Can we Regulate "Good" People in Subtle Conflicts of Interest Situations

Yuval Feldman[1] and Eliran Halali[2]

Growing recognition in both the psychological and management literature of the concept of "good people" has caused a paradigm shift in our understand of wrongful behavior: wrongdoings that were previously assumed to be choice-based, conscious, and deliberate decisions, are often the product of intuitive processes that prevent people from recognizing the wrongfulness of their behavior. This process has been dubbed by several leading scholars as an ethical blind spot. This study will explore the main implications of the good people paradigm on how to . We examined, experimentally, using two m-Turk studies, 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 mind-set (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 create a risk to the integrity of individuals, but they also suggest that a modest explicit/deliberative intervention can easily remedy much of the wrongdoing.

Regulating ethicality, conflict of interest, deterrence and legitimacy, behavioral ethics

Introduction

Traditionally, states and organizations have used their powers to prevent people from engaging in

wrongful conducts, breaching contracts, and eschewing their duties. The assumption of most theoretical

accounts of both law and management has been that people willingly choose to engage in these behaviours, and that certain techniques, most notably incentives, can be used to change such decisions[3]. Most 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. In contrast, many theories of the behavioural approach to human judgment and decision-making have challenged the basic assumptions of the neo-classical economic doctrine of rational choice[4]. Among these, the literature related to the rising role of non-deliberative choice in people's behaviour stands as a central and dominant alternative[5]. This "behavioural" approach has gained prominence in an increasing number of "applied" fields, where discussion of automaticity, implicit vs. explicit attitudes, responsibility, self-awareness, and motivated reasoning is gaining increasing recognition. A common theme in these paradigms is the view that many of the undesirable behaviours, traditionally the focus of prevention using rational choice mechanisms, have to do with "good" people who do not necessarily engage in any deliberative process before doing a "bad" action. Therefore, the ability of current explicit mechanisms to curb unethical behaviour may be limited. Thus for example, fines, damages or organizational sanctions will not cause people to stop doing what

---

[3]   See Gneezy, U, Meier S,, Rey-Biel P. (2011). When and why incentives (don't) work to modify behavior: *The Journal of Economic Perspectives*: 191-209, for a review and discussion of the contexts in which incentives are less effective. See also Camerer, Colin 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.

[4]   See Feldman, Y. (2011). For Love Or Money? Defining Relationships In Law And Life: The Complexity of Disentangling Intrinsic and Extrinsic Compliance Motivations: Theoretical and Empirical Insights from the Behavioral Analysis of Law.: *Wash. UJL & Pol'y* 35, 11-547.

[5]   The concept of two systems of reasoning has gained popular recognition in Kahneman, Daniel. *Thinking, fast and slow*. Macmillan, 2011. This concept lies at the core of much of the research in behavioral law and economics. The general concept differentiates between an automatic, intuitive, and mostly unconscious process (labeled System 1) and a controlled and deliberative process (labeled System 2) (see also Stanovich, Keith E., and Richard F. West. "Advancing the rationality debate."*Behavioral and brain sciences* 23.05 (2000): 701-717; Evans, Jonathan St BT. "In two minds: dual-process accounts of reasoning."*Trends in cognitive sciences* 7.10 (2003): 454-459.

they do not perceive as an ethical problem. If indeed most unethical behaviour is attributed to such moral blind spots, it may be necessary to find new ways to change people's behaviour and their implicit judgments.

The rise of good people

The focus on "good people" represents the growing recognition that many ethical decisions are the result of implicit rather than explicit choices and affect normative citizens. Simply reviewing the titles of current papers shows how central the theme has become.[6] This theme of good people suggests a growing recognition that many ethically relevant behaviours 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 (Chugh, Bazerman & Banaji, 2005; Terbenusel et al., 2010; Banaji and Greenwald, 2013).[7] Given the growing focus on good people in psychology and management, the dearth of discussions of its implications on law and regulation is quite puzzling (see Feldman, 2014 for a review). The main question this study will attempt to explore is what are the implications of this literature to questions of legal enforcement policy. While the classical debate in enforcement strategies was usually

---

[6] For example, Mazar, Nina M, On Amir O, and Dan Ariely D. (2008). The dishonesty of honest people: A theory of self-concept maintenance:." *Journal of marketing research* 45.6, (2008): 633-644. Also, Bersoff, D,avid M. (1999). Why good people sometimes do bad things: Motivated reasoning and unethical behavior:." *Personality and Social Psychology Bulletin* 25.1 (1999),: 28-39.
Pillutla, Madan M. P. (2011). "When good people do wrong: Morality, social identity, and ethical behavior:." *Social Psychology and Organizations* 353 (2011 ; Hollis, James. (2008). Why good people do bad things: Understanding our darker selves. Penguin, 2008; Banaji, Mahzarin R. B., and Anthony G. Greenwald' A,G.. (2013). Blindspot: Hidden biases of good people. Delacorte Press, 2013, note that the "good people" scholarship is usually different from the type of research conducted by Zimbardo on the Lucifer effect [Zimbardo, Philip Z.. (2007). "The Lucifer effect: Understanding how good people turn evil." New York (2007)]. These works generally try to explain how ordinary people end up doing evil or engage in gross criminal behavior.

[7] For a recent review see Sezer, Ovul, Francesca Gino, and Max H. Bazerman. "Ethical blind spots: explaining unintentional unethical behavior." *Current Opinion in Psychology* 6 (2015): 77-81.

related to comparing deterrence and morality (for a review see Feldman, 2011), the good people paradigm might require a new way of dealing with wrongdoing. The current focus in the literature on both moral suasion and deterrence assumes deliberative choice . Hence, the growing research of the role of automaticity might demand that management and legal policy makers adapt their enforcement strategies to these new views of human agency.

