**Behaviorally Informed Regulation of Dishonest Behavior:**

**Understanding How, When, and Why Pledges Can Reduce Dishonesty while Promoting Trust and Regulatory Efficiency**

Eyal Pe’er (School of Public Policy, Hebrew University of Jerusalem)

Yuval Feldman (Faculty of Law, Bar-Ilan University)

**Abstract**

A common dilemma in regulation is determining how much trust authorities can place in people’s self-reports. Regulators, who are typically risk averse, do not readily confer trust, resulting worldwide in excessive requirements when applying for permits, licenses, and the like. However, recent research in behavioral ethics suggests that asking people to pledge ex ante to behave ethically reduces their level of dishonesty and incompliance. If such pledges indeed prevent people from cheating on their self-reports, regulators could then relax many bureaucratic hurdles—with their huge costs to market efficiency, voluntary compliance, and trust—without incurring major risks to the public interest. Although some evidence (including our own preliminary results) shows that ex-ante pledges can curb unethical behavior, no study to date has systematically examined the conditions and factors that determine when and how pledges reduce dishonesty and when they may be counterproductive. It is unclear whether and how pledges’ fare better than command-and-control regulation (e.g., fines or sanctions on cheating) and whether their effect may decay over time, which would put their efficacy in question. Moreover, it is theoretically unclear why pledges work and what psychological mechanisms underlie their effects. There is also no known theoretical framework or principles guiding pledges’ optimal design, specifically in relation to their language, scope, and content. Lastly, the prediction that pledges, being trust-based means, could increase trust between regulated parties and their regulators has never been put to an empirical test. Understanding when and why pledges curb unethical behavior and enhance voluntary compliance could both advance our scientific knowledge on people’s (un)ethical behavior and provide valuable behavioral insights for regulatory policy and governance. Our proposed research aims to provide both theoretical and applicable knowledge on when, how, and why to use pledges to prevent dishonesty, improve regulatory and enforcement decisions, and help rebuild trust in governance and public policy.

# **A. Scientific Background**

Regulators’ need and efforts to control human behavior are related to the degree to which governments trust their citizens (Moyson, 2017). A common dilemma faces many policy makers: Can the public be trusted to provide accurate and honest reports of their actions, intentions, and behavior, or should the government invest resources in measures that prevent people from behaving dishonestly, often at high procedural costs and posing an increased regulatory burden (Anania & Nisticò, 2004; Gilligan, 2018)? Because regulators tend to be risk averse, states often prefer not to confer trust upon those regulated and instead to do whatever they can to prevent risks to the public interest and ensure public safety (e.g., Bews & Rossouw, 2002; Cohn, Fehr & Maréchal, 2014). For instance, the World Bank’s 2019 *Doing Business* report identifies many hurdles and obstacles to getting permits to open a business, registering a property, or obtaining financial credit (World Bank Group, 2019). These situations are inherently risky for regulators and so encompass a whole array of burdensome requirements to reduce the likelihood of mistakes. Mistrust also makes governments suspicious of the veracity of individuals’ self-reports: states cannot infer ex ante the proportion of the population who will exploit the option to self-report to make fraudulent claims (e.g., Feldman, 2018). The end result is often a highly burdensome bureaucratic mechanism that reduces risk to regulators, but hampers growth.

One solution to this problem can be found within the responsive regulation paradigm (Ayres and Braithwaite, 1992). This paradigm adopts a dynamic regulatory strategy, which first chooses trust between regulators and regulatees, followed by an escalation to more punitive regulation if that trust is abused (Braithwaite & Makkai, 1994). To tackle dishonesty under this paradigm, policy makers move beyond the one-size-fits-all command-and-control policies that typically require costly monitoring and enforcement and identify cases in which it could be possible to trust people and use less forceful and less coercive measures of ensuring honest and ethical conduct. An important concept is the “enforcement pyramid,” which outlines a careful, stepped-up escalation in means used by enforcers: this scheme informs regulatees that regulators could move to impose harsher means at every compliance dilemma (Ayres & Braithwaite 1992). Focusing first on regulatory means that allow people to feel trustworthy and that enable them to engage in voluntary compliance facilitates the formation and growth of trust (Feldman 2018; Möllering 2006). For example, when an entrepreneur wishes to open a new business or when citizens report their annual expenses and income for tax purposes, the government does not always have to insist that they provide all relevant materials beforehand and then scrutinize their documents before approving their application. In some cases, governments may simply ask applicants to guarantee, in advance, that their reports or applications are accurate and honest; then the state can invest more resources in auditing and sanctioning afterward. Such ex-ante pledges can clearly reduce the administrative burden imposed not only on citizens but also on regulation and licensing authorities (e.g., Kucher & Götte, 1998; Torgler, 2003).

