue	virtue
t_ _th	truth
_ _irness	fairness
neut_ali_ _	neutrality
jus_ic_	justice
cro_ _ing	crossing
rotat_ _ _	rotation
_ miling	smiling
s_ll	sill
fi_ _y	fiery
flou_	flour
b_ld	bold
r_ _t	root
fe_er	fever
w_ _ds	weeds
fema_ _	female
_ _ gineer	engineer
al_gn	align
d_sconn_ _ted	disconnected
catal_ _	catalog
_ orn	corn
mer_ e	merge
fantast_ _	fantastic
_uman	human
exc_ll_nt	excellent
cop_ _r	copper
tra_	trap
bl_e	blue
effic_ _nt	efficient

References

Achtziger, A., Alós-Ferrer, C., & Wagner, A. K. (2015). Money, depletion, and prosociality in the dictator game. *Journal of Neuroscience, Psychology, and Economics, 8*(1), 1.

Adams, J. S., Tashchian, A., & Shore, T. H. (2001). Codes of ethics as signals for ethical behavior. *Journal of Business Ethics, 29*(3), 199–211.

Alemanno, A., & Sibony, A. L. (2015). *Nudge and the law: A European perspective*. London: Bloomsbury Publishing.

Ayal, S., & Gino, F. (2011). *Honest rationales for dishonest behavior. The social psychology of morality: Exploring the causes of good and evil*. Washington, DC: American Psychological Association.

Banaji, M. R., & Greenwald, A. G. (2013). *Blindspot: Hidden biases of good people*. New York, NY: Delacorte Press.

Bazerman, M. H., & Tenbrunsel, A. E. (2011). *Blind spots: Why we fail to do what's right and what to do about it*. Princeton, NJ: Princeton University Press.

Bereby-Meyer, Y., & Shalvi, S. (2015). Deliberate honesty. *Current Opinion in Psychology, 6,* 195–198.

Bersoff, D. M. (1999). Why good people sometimes do bad things: Motivated reasoning and unethical behavior. *Personality and Social Psychology Bulletin, 25*(1), 28–39.

Buhrmester, M. D., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science, 6,* 3–5.

Cain, D. M., Loewenstein, G., & Moore, D. A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies, 34*(1), 1–25.

Camerer, C. F., & Hogarth, R. M. (1999). The effects of financial incentives in experiments: A review and capital–labor–production framework. *Journal of risk and uncertainty, 19*(1-3), 7–42.

Che, Y. K. (1995). Revolving doors and the optimal tolerance for agency collusion. *The Rand Journal of Economics, 26*(2), 378–397.

Chugh, D., Bazerman, M. H., & Banaji, M. R. (2005). Bounded ethicality as a psychological barrier to recognizing conflicts of interest. In D. A. Moore, D. M. Cain, G. Loewenstein, & M. H. Bazerman (Eds.), *Conflict of interest: Challenges and solutions in business, law, medicine, and public policy* (pp. 74–95). New York, NY: Cambridge University Press.

Cornaggia, J., Cornaggia, K. J., & Xia, H. (2016). Revolving doors on wall street. *Journal of Financial Economics, 120*(2), 400–419.

Craswell, R., & Calfee, J. E. (1986). Deterrence and uncertain legal standards. *Journal of Law, Economics and Organization, 2*(2), 279–303.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–688.

Evans, J. S. B. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7*(10), 454–459.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59,* 255–278.

Feldman, Y. (2009). The expressive function of the trade secret law: Legality, cost, intrinsic motivation and consensus. *Journal of Empirical Legal Studies, 6*(1), 177–212.

Feldman, Y. (2011). The complexity of disentangling intrinsic and extrinsic compliance motivations: Theoretical and empirical insights from the behavioral analysis of law. *Washington University Journal of Law and Policy, 35,* 11–52.

Feldman, Y. (2014). Behavioral ethics meets behavioral law and economics. In Zamir, E., & Teichman, D. (Eds.), *Oxford handbook of behavioral law and economics* (pp. 213–241). Oxford University Press.

Feldman, Y., Gauthier, R., & Schuler, T. (2013). Curbing misconduct in the pharmaceutical industry: Insights from behavioral ethics and the behavioral approach to law. *The Journal of Law, Medicine and Ethics, 41*(3), 620–628.

Feldman, Y., & Lobel, O. (2015). Behavioral trade-offs: Beyond the land of nudges spans the world of law and psychology. In Alemanno, A. & Sibony, E. (Eds.), *Nudge and the law: A European perspective*. Oxford: Hart Publishing.

Friedberg, M., Saffran, B., Stinson, T. J., Nelson, W., & Bennett, C. L. (1999). Evaluation of conflict of interest in economic analyses of new drugs used in oncology. *JAMA, 282*(15), 1453–1457.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*(4), 650–669.

Gino, F., & Desai, S. D. (2012). Memory lane and morality: How childhood memories promote prosocial behavior. *Journal of Personality and Social Psychology, 102*(4), 743–758.

Gino, F., Schweitzer, M., Mead, N., & Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes, 115*(2), 191–203.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, *25*(4), 191–209.

Gormley Jr, W. T. (1979). A test of the revolving door hypothesis at the FCC. *American Journal of Political Science*, *23*(4), 665–683.

Halali, E., Bereby-Meyer, Y., & Meiran, N. (2014). Between self-interest and reciprocity: The social bright side of self-control failure. *Journal of Experimental Psychology, 143,* 745–754.