Dual reasoning and Ethical Decision Making

A deeper understanding of good people can be achieved based on the concept of dual reasoning, the assumption of two systems of reasoning. This gained popular recognition in Kahneman's book, Thinking Fast and Slow (2011). The general concept, which stands at the core of much of the research in behavioural 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 Stanovitch and West, 2000; Evans, 2003). 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 behavioural economics (e.g., Halali, Bereby-Meyer, and Ockenfels, 2013; Halali, Bereby-Meyer, and Meiran, 2014; Sanfey, Rilling, Aronson, Nystrom, and Cohen, 2003) and behavioural law and economics (e.g. Jolls, Sunstein, and Tahler, 1998), and it is the foundation for a new understanding and approach to self-interest (see Gigerenzer and Goldstein, 1996).

Theoretical foundations of Implicit Corruption

It has been suggested that moral judgments and decisions are the result of reasoning and deliberation. While self-interest is an automatic primary motive that needs to be constrained by appropriate inhibitory mechanisms (e.g., Moore and Loewenstein, 2004). Moore et al. (2010) showed that people truly believe their own biased judgments, with limited ability to recognize that their behaviour was

affected by self-interest.[8] A similar view was advanced by Gino et al. (2011), who showed that the level

of control needed to behave ethically is much higher than the level needed to behave unethically. Similar

support of this view could be seen in the work of Shalvi et al. (2012) on the idea that honesty requires

time, demonstrating that behaving ethically requires more deliberative resources. Similarly, Halali et al.

(2013) have showed a similar affect with regard to fairness considerations in dictator games settings,

which seem to be much more deliberative than self-interest considerations which are more intuitive. An

additional example of the automatic effect of self-interest on behaviour may be seen in the concept of

implicit egotism. In contrast to the automaticity of self-interest, which focuses on what is good for the

decision maker, egotism focuses on the tendency of people to appreciate and unconsciously prefer things

they associate with themselves even when the outcome doesn't include personal benefit. Egotism can be

associated with various personal details such as individual's name and even items to which one was

exposed to in the past (Nuttin, 1987; Jones et al., 2002, 2004). While indeed the literature in this area is

still debated (e.g. Cushman, Young, and Hauser theoretically, 2006), Clearly from an applied perspective,

the behaviour of "good people" attest to the need of understanding implicit corruption (see also Lessig,

2011, Feldman, Gauthier & Schuler, 2013).

Behavioural ethics and the enforcement of unethicality in the workplace and beyond

The evidence regarding the automaticity of self-interest suggests that when dealing with unethical

behaviour, a new approach is required in order to create effective enforcement methods, especially

regarding people in situations of conflict of interest. The ability of incentives and deterrents to affect non-

deliberate behaviour has been discussed by scholars such as Bazerman and Terbnusel (2011), who suggest

that "such measures simply bypass the vast majority of unethical behaviours that occur without the

---

[8] The measurement of "private" evaluations was done by giving participants incentives to be accurate in their predictions.

conscious awareness of the actors, who engage in them". Following the same line of thought, psychologists who study ethical decision-making have challenged the classic enforcement approach that focuses on external measures and incentives to control unethical behaviour, as such an approach provides an unjustified central role to self-control, autonomy, and responsibility for action (Banaji & Greenwald, 2013).

Implicit interventions

The ability of states and organizations to modify the behaviour of their employees is not limited only to explicit intervention. Much of the research on priming has shown the ability of primes to lead to behavioural changes. For example, Money as a priming tool has been proven effective both methodologically and theoretically in experiments conducted by many scholars examining both automatic reasoning in the face of financial self-interest and subtle curbing methods. In a series of nine experiments, Vohs et al. (2006, 2008) demonstrated how the behaviour of people in areas of cooperation and social interaction changes entirely with exposure to subtle reminders of money (for a related study, see Kay et al., 2004). In comparing the effect of morality and money, Aquinno et al. (2009) found that financial incentives had a stronger effect on such behaviours as contribution to public causes and dishonesty than did stimuli such as the Ten Commandments.[9] Kouchaki et al. (2013) demonstrated how the inclination to engage in unethical behaviour is increased by priming with money. These studies show the relevance of priming to the understanding of automatic human behaviour and to the law.

The Current Study:

The above arguments lie at the basis of the current study. If indeed much of the corruption of people could be explained by automatic processes, what is the ability of existing enforcement strategies to curb such corruption? Is shaping people's morality unlikely to change their behaviour? Hence, if indeed

---

[9] Amir, Mazar, and Ariely (2008) provided an earlier demonstration of the effect of the Ten Commandments.

people are unaware of the problematic nature of their behaviour, is the explicit request to behave in an ethical way, relevant?

We designed two studies to examine how people behave in subtle conflict of interests under an intuitive/automatic vs. analytical/deliberative mind-sets and then tested which enforcement mechanisms work best. The research focuses on conflict of interest because it lies at the core of most unethical behaviours in the public and private sectors (Tyler, 2005; Dana et al., 2006). We used subtle conflicts of interest in which the contrast between self-interest and professional integrity is somewhat ambiguous, rather than actual bribes because these leave less room for implicit corruption. This division appears in the literature on conflict of interest that discusses both the more visible and obvious conflict of interest and the more subtle one (Moore, Tanlu & Bazerman, 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, Bazerman & Banaji, 2005).

From understanding to intervening

While the Behavioral ethics literature focused much of its attention on what mechanisms are responsible for unethical behaviors, Chugh, Bazerman & Banaji, (2005) argue that incentives and similar concepts fail to correct a large portion of unethical behaviours, because "such measures simply bypass the vast majority of unethical behaviours that occur without the conscious awareness of the actors, who engage in them." The second study reported in this paper have focused on this question. We examined how material interests imperceptibly effect decision-making , and the effectiveness of negating self-interests in conflict with organizational duty through classical and new intervention approaches.

Awareness constitutes a powerful medium for change when dealing with subtle conflict of interest and unethical behaviour at large. The legal scholarship in particular must better understand the

scope and potential effect of awareness in order to craft appropriate legal enforcement mechanisms. Currently, the research literature is still far from reaching a consensus on how aware people are of the motivation of their (un)ethical behaviour and how effectve explicit classical enforcement mechanisms are in situations of subtle conflict of interest.