Yet relying on pledges, rather than imposing mandatory checks, raises the risk that some people may take advantage of the situation and make false reports to claim higher benefits for themselves (Feld & Frey, 2018). Recent research on unethical behavior has shown that indeed many people would cheat if given the opportunity (e.g., Gerlach, Teodorescu, & Hertwig, 2019; Jacobsen, Fosgaard, & Pascual‐Ezama, 2018) and that this cheating results in excessive damage to the social fabric of society (Gächter, & Schulz 2016). However, some research has also suggested that pledges can, under some conditions, make the ethical requirement more salient, which reduces the likelihood that people will behave dishonestly (Bazerman & Gino 2012). Namely, when participants in a study were asked to make a pledge by adding their signature to an honesty statement in advance, they claimed less unwarranted rewards in a laboratory task. Applying this “signing-at-the-beginning” design to car insurance application forms in one company also resulted in claimants reporting, in their renewal application, a higher annual mileage (which results in higher premiums and thus is considered more honest; Shu, Mazar, Gino, Ariely, & Bazerman, 2012).

Although it has yet to be researched systematically, the ability of pledges to reduce dishonesty has been shown by several additional studies. Beck, Bührn, Frank, and Khachatryan (2018) used the “die under a cup” paradigm, in which participants roll a die secretly to determine their payment for participating in the experiment (e.g., Shalvi, Eldar, & Bereby-Meyer, 2012); they found that dishonest reporting decreased considerably when participants had to promise, with their signature, that the data they would provide regarding their performance during the experiment would be in line with the principle of honesty and that they would not lie to enrich themselves. Similarly, Jacquemet, James, Luchini, Murphy, and Shogren (2019) had participants (freely) sign a pledge that states, “[I] swear upon my honor that, during the whole experiment, I will tell the truth and always provide honest answers,” and then play a sender–receiver game (Erat & Gneezy 2012) with different payoff schemes. They found that the oath reduced lying, and the effect was significant when lies were made explicit in the instructions. In contrast, one recent study showed that students who were asked to sign a commitment form (pledge) before starting the exam actually showed an *increased* rate of cheating, measured by their propensity to give incorrect answers that were identical to those of their neighbors (Cagala, Glogowsky, & Rincke, 2019).

Other studies have examined the indirect effects of pledges on reducing biases in preference elicitation and survey responses. Carllson et al. (2013) asked survey respondents in Sweden and China to “promise to answer the questions in the survey as truthfully as possible,” which measured how much they would be willing to pay in increased taxes to reduce carbon emissions. They found that the oath led participants to provide willingness-to-pay (WTP) values that are considered more realistic (with less zeros or maximum values, and a lower variance). Similarly, Kemper, Nayga, Popp, and Bazzani (2016) found that an honesty oath led to significantly lower WTP values, which are regarded as less biased. Thus, it appears that, when asked to pledge their honesty ex-ante, people take their promises seriously and curb their typical inclination to provide biased answers.

Thus, pledges may be a very useful tool for regulators in their attempts to balance regulatory efficiency and the “ease of doing business” with the need to protect the public interest from unethical behaviors. Indeed, some countries have already installed responsive regulatory approaches in some of their policies, using versions of ex-ante pledges or affidavits (Ivec & Braithwaite, 2015). However, because they were part of larger reforms that included many other changes, it is hard to discern the actual impact of using pledges. Before we can argue for or against the use of pledges in regulation, it is critical to ascertain how, when, why, and to what extent they would prevent dishonest behavior, as well as to understand under which conditions pledges would be counterproductive and should thus be avoided. Clearly, the studies conducted to date do not yet allow us to answer these important questions.