Halali, E., Bereby-Meyer, Y., & Ockenfels, A. (2013). Is it all about the self? The effect of self-control depletion on ultimatum game proposers. *Frontiers in Human Neuroscience, 7,* 240.

Hillman, A. L. (1987). Financial incentives for physicians in HMOs. Is there a conflict of interest? *The New England Journal of Medicine, 317*(27), 1743–1748.

Hollis, J. (2008). *Why good people do bad things: Understanding our darker selves*. New York, NY: Gotham Books.

Jolls, C., Sunstein, C. R., & Thaler, R. (1998). A behavioral approach to law and economics. *Stanford Law Review*, *50*(5), 1471–1550.

Kahneman, D. (2011). *Thinking, fast and slow*. London: Macmillan.

Kasser, T., & Ryan, R. M. (1996). Further examining the American dream: Differential correlates of intrinsic and extrinsic goals. *Personality and Social Psychology Bulletin, 22*(3), 280–287.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review, 118*(1), 97–109.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2,* 151–160.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480.

Lessig, L. (2011). *Republic, lost: How money corrupts congress—and a plan to stop it*. New York, NY: Hachette Digital Inc.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*(6), 633–644.

Mead, N., Baumeister, R. F., Gino, F., Schweitzer, M., & Ariely, D. (2009). Too tired to tell the truth: Self control resource depletion and dishonesty. *Journal of Experimental Social Psychology, 45,* 594–597.

Moore, D. A., & Loewenstein, G. (2004). Self-interest, automaticity, and the psychology of conflict of interest. *Social Justice Research, 17*(2), 189–202.

Moore, D. A., Tanlu, L., & Bazerman, M. H. (2010). Conflict of interest and the intrusion of bias. *Judgment and Decision Making, 5*(1), 37–53.

Norenzayan, A., & Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science, 322*(5898), 58–62.

Paolacci, G., & Chandler, J. (2014). Inside the turk understanding mechanical turk as a participant pool. *Current Directions in Psychological Science, 23*(3), 184–188.

Pillutla, M. M. (2011). When good people do wrong: Morality, social identity, and ethical behavior. In D. De Cremer, R. van Dijk, & J. K. Murnighan (Eds.), *Social psychology and organizations* (pp. 353–370). New York, NY: Routledge.

Pittarello, A., Leib, M., Gordon-Hecker, T., & Shalvi, S. (2015). Justifications shape ethical blind spots. *Psychological Science*, *26*(6), 794–804.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*(7416), 427–430.

Rodwin, M. A. (1989). Physicians' conflicts of interest: The limitations of disclosure. *New England Journal of Medicine, 321*(20), 1405–1409.

Rodwin, M. A. (2012). Conflicts of interest, institutional corruption, and pharma: An agenda for reform. *The Journal of Law, Medicine and Ethics, 40*(3), 511–522.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300,* 1755–1758.

Schwartz, M. S. (2002). A code of ethics for corporatecode of ethics. *Journal of Business Ethics, 41*(1–2), 27–43.

Schweitzer, M. E., & Hsee, C. K. (2002). Stretching the truth: Elastic justification and motivated communication of uncertain information. *Journal of Risk and Uncertainty, 25,* 185–201.

Sezer, O., Gino, F., & Bazerman, M. H. (2015). Ethical blind spots: Explaining unintentional unethical behavior. *Current Opinion in Psychology, 6,* 77–81.

Shalvi, S., Dana, J., Handgraaf, M. J. J., & De Dreu, C. K. W. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes, 115,* 181–190.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological Science, 23,* 1264–1270.

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications doing wrong and feeling moral. *Current Directions in Psychological Science, 24,* 125–130.

Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in god. *Journal of Experimental Psychology: General, 141,* 423–428.

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences, 109*(38), 15197–15200.

Somers, M. J. (2001). Ethical codes of conduct and organizational context: A study of the relationship between codes of conduct, employee behavior and organizational values. *Journal of Business Ethics, 30*(2), 185–195.

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology, 37*(10), 1660.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah. NJ: Erlbaum.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(5), 645–665.

Stapenhurst, R., & Kpundeh, S. J. (Eds.). (1999). *Curbing corruption: Toward a model for building national integrity*. Washington, DC: World Bank Publications.

Stevens, B. (1994). An analysis of corporate ethical code studies: "Where do we go from here?". *Journal of Business Ethics, 13*(1), 63–69.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions using the architecture of choice*. New Haven, CT: Yale University Press.

Uziel, L., & Hefetz, U. (2014). The selfish side of self-control. *European Journal of Personality, 28*(5), 449–458.

Weaver, G. R. (1995). Does ethics code design matter? Effects of ethics code rationales and sanctions on recipients' justice perceptions and content recall. *Journal of Business Ethics, 14*(5), 367–385.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60*(2), 181–192.

Xu, H., Bègue, L., & Bushman, B. J. (2012). Too fatigued to care: Ego depletion, guilt, and prosocial behavior. *Journal of Experimental Social Psychology, 48*(5), 1183–1186.

Zamir, E., & Sulitzeanu-Kenan, R. (2016). Explaining Self-Interested Behavior of Public-Spirited Policymakers (November 28, 2016). Hebrew University of Jerusalem Legal Research Paper No. 17–8. Available at https://ssrn.com/abstract=2876437.