For example, the medical field is a fertile ground for conflicts of interest. A frequent example is clinical studies financed by pharmaceutical companies, providing an incentive for physician-researchers to reach certain results (e.g. Rodwin, 2012; Friedberg et al., 1999; Rodwin, 1989; Hilman, 1987; Feldman et al., 2013). The transition of professionals from the public to the private sector is a common phenomenon (Che, 1995). This process, referred to as the "revolving doors," presents a possible conflict of interest if officials abuse their public position for the purpose of personal benefits within the private sector.[10] A prime example of such concern is the hypothesis that anticipation of future opportunities in regulated firms may cause regulators to be less aggressive in administering regulatory policy (e.g., Che, 1995 at p. 379). 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.

Although in various situations "good people" may not be affected by incentives, it is not clear to what likelihood people will ignore entirely the existence of incentives. For example, even if the process of self-deception might block people's full awareness to the unethicality of their own behaviour, it is possible to predict that introducing sanctions will cause people to be more aware of their behaviour and lessen the effect of automaticity. Therefore, traditional intervention techniques that target awareness should

---

[10] See General Motors Corp. v. City of New York, 5oi F.2d 639, 65o n.20 (2d Cir. I974); D.C. BAR COMM. ON LEGAL ETHICS, INQUIRY No. I9 (Tent. Draft I976).

not be disregarded but rather re-examined in light of the understanding of the literature on behavioural ethics.

Thus, the question of how policy makers can prevent misbehaviour that is neither willful nor intentional, both at the state level and at the organizational level, is still an open both from a theoretical and from an applied perspective.

For that reason, the second study focused on comparing the efficacy of two common intervention practices used by the authorities: *deterrence,* which serves as a traditional function of the law, and where extrinsic motivation is used to shape behaviour; and *morality,* which focuses on changing people's intrinsic motivation. Extrinsic motivation refers to actions driven by external commands or incentives. Extrinsic motivation can be achieved through targeting the potential offender's financial status through rewards and fines. Conversely, intrinsic motivation is linked to behaviour driven from within the individual, usually out of a sense of moral or civic duty,[11] and it is affected by targeting the sense of morality (see Feldman, 2009, 2011). In our research we provide a first look into the ability to curb the behaviour of "good" people in subtle conflict of interest. The focus on people in subtle conflict 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 behaviour is unethical. Clearly, in obvious situations, where people are often bribe to act against their duty of loyalty to the state or to the corporation, the now common "blind spot" argument is less likely to occur.

---

[11] 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 wellbeing associated with focusing on extrinsic and intrinsic goals.

Design

The case study we focus on deals with people in a subtle conflict of interest, where the gap between ethical demands and self-interest is minimal. Many questions in this field remain open, as most of the new research on decision-making and on behavioural 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 determining which intervention method legal policy-makers should focus on in various context of interest.

To answer these questions, we have designed two studies that focus on people's behaviour in conflict of interest situations. In the first study, we have focused on understanding the process through which conflict of interest might affect people more, in intuitive/automatic or analytical/deliberative mind-sets. In the second study we focused on what are the best intervention methods to curb such behaviours, deterrence or morality, explicitly or implicitly.

In both studies we have created a conflict of interest situation by asking participants to provide their *objective* evaluation of the need to fund and help a research institute, 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. When participant provided a biased evaluation of the research institution, relative to a control group who had no financial interest in the evaluation, we can point on deviation from objective and neutrality. Clearly even such behaviors could not be defined as corrupted or unethical. However, as suggested in the introduction, we have intentionally focused on such contexts as a way to account for the Blind Spot argument. Participants were then assigned to a few randomized groups where they were informed by the experimenter, implicitly and 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 males, 47 females; 74 White, 7 Black, 4 Hispanic, 10 Asian, 4 Other) completed the experiment online, through Amazon Mechanical Turk (mTurk) in exchange for 1$ (see Buhrmester, Kwang & Gosling (2011) for a full description of Mechanical Turk sampling). We excluded responses from one participant who attempted to complete the study multiple times[12]. All participants signed an informed consent form before participating in the study. We randomly assigned participants to one of two experimental mind-set conditions (intuitive/analytical).

Materials and Procedure. After signing the consent form, participants went through the 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 an 18-item questionnaire in which they had to provide an objective evaluation of the need to fund and help the described research institute, focusing both on the research conducted at the center and on the scientists working there. After completing the questionnaire, participants received a set of three statements about the research institute. Each statement was followed by four binominal questions assessing participants' agreement with the statement and their "willingness to pay" for that support. Next, participants were presented

---

[12] These responses was identified based on duplicated IP addresses and GPS locations.

with three yes/no questions assessing their sense of objectivity with regard to the research institute when answering the questionnaires. Finally, participants answered a demographic questionnaire.

Mind-set manipulation. Relying on Shenhav et al. (2012; also used by Rand et al., 2012), we manipulated mind-set 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 mind-set) or *carefully reasoning through a situation* (i.e., analytical/deliberative mind-set) led them in the right direction and resulted in a good outcome. Prior to the 18-item questionnaire, participants were asked to rely on intuition/reasoning when making their responses. Additionally, in the analytical condition we told the participants that, following their responses, they will be required to describe in writing the reasons for their responses, in order to further strengthen their reliance on reasoning (Wilson & Schooler, 1991).

The conflict of interest. We created a potential for conflict of interest (COI) by telling participants that if their answers to the current survey demonstrate interest in the research conducted at the center, they might be selected for an additional survey on that topic with higher pay. Following this statement, we asked participants to indicate whether they would like to be considered for this additional experiment.[13]

The 18-item questionnaire. The 18-item questionnaire included 9 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 8 items that focused on the scientists

---

[13] Eleven participants indicated that they did not want to be considered for the additional experiment (5 in the intuitive mindset condition, and 6 in the analytic mindset condition). We excluded these participants from further analysis as if they did not want to be considered for the additional experiment most likely they did not experienced a 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.

working at the institute, such as "The salaries of scientists at this center should be higher than other scientists' salaries.[14]" Participants were asked to indicate their answers on a scale of 1 (strongly disagree) to 6 (strongly agree). Agreement on all items indicated an evaluation favorable to the research institute.