**B. Research Objectives and Expected Significance**

Although the studies described earlier suggest that pledges can reduce dishonesty in some cases, the actual effectiveness, scope, scalability, and possible boundaries of pledges are still unknown. This has limited the use of pledges for several reasons. First, the effects of pledges have only been tested on one-shot or short-term decisions. However, in everyday situations, the temptation or opportunity to cheat may occur frequently and even several times a day; it may also occur a long time after the individual has taken the pledge. Second, earlier studies focused only on showing that pledges can reduce dishonesty compared to when no pledge is taken and when cheating does not lead to any major negative consequences. However, in the real world, cheating and lying are often followed by traditional mechanisms of enforcement and deterrence, such as financial sanctions (fines), legal sanctions, and other penalties, despite the individual having signed a pledge or affidavit. Cheating in those situations would entail serious repercussions if caught, and the threat of financial or other sanctions can itself reduce cheating (e.g., Laske, Saccardo, & Gneezy, 2018). Adding a pledge might then not offer a benefit when such fines are present. Thus, the effect of a pledge must be tested and evaluated against such traditional sanctions to determine its ecological validity and implementability. Moreover, research to date has yet to provide conclusive evidence that explains why pledges actually work (when they do) and what are the underlying mechanisms that could be driving their ability to reduce dishonesty. Clearly, a more systematic research program is required to fully understand the scope, boundaries, and moderating factors for the potential ability of pledges to reduce dishonesty.

Studying the mechanisms through which pledges operate would provide insights not only about how to design them in ways that would maximize their efficacy; it also would help determine in which conditions pledges would be most effective, so that they could achieve their ultimate aim: to contribute to building trust between those people who are interested in behaving ethically and their government. We could then create a situation in which the regulatory burden on “good” people could be lessened without a heavy reliance on enforcement mechanisms that reduce intrinsic compliance motivation (Bowles & Polania-Reyes, 2012)—while not jeopardizing public safety and enabling the building of mutual trust (e.g., Hardin, 2002).

Our research has five main components:

1. Evaluating the *effect of pledges* *in reducing dishonesty* in frequently encountered situations, in comparison to traditional measures of fines and sanctions
2. Uncovering the *temporal pattern of the effect* of pledges over time
3. Understanding the *psychological mechanism(s) underlying the (discovered) effects* of pledges on dishonesty
4. Systematically exploring and identifying *how to optimally design pledges* (e.g., in respect to language and content) and, lastly, generalizing to a more institutional level
5. Testing *how pledges can help rebuild or foster trust* between the regulated parties

In the following subsections we elaborate on each research component.

**C. Detailed Description of the Proposed Research**

1. **Evaluating the effectiveness of pledges**

The first component of the research will empirically test and determine the effectiveness of pledges in reducing dishonesty in repeatedly made decisions—instead of one-shot choices, the focus of earlier studies—compared to traditional measures of preventing dishonesty such as fines and sanctions. Even those who believe that many people can be trusted to abide by their pledges recognize the need to maintain some monitoring mechanisms, with penalties for transgressions. Thus, it is crucial to understand the interaction between pledges and sanctions. The current research on sanctions provides evidence for two competing types of interaction effects: crowding out and crowding in (Bowles et al., 2012). The most common finding is that sanctions, especially if they are not strong enough, can undermine compliance by crowding out prosocial motivations (Tenbrunsel & Messick, 1999). In contrast, other studies suggest that, in some contexts, incentives such as monetary rewards increase voluntary compliance and prosocial behavior (e.g., Galbiati & Vertova 2014). To date, studies on crowding out and crowding in have focused on altruism—participants in public good experiments helping others and giving charitable donations—and not on honesty. Studying the joint effect of incentives and pledges on honesty is hence an important step toward understanding how pledges operate in the real world.

To examine the effect of pledges compared to traditional fines and sanctions, we will build upon the basic design that is detailed in the later section, Preliminary Results. We will have participants perform a task that provides an opportunity to cheat and earn more money by overstating their performance (Mazar, Amir, & Ariely, 2008). The task is to solve 10 simple calculus problems sequentially, for which the participants earn a small bonus payment per problem; these bonuses, however, are earned differently in the control and experimental conditions. In the control condition participants have to report how they solved each problem, and their reports are checked thoroughly. In the experimental (cheating) conditions, participants only need to mark the box that says they solved a problem to earn a bonus for it, without having to actually provide the solution. Although this design does not enable us to measure how much (if at all) each participant cheated individually, it does allow the comparison of report rates between conditions to arrive at the more important aggregate cheating rates. Such tasks have been used extensively in studies of behavioral ethics in the past two decades (Gerlach et al., 2019). Recently, using a conceptually similar task paradigm of the “die under a cup,” researchers were also able to prove that such behaviorally abstract tasks, with little relevance to everyday life, are strongly correlated with actual dishonest behaviors by public sector employees (Olsen, Hjorth, Harmon, & Barfort, 2019) and that they predict degrees of rule violations between countries (Gächer & Schultz, 2016).

We will build on our basic design to include conditions that measure how adding pledges can reduce the over-reporting gap found between the cheating and control conditions. Specifically, in the pledge condition participants will be asked to pledge their honesty before performing the task (as in Shu et al., 2012); we will then measure how the addition of the pledge curbed the dishonesty gap between the cheating and the control groups. Then, we will vary both the probability of participants’ responses being audited for correctness (as in Laske et al., 2018) and the threat and level of the sanction (e.g., a fine) that would be imposed on those caught cheating in the audit. For example, participants in one of these conditions would be told that a random 10% of their answers will be audited and that if they fail to provide the actual solution, they will suffer a fine of, say, half of their earnings or a fixed pre-determined sum.

Our preliminary results show that a pledge can indeed reduce dishonest over-reporting in a considerable and consistent manner—even when the chance of an audit is small (10%) and the fine is minimal (participants would only lose the amount earned for the audited problem). In fact, increasing the fine to its maximum level (losing all their earnings) did not, in our preliminary study, increase the effect of the pledge. However, it is necessary to explore other levels of probabilities and fines, as well as other sanctions, such as bans from future participation. Thus, our first set of studies will follow the basic design of the preliminary study, but will vary (a) the gains to be earned from cheating, (b) the probability of being audited, and (c) the size of the fine. This should give us a clearer understanding of the boundaries within which pledges can effectively reduce dishonest behavior and when, if ever, they may not be effective or even counterproductive.

1. **Uncovering temporal decay**

The second component of our research will examine whether and how the effectiveness of pledges decays over time. In many situations, people are routinely asked to give pledges before they do a task or fulfill their responsibilities. For example, students are often asked before they begin a test to pledge that they will not cheat on it, managers and public servants pledge their ethicality before taking office, and witnesses in trials take an oath of truthfulness before providing their testimony. Sometimes such a pledge is solicited in close temporal proximity to the action, and sometimes a large span of time passes between the time the pledge is taken and the person is confronted with the temptation to behave dishonestly. To date, studies on the effectiveness of pledges have only examined those taken right before the action (e.g., Shu et al., 2012). However, in many situations there is often a significant time interval between the pledge and the opportunity to cheat. Understanding the time frame of the effect of pledges on ethical behavior is crucial to understanding the nature of their effect both theoretically and from a prescriptive point of view of when pledges should be taken and then be renewed or reinforced.

Research on ethical decision making provides important insights into the potential fading over time of the power of ethical nudges. On the one hand, being aware of the ethical meaning of one’s behavior can increase one’s honesty (Ayal, Gino, Barkan & Ariely, 2015; Ruedy & Schweitzer 2010). Thus, if one assumes that such awareness decreases over time, then so should the effect of the pledge on curbing dishonesty. In contrast, research on preference change (e.g., Vanberg, 2008), as well as on personal commitment and internalization of morality (O'Reilly & Chatman, 1986), supports the claim that the process of making a commitment to ethicality will have a sustainable effect on people’s ethical decision making, unrelated to their level of activated awareness of morality. This suggests that pledges’ effects will not decrease over time, at least not substantially, because the person taking the pledge would internalize and commit to the honest course of action and in lieu of any new experience or information, would prefer to stick to this behavior. Clearly, both accounts seem plausible theoretically, and thus it is important to discern the actual temporal effect of pledges over time and whether it depends on any observable moderating factors.