The binominal questionnaires. The three statements in the binominal questionnaire were as follows: (a) "Research conducted by the Safra Center is crucial for the wellbeing 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: (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 for the survey completion in the intuitive and the analytic mindset conditions was 11:33 and 12:41 minutes, respectively. We excluded three outlier participants (two in the intuitive mindset condition and one in the analytic mindset condition) from all analysis since their completion time (41 and 46 minutes in the intuitive condition and 44 in the analytic condition) was more than three

---

[14] To make sure participants read each item before answering, one item (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. Only one participant missed the answer. All participants respond right on this item.

standard-deviations away from the mean in their condition (Z values were 3.26, 3.82, and 4.35, respectively).[15]

The 18-item questionnaire. For each participant we calculated a separate mean score for the research-related items and for the scientist-related items. Cronbach's alpha reliability of these items were .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, $F_{(1, 82)} = 86.88$, $p < .001$, $\eta^2_p = .51$, 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, as expected, participants' mean evaluation in the intuitive (M = 3.81, SD = .83) compared to the analytic (M = 3.44, SD = .86) mindset condition was significantly more positive, $F_{(1, 82)} = 3.99$, $p < .05$, $\eta^2_p = .05$, with no Condition × Item issue interaction ($F<1$, n.s.).

The binominal questionnaires. The Kuder-Richardson Formula 20 reliability of the 12 items in the binominal questionnaires 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 $F_{(1, 85)} = 1.23$, $p = .27$, $\eta^2_p = .02$.

---

[15] The pattern of the following reported results was similar when all the total of 15 excluded participants in this experiment was included in the analysis.

The objectivity questionnaire. Because the objectivity questionnaire included only three binominal items, we entered the meaning of the participants' answer (0: objective, 1: nonobjective) into a generalized probit estimation equation for binominal data, with Condition (depletion, no-depletion) as a between-participant independent variable, Question (1, 2, 3) as within- participant 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 behaviour 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 were not significant: all *Wald χ^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 are less objective.

Thus, the results of the first study show that lack of deliberation is related not just to dishonesty as much of the literature has focused on, but also to behavior in subtle conflict of interest situations and hence extend the findings of the literature on behavioural ethics. Intuitive/automatic mind-set seems to make it more likely that people will provide favorable reviews toward the Safra center relative to the participants who were put in the analytical/deliberative mind-set. However, this study doesn't inform us on who should change people's behaviour, given that we are more like to behave in a non-objective way in an intuitive mind-set, an explicit intervention or an implicit one?

Experiment 2

Participants. Three hundred and twenty participants (165 males, 155 females; 244 White, 24 Black, 15 Hispanic, 24 Asian, 13 Other) completed the experiment online through Amazon Mechanical Turk (mTurk) in exchange for 1$. 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 6 groups:

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 conflict of interest with no intervention (n=56)

6. Control group: no conflict of interest and no intervention (n=49)

Materials and 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 questionnaires, objectivity questionnaire, and demographic questionnaire as in Experiment 1. In the control group, participants answered the set of questionnaires immediately after they read the paragraph describing the research institute; in the control conflict of interest with no intervention, they answered the questionnaires after being exposed to the description of the conflict.

The conflict of interest[16]. We have created a potential for conflict of interest (COI) using the same way as in Experiment 1.[17] *Explicit deterrence.* We manipulated explicit deterrence by asking participants to read a paragraph on government crackdown and stating that participants might lose some of their compensation if their answers reveal conflicting interest. Next, we asked participants to answer a three-item questionnaire to verify their understanding of the explicit deterrence intervention.[18]

Explicit morality. We manipulated the 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 participants would harm the public good if they answered our questions based on their self-interest. Next, we asked participants to answer a three-item questionnaire to verify their understanding of the explicit morality intervention.[19]

Implicit deterrence. We manipulated the implicit deterrence using a 36-items word completion test in which 12 of the items were words related to deterrence (e.g. punishment, subpoena, indictment)

---

[16] In the original design, we have included a different type of conflict of interest (identity-based conflict of interest) and we have measured its effect on participants. Since there was no effect of the identity-based interest on participants (as compared to the control no COI condition, there was no point in including these conditions in the paper, so we focused only on the money-based conflict of interest conditions.

[17] Fifteen participants in the COI conditions indicated that they did not want to be considered for the additional experiment (3 in explicit deterrence, 4 in explicit morality, 3 in implicit deterrence, 2 in implicit morality, and 3 in no intervention). We excluded these participants from further analysis as if they did not want to be considered for the additional experiment most likely they did not experienced a 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.

[18] Three participants in the explicit deterrence intervention failed to answer the three deterrence comprehension items correctly. We excluded these participants from further analysis.

[19] Nine participants in the explicit morality intervention failed to answer the three morality comprehension items correctly. We excluded these participants from further analysis.

in order to prime participants with deterrence.[20] Methodologically in priming the targets words we have

followed a rationale similar to that of used in other studies on priming words to induce a state of mind

(e.g. Srull and Wyer; 1979; Norenzayan & 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.

Implicit morality. The implicit morality intervention was the same as the implicit deterrence

intervention except that the 12 prime words were related to morality (e.g. integrity, morality, honesty)

rather than to deterrence. [21]

The 18-item questionnaire. The 18-item questionnaire was the same as in Experiment 1.[22]

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.

---

[20] We excluded two participants from further analysis because they failed to identify six or more of the 12 morality prime words.

[21] We excluded from any further analysis two participants who failed to identify six or more out of the 12 deterrence prime words.

[22] 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.

Results and discussion

Table 3 displays the average time for survey completion in all experimental conditions. We excluded three outlier participants (one in the explicit deterrence condition, one in the explicit morality, and one in the Control COI-no intervention condition) from all analysis since their completion time (48, 45 and 188 minutes, 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). [23]

The 18-item questionnaire. For each participant we calculated a separate mean score for the research-related items and for the scientist-related items. Cronbach's alpha reliability of these items were .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, $F_{(1, 280)}$ = 443.58, p < .001, η^2_p = .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, η^2_p = .09, as was the Condition × Item issue interaction, $F_{(5, 280)}$ = 3.15, p < .01, η^2_p = .05.

As can be seen in Figure 1, subsequent analyses of the Condition × Item issue interaction revealed that participants' evaluations of the research conducted at the institute in the Control COI-no intervention condition and the two implicit intervention conditions (deterrence, morality) were significantly higher than those of the Control-no COI condition and the two explicit interventions conditions (deterrence, morality), $F_{(1,462)}$ = 20.52, p < .001, η^2_p = .07. These results indicate that the opportunity to earn extra money

---

[23] The pattern of the following reported results was similar when all the total of 34 excluded participants in this experiment was included in the analysis.

by participating in another experiment of the research institute caused participants to be more favorably disposed toward the research conducted at the institute. 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 these 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-no intervention condition.

In contrast to the effect of the COI-on participants' evaluations of the research (as illustrates by the difference between the Control COI-no intervention 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-no intervention condition were not significantly different from the Control-no COI condition ($F < 1$, n.s.).[24]

The binominal questionnaires. The Kuder-Richardson 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, $F_{(5, 280)} = 2.31$, $p < .05$, $\eta^2_p = .04$. Subsequent analysis revealed similar results to those observed in the analysis of the participants' evaluations of the research conducted at the institute, as measured by the 18-item questionnaire. Specifically, participants' favoritism toward the research institute in the COI-no intervention condition and

---

[24] Interestingly, as can be seen in Figure 1, while participants' evaluations of the scientists in the COI conditions with implicit interventions (deterrence, morality) did not differ from the two control groups (no COI, COI-no intervention), 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-no intervention, implicit deterrence, implicit morality), $F_{(1,280)} = 21.81$, $p < .001$, $\eta^2_p = .07$. This pattern of results suggests a chilling effect (Ferguson and Peters, 2000; Calfee and Craswell, 1986), that is, even more "objective" evaluations than in the control group, where there was no conflict.

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, η^2_p = .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 of the research institute 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: nonobjective) into a generalized probit estimation equation for binominal data, with Condition (depletion, no-depletion) as a between-participant independent variable, Question (1, 2, 3) as within- participant 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 behaviour 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 for the condition effect revealed that in the Control COI-no intervention condition and the two morality intervention conditions (explicit, implicit) more participants indicated nonobjective behaviour than in the Control-no COI condition and 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, in the implicit

morality condition participants were less objective comparing the Control-no COI (more in favor of the Safra Center), whereas in the explicit morality condition they were not. It appears, therefore, that in these two conditions participants' reports about how objective they were during the study may be better explained by self-justification following the morality intervention rather than by their actual behaviour. In contrast, the reports of non-objective behaviour by participants in the Control COI-no intervention condition cannot be explained by self-justification, as they are consistent with these participants' exaggerated favorable evaluations of the research conducted at the institute. It appears, therefore, that these participants not only behaved unethically but were also aware of it.

Summary of finding and discussion:

In the current study we examined how different intervention techniques affected people when they were in a subtle conflict of interest (opportunity to earn additional money if invited to participate in a future study) and with very limited ability to behave in a biased way (rating in somewhat favorite way the Safra center). Explicit measures (both deterrence and morality) had a much greater constraining effect on participants' judgment than did implicit measures (again, with a similar effect of deterrence and morality). The same pattern was obtained with two different dependent variables, an 18-item Likert-type scale questionnaire and a binominal questionnaire. Thus, raising the awareness of participants to the fact that there may be a problem was sufficient even in the context of subtle conflict of interests, and the type of the explicit intervention was less important. These 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 intervention used is less important (deterrence or morality), as long as it is conducted explicitly so that it triggers deliberate, System 2 processing. By contrast, at least in the context studied in the current study, implicit interventions did not have any effect on participants' unethical behaviour, suggesting that

interventions based on System 1 processing cannot override System 1 unethical behaviour. Yet, more research is needed regarding the automatic effect of morality on behaviour. In light of Gino and Desai's (2012) findings on exposure to childhood cues, it might be that a stronger implicit method would have affected participants' behaviour as such exposure resulted in reduced unethical behaviour. Nevertheless, taken together, the current findings strengthen our claim that traditional intervention techniques that target awareness should not be completely washed away by the more innovative nudge like techniques advocated in the behavioural ethics literature. Instead, the classic deterrence literature should be modified in light of the literature on behavioural ethics but it should not be abandoned all together even when dealing with "good" people. It might be the case that, as Banaji and Bazerman (2005) argue, incentives won't change the behaviour 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 will be done in specific organizational and legal contexts.

Several relatively robust findings regarding bounded ethicality in general emerged from the two experiments:

First, we have seen that under an intuitive/automatic mindset, participants who were offered an opportunity to participate in a second study were more likely to favor that institution, relative to participants who were in an analytical/deliberative mindset.

Second, in a the second study we have confirmed this interpretation of our finding from the first experiment by manipulating the self-interest of people (potential for participating in an additional experiments for a small fee) relative to a control group of participant who were not offered this offer, and found a substantial "corrupting" effect (reporting a more favorable view of the target stimuli than did the control group, which had no conflict of interest). This effect was manifested both in the 18-item and the

binominal questionnaires. Although the potential effect of money on behaviour is neither new nor surprising, the fact that an opportunity to earn such small amounts of extra money in a future research, subtly mentioned to the participants, created an increase in their evaluation of the research institute (which they were explicitly asked to conduct objectively) is rather alarming.