Specifically, in this part of the research, we will examine whether the effect of pledges decays over short versus long periods of time and to what degree. Moreover, we will analyze how the time interval between the pledge and the action affects the efficacy of the pledge. In this context, we will also explore whether reminders, which have often been found to be effective nudges of desired behaviors (e.g., Nickerson & Rogers, 2010) can help mitigate any decay found in pledges’ effects over time. To accomplish this, we will repeat the basic design (as detailed in the Preliminary Results section) of our study while we (a) increase the number of trials (problems) in which participants can cheat and (b) add varying time intervals between the time when the pledge is taken and the opportunity to cheat occurs. The post-pledge time interval will be filled with either an unrelated task (e.g., a reading comprehension task) or a task that would be aimed at increasing or reducing the saliency of moral values to the participant. In this, we aim to simulate real-life scenarios that might have a positive or negative effect on the potency of the pledge so we can estimate these moderators’ impact on the effects of pledges.

1. **Understanding the psychological mechanism(s) underlying the effects of pledges**

Next, we aim to identify the psychological and behavioral mechanism(s) that can explain the effects of pledges in an attempt to better understand why and when pledges should work in real-life situations (cf. Jacobsen, Fosgaard, & Pascual‐Ezama 2018). Current research suggests that pledges work by reminding people about their morality and values. Shu et al. (2012) found that signing a pledge at the beginning of a self-report leads to less dishonest reporting or over-claiming: in another study they found that signing at the beginning increased the frequency of ethicality-related words in a word-completion task, suggesting that participants who were asked to pledge (sign) in advance had more moral concepts activated in their minds. However, the notion that pledging reduces unethicality by invoking moral values lacks further empirical support, and alternative mechanisms have not yet been ruled out. Pruckner and Sausgruber (2013), for instance, conducted a field experiment on self-payment newspaper booths in Germany, to which they added two types of moral reminders. They found that a reminder that included reference to moral norms increased self-payments, whereas one that referred to legal norms did not. Mazar et al. (2008) reported that asking people to recall the Ten Commandments reduced cheating in the anonymous matrices task. However, a recent large replication project of this experiment, done by 19 labs worldwide, showed disconfirming results: in none of the cases, did this type of moral reminder reduce cheating significantly, and it even increased cheating in one (Verschuere et al., 2018).

The literature on behavioral ethics identifies several behavioral and psychological mechanisms that might explain pledges’ effects. First, despite the failure to replicate the effectiveness of the Ten Commandments nudge, pledges could indeed simply act as morality reminders, highlighting the importance and desirability of behaving in an ethical manner (Shu, Gino, & Bazerman 2011). A different possibility is that pledges act as signals of social norms, encouraging people not to deviate from others’ behavior and thereby maintain their ethicality. Indeed, much research has showed that highlighting social norms can help curb undesired behavior, such as littering (Reno, Cialdini, & Kallgren,1993) or excessive energy consumption (e.g., Allcott (2011). Thus, it is possible that, when asking people to pledge, the perception of an ethical social norm gains more saliency than it had before. A third account proposes that pledges are pre-commitment devices (Baca-Motes, Brown, Gneezy, Keenan, & Nelson, 2012) and exert their influence because they appeal to people’s inherent desire to act self-consistently (Swann, Rentfrow, & Guinn, 2003) . A fourth reason why pledges work involves how they affect the choice architecture of the decision maker. Namely, pledges may act as cues that the default behavior expected in the situation is an honest one, and because people typically prefer to stick to defaults (e.g., Korobkin, 1997; Sunstein 2013), pledges promote more honest behavior. Lastly, pledges may make some aspects of the situation, such as the perceived risk of getting caught or the perceived penalty for dishonest behavior or both, more salient and clearer than when people are not required to actively sign or affirm their honest behavior ex ante.

Theoretically, any one or more of these accounts may explain pledges’ effects. Thus, in this component of research, we will conduct a series of experimental studies to isolate and test each account. For instance, to test whether pledges affect perceptions of social norms, we will ask participants in the study to indicate what is their personal prediction about how many people would cheat under this situation. We predict that the more people believe that others would cheat, the less of an effect the pledge would have. We will then manipulate the descriptive and injunctive social norm by explicitly giving participants different information about how many participants have cheated in the past (e.g., by highlighting either the percent of people who cheated to invoke a dishonest social norm or the percent of people who did not cheat to emphasize a social norm of honesty). We will design additional experimental manipulations to test the role of other factors that could explain the effects of pledges, including their effects on the saliency of moral values, self-consistency, or perceived risks and penalties.

1. **Identifying how to optimally design pledges**

After we achieve an understanding of when and why pledges reduce dishonesty, we will explore the optimal language and content of pledges. The major issue is how clear, specific, and particular the statement should be. When creating pledges, choice architects (e.g., policy makers) must decide on the level of specificity in describing the desired (honest) behavior. Research on standards versus rules (Kaplow, 1992) suggests there is a trade-off between detailed and general commitments, in which broadly defined commitments are more effective in uncertain, ambiguous circumstances (Feldman & Smith, 2014). In contrast, detailed pledges might reduce self-deception about the meaning of the promise (cf. Dana, Weber, & Kuang, 2007) while focusing people’s attention on particular aspects of their tasks (Boussalis, Feldman, & Smith, 2018). We thus propose that specific pledges would be more effective in preventing unethicality in actions similar to the one described in the pledge. Furthermore, specific pledges should work better with sanctions, because they will make clearer how the violation and deviation from the pledge will be measured (e.g., Cramton, 1969). We also predict that people would be more committed to behaving honestly when all facts are known when they make the pledge and when there are fewer contingencies and less reliance on vague statements, such as “I will do my best to conform to the rules”.

Another important issue in the design of pledges is whether it is better to use legal or formal language or more accessible wording that laypeople could more easily relate to and comprehend. The importance of language in ethical decision making specifically, and in decision making in general, has been widely acknowledged (e.g., Stevens, 1994). Some studies have shown the advantages of the use of formal over informal language in the relationship between employees and employers (Kouchaki, Gino & Feldman, in press). We thus propose that formal language pledges would be more likely to increase the likelihood that people will remain committed to their pledges. To confirm this hypothesis, we will devise several different, yet equivalent, phrasings for the same pledge and randomly assign participants to conditions with the differently phrased pledges to examine which type of language (i.e., formal vs. informal, specific vs. general) is most effective in curbing dishonesty.