Second, we have seen that across most conditions, when participants have changed their views relative to the control group, it was only with regard to the dependent variables that measured their views on the importance of the research, not on the integrity of the researchers themselves. This can be explained by the fact that the latter judgment was based on far less information then the former and gave participants greater leeway (there was no reason for them to suggest an increase in the salary of the scientists or that there should be less scrutiny of their actions, except for some irrelevant considerations). The fact that it was more difficult to "corrupt" participants on this 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). Specifically, although the prospect of a smaller profit did change their judgment, it did so only in cases when it made sense for them to have an opinion. Participants 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, in that regard our findings do support the "good" people focus, as from a rationale choice perspective there was less room to have views that are more favorable to Safra in the "conflict of interest" condition, relative to the control group in only some measures but not in others.

Another important finding that emerges from the study has to do with the self-awareness of participants as to whether or not they were objective in their evaluation. As suggested in the introduction, this is an important area of research for the interaction between behavioural 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 appear to have led people in our manipulated conflict of interest to rate their objectivity lower than did members of the no conflict control group, regardless of their actual level of objectivity, as measured by their evaluations of the research institute. More important, however, participants in the manipulated conflict of interest group with no intervention condition, who indeed where less objective in their evaluations than their peers, admitted to being fully objective. This finding suggests that members of this group of participants were aware of their unethical behaviour, raising the question of how "good" are these people who do "bad things."

Limitations

The findings of this exploratory study described above, suggest several potential implications for theory regarding the interplay between behavioural ethics and law, as will be discussed in the following sections. But before proceeding with the substantive discussion, we must draw attention to some of the limitations of the study.

Naturally, considering the exploratory nature of the current research, its results should be treated with caution. Nevertheless, our main findings about 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 the introduction, was not to compare what intervention is stronger, as such findings are limited to the specific 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 behavioural ethics. In such contexts, the mere fact that we have found effects and shown some consistency in the effects of certain intervention should be seen as laying the ground for further explorations across different context of what types of intervention work better for each context.

Policy Implications

The design and findings presented above have several normative implications. First, the realization of how little is needed to change people's behaviour 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, the paradigm of "good people doing bad things," the study of which was the underlying motivation for this project, has been confirmed: it takes little for people to do bad things. At the same time, our findings also suggest something about the first part of the phrase: people were good in the sense that they did not go all out in an attempt to increase their chances of being selected for further employment. Our measure of objectivity had two levels: the first one was related to

evaluating the research, the second to how people evaluate the scientists. Participants were not given any information that would allow them to judge the integrity of the scientists. If people had been motivated strictly by a desire to maximize their self-interest, and small amounts of money could corrupt them, they should have been biased across all the measures of their judgment. The fact that the majority of participants did not go all out, even without any intervention, suggests that people used some self-restrain and set self-imposed limits on how far they would go in shifting their objectivity.

Second, another important component in paradigm of 'good people doing bad things' is the people's absence of awareness of their lack of objectivity in their evaluations. Naturally, the context of the present study makes it difficult for them to know whether their behaviour had crossed any lines because no such lines existed, but it seems that with sufficiently strong communication regarding a potential wrongdoing, many people would immediately change their behaviour, regardless of the nature of the intervention. This would not lead to a change in the behaviour of "bad" people, who would engage in further cost-benefit calculations to assess the wisdom of engaging in bad behaviour. In other words, much of the concern with the inability to deter the unaware individual (e.g. Chough, Bazerman & Banaji, 2005) might prove premature. People might indeed be not fully aware to the unethicality of their behavior, but possibly traditional explicit reminders of both deterrence and morality might be sufficient to cause them, to at least start correcting for it.

Third, our findings did not demonstrate much success both of the moral framing and of the implicit intervention measures, which leaves an open question of whether the celebration of ethical nudges as a trustworthy instrument to changing behaviour (as being argued for by behavioural ethics researchers who focused on the advantage of framing behaviour in moral terms) is indeed working.

Fourth, 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.25 BIT advises governments on how to use knowledge from psychology and behavioral economics in shaping people's behavior in socially desirable ways,. Generally speaking,26 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 doesn't 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. It seems that future research should examine how to combine traditional explicit interventions with implicit interventions, when attempting to shape ethical behavior.

References

Balabanis, G., Diamantopoulos, A., Mueller, R. D., & Melewar, T. C. (2001). The impact of nationalism, patriotism and internationalism on consumer ethnocentric tendencies. *Journal of International Business Studies, 32(1)*,157-175.

Banaji, M. R., & A. G. Greenwald. (2013). *Blindspot: Hidden Biases of Good People*. New York: Delacorte Press.

---

25   See Thaler, R H., Sunstein C.R. (2008), *N*udge: Improving Decisions Using the Architecture of Choice: Yale University Press. For a recent collection of papers on Nudge see Alemanno, Alberto, and Sibony A.L (2015). Nudge and the Law: A European Perspective: Bloomsbury Publishing.

26 Both UK and US governments rely increasingly on insights from behavioral sciences, especially with regard to non-deliberative choice by individuals. See http://www.behaviouralinsights.co.uk/wp-content/uploads /2015/07/ BIT_ Update-Report-Final-2013-2015.pdf. See also https://www.whitehouse.gov/the-press-office/2015/09/15/executive-order-using-behavioral-science-insights-better-serve-american. JRS SCIENTIFIC AND POLICYREPORTS: Applying Behavioural Science to EU Policy-Making:

http://ec.europa.eu/dgs/health_consumer/information_sources/docs/30092013_jrc_scientific_policy_report_en.pdf; *See* also the "Nudge and the Law " infra note no.7 The EU, a late-comer to this game, has recently established a task force that published a guideline report, describing how policy makers should use psychology when implementing behaviorally-based legal policy in areas related to health and consumer choices. In the last two years, many organizations such as the World Bank and the OECD have prepared reports identifying behavioral sciences as one of the main focuses of governments. For a discussion of the different enforcement trade-offs that legal policy makers should account for in choosing nudges of traditional enforcement mechanisms Feldman, Yuval, and Orly Lobel. "Behavioral Trade-Offs." *In Nudge and the Law: A European Perspective Almano and Siboni 2015*.

Baumeister , R. F.& Leary, M. R. (1995). The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation, *Psychological bulletin*, *117*(3), 497-529.

Bazerman, M. H., & A. E. Tenbrunsel. (2011). *Blind spots: Why We Fail to Do What's Right and What to Do about It*. Princeton, NJ: Princeton University Press.

Bazerman, M. H., G. Loewenstein, & D. A. Moore. (2002). Why good accountants do bad audits. *Harvard Business Review*, *80(11),* 96–103.

Bennedsen, M., &Feldmann, S. E. (2006). Informational lobbying and political contributions. *Journal of Public Economics*, *90(4)*, 631-656.

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.

Braithwaite, J., & Makkai, T. (1991). Testing an expected utility model of corporate deterrence. *Law and Society Review*, *25(1)*, 7-40.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love and outgroup hate?. *Journal of social issues*, *55(3*), 429-444.

Buchner, A. & Werner W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology*, *40(3),* 227-259.

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.

Campos, N. F., & Giovannoni, F. (2007). Lobbying, corruption and political influence. *Public Choice*, *131(1-2)*, 1-21.

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 Don 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: Cambridge University Press.

Cushman, F., Young, L. & Hauser, M. 2006. The role of conscious reasoning and intuition in moral judgment testing three principles of harm. *Psychological science, 17(12)*, 1082–89.

Dana, J., Cain, D. M., & Dawes, R. M. (2006). What you Don't Know Won't Hurt me: Costly (but quiet) Exit in a Dictator Games. *Organizational Behavior and Human Decision Processes*, *100(2),* 193–201.

Deci, E. L. & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological inquiry*, *11*(4), 227-268

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.

Druckman, D. (1994). Nationalism, patriotism, and group loyalty: A social psychological perspective. *Mershon International Studies Review*, *38(1)*, 43-68.

Epley, N., & Caruso, E. M. (2004). Egocentric ethics. *Social Justice Research, 17(2)*, 171–87.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7(10*), 454–59.

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 & Policy*, *35*, 11–52.

Feldman, Y., & Teichaman, D. (2009). Are all legal probabilities created equal? *NYU Law Review 84(4),* 980–1022.

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 & Ethics*, *41*(3), 620-628.

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*), 743–58.

Gino, F., & Galinsky, A. D. (2012). Vicarious dishonesty: When psychological closeness creates distance from one's moral compass. *Organizational Behavior and Human Decision Processes, 119*, 15–26.

Gino, F., M. Schweitzer, 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.

Gino, F., Ayal S., & Ariely D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological* Science, *20(3),* 393–98.

Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *The American economic review*, *96(2)*, 212-216.

Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences,* 106(30), 12506–11.

Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization, 3*, 367–388.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108(4),* 814–34.

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.

Heap, S. P. H., & Zizzo, D. J. (2009). The value of groups. *The American Economic Review*, *99(1)*, 295-323.

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: Gotham Books.

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

Jones, J. T., Pelham, B. W., Carvallo, M., & Mirenberg, M. C. (2004). How do I love thee? Let me count the Js: implicit egotism and interpersonal attraction. *Journal of personality and social psychology*, *87*, 665.

Jones, J. T., Pelham, B. W., Mirenberg, M. C., & Hetts, J. J. (2002). Name letter preferences are not merely mere exposure: Implicit egotism as self-regulation. *Journal of Experimental Social Psychology*, *38*, 170-177.

Kahan, D. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making,* 8, 407–424.

Kahneman, D. 2011. *Thinking, Fast and Slow*. London: Macmillan.

Kahneman, D., Knetsch, J. L., &Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business, 59(4)*, 285–300.

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.

Kay, A. C., Wheeler, S. C., Bargh, J. A., & Ross L. (2004). Material priming: The influence of mundane physical objects on situational construal and competitive behavioral choice. *Organizational Behavior and Human Decision Processes*, *95(1*), 83–96.

Kidder, R. M. (2009). *How Good People Make Tough Choices: Resolving the Dilemmas of Ethical Living*. New York: HarperCollins.

Kouchaki, M., Smith-Crowe K., Brief, A. P. & Sousa, C. (2013). Seeing green: Mere exposure to money triggers a business decision frame and unethical outcomes. *Organizational Behavior and Human Decision Processes*, *121(1)*, 53–61.

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.

Lessig L. (2011). *Rupblic, Lost: How Money Corrupts Congress--and a Plan to Stop It*. 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),* 633–44.

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.

Moore, D. A.,Tetlock, P. H., Tanlu, L. & Bazerman, M. H. (2006).Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. *Academy of Management Review*, *31(1)*, 10-29.

Morin, K., Rakatansky, H., Riddick Jr, F. A., Morse, L. J., O'Bannon III, J. M., Goldrich, M. S., Ray, P. et al. (2002). Managing conflicts of interest in the conduct of clinical trials. *JAMA, 287(1)*, 78-84.

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

Nuttin, J. (1987). Affective consequences of mere ownership: The name letter effect in twelve European languages. *European Journal of Social Psychology*, *17(4),* 381–402.

Paternoster, R., & Iovanni, L. (1986). The deterrent effect of perceived severity: A reexamination. *Social Forces*, *64(3),* 751-777.

Paternoster, R., & Lovanni L. (1986). The deterrent effect of perceived severity: A reexamination. *Social Forces, 64(3),* 751-777.

Paternoster, R., Saltzman, L. E., Waldo, G. P., & Chiricos, T. G. (1983). Perceived risk and social control: Do sanctions reallydeter. *Law & Society Rev*iew, *17(3)*, 457-480.

Piliavin, I., Gartner, R., Thornton, C., & Matsueda, R. L. (1986). Crime, deterrence, and rational choice. *American Sociological Review*, *51(1)*, 101-119.

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: Routledge.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489(7416)*, 427–30.

Robinson, P. H., & Darley, J. M. (2004). Does criminal law deter? A behavioural science investigation. *Oxford Journal of Legal Studies*, *24*(2), 173-205.

Rodwin, M. A. (2007). Medical commerce, physician entrepreneurialism, and conflicts of interest. *Cambridge Quarterly of Healthcare Ethics*, *16*(04), 387-397.

Rodwin, M. A. (2012). Conflicts of Interest, Institutional Corruption, and Pharma: An Agenda for Reform. *The Journal of Law, Medicine & Ethics*, *40(3*), 511-522.

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

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.

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.

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*, 1264–70.

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.

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)*, 645–65.

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.

Tenbrunsel, A. E., Diekmann , K. A., Wade-Bemzoni, K. A., & Bazerman M. H. (2010) The ethical mirage: A temporal explanation as to why we are not as ethical as we think we are*. Research in Organizational* Behavior, *30*, 153-173.

Thompson, D. F. (1993). Understanding financial conflicts of interest. *New England Journal of Medicine, 329(8),* 573-573.

Tyler, T. R. (2005). Managing conflicts of interest within organizations: Does activating social values change the impact of self-interest on behavior. In D. A. Moore, D. M. Cain, G. Loewenstein & M. H. Bazerman (Eds.), *Conflicts of interest: Challenges and solutions in business, law, medicine, and public policy* (pp.13-35). New York: Cambridge University Press.

Tyler, T. R. (2006), *Why People Obey the Law*, Princeton University Press, Princeton (NJ).

Tyler, T. R., & Fagan, J. (2008). Legitimacy and cooperation: Why do people help the police fight crime in their communities?. *Ohio St. J. Crim. L.*, *6*, 231.

Valdesolo, P., & DeSten, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17(6)*, 476-477.

Vohs, K. D., Mead N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science 314(5802),* 1154–56.

Vohs, K. D., N. L. Mead, & M. R. Goode. (2008). Merely activating the concept of money changes personal and interpersonal behavior. *Current Directions in Psychological Science*, *17(3)*, 208–12.

Wang, L., Murnighan, J. K. (2011). On greed. *The Academy of Management Annals*, *5(1)*, 279-316.

Zimbardo, Philip. (2007). *The Lucifer Effect: Understanding How Good People Turn Evil*. New York.
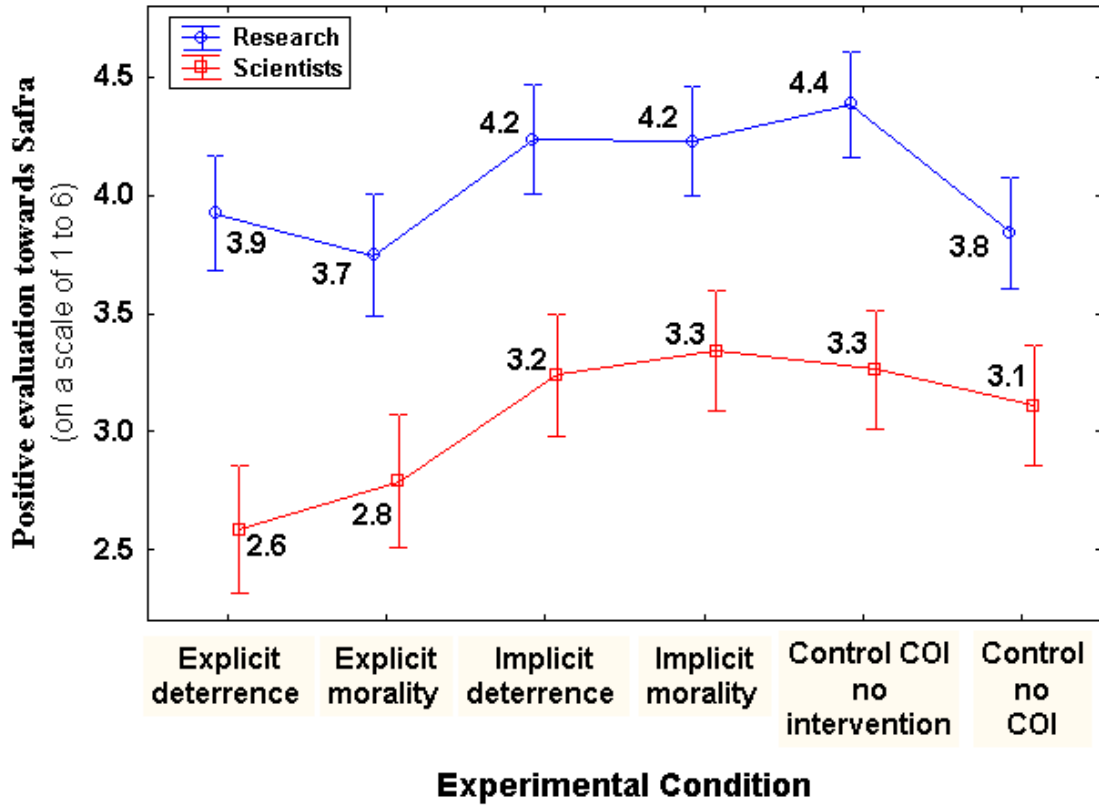

Figure 1

Table 1

	Average survey completion time
1. Explicit deterrence	20.2 min
2. Explicit morality	18.7 min
3. Implicit deterrence	21.0 min
4. Implicit morality	23.1 min
5. Control COI-no intervention	22.7 min
6. Control–no COI	20.5 min

Table 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-no intervention	73.6%
6. Control–no COI	62.4%

Table 3

	Proportion of nonobjective answers
1. Explicit deterrence	9.6%
2. Explicit morality	21.1%
3. Implicit deterrence	10.9%
4. Implicit morality	17.3%
5. Control COI-no intervention	19.9%
6. Control–no COI	9.5%