1. **Testing how pledges can help rebuild or foster trust**

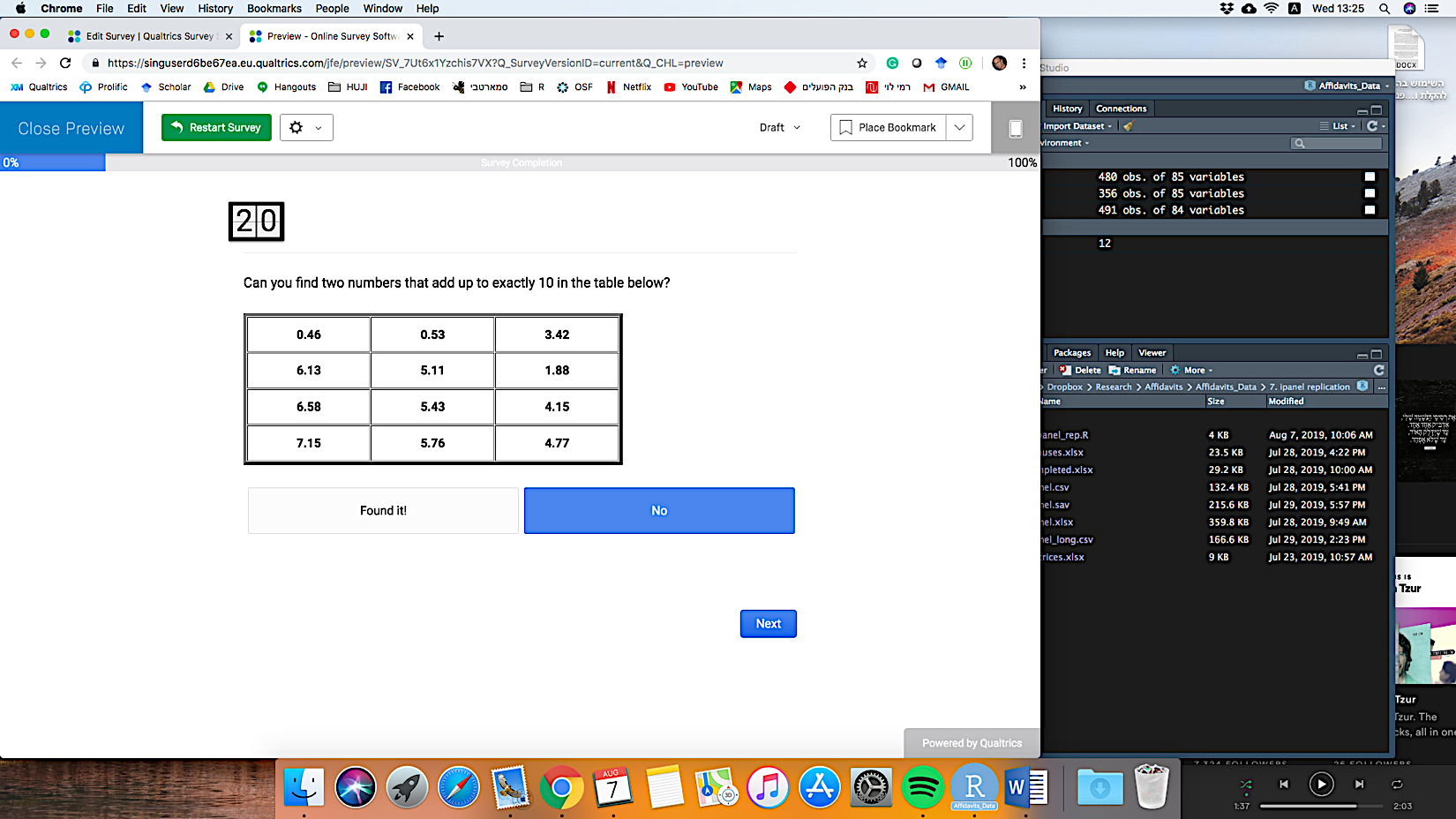
In our final component, we will examine the impact of making the transition to a regulation model that frequently uses pledges. Specifically, we will explore how pledges would affect trust between people and regulators over time. Could pledges, being a trust-giving mechanism, invoke more trust from people, and would that increased trust lead to more uses of trust-giving regulation, creating a repeating virtuous cycle of reinforcing feedback that would promote trust between both parties in a sustainable manner? Or, alternatively, would pledges actually increase the salience of the opportunities to cheat for increased gains, which could lead more people, in the long run, to try and exploit pledges more, leading regulators to distrust people more and avoid using pledges? To examine these questions, we will add, at the end of our experimental design, existing measures of trust and ask participants to rate the experimenters on those measures. We will then compare trust ratings between those who were asked to pledge their honesty and those who were not, controlling for actual performance (and payment) in the experimental task.

Additionally, to explore the building of trust on an institutional level, we will work together with several government ministries in Israel that plan to introduce pledges in some of their regulatory reforms. We are currently working with agencies and departments within the Israeli Ministries of Justice, Interior, and Finance. With their cooperation, we plan to compare the level of people’s trust in agencies before and after the introduction of pledges in the issuance process of permits, licenses, and the like. We predict that as long as the use of pledges indeed reduces the bureaucratic burden, they will produce increased levels of trust for the relevant agency.

**Preliminary Results**

To test the effect of pledges on dishonest behavior, we conducted an online experiment that included 491 participants (42%, Mage = 39, SD=11.8) sampled from an online participant pool. Participants were asked to complete 10 simple calculus problems, adopted from Mazar et al.’s (2008) “Matrix Task.” For each problem, participants are presented with a table containing 12 numbers from 0 to 10 with two decimal digits each (see Figure 1) and are given 20 seconds to find two of those numbers that, when added together, equal 10. For each problem they solve they are awarded a monetary bonus of 0.5 NIS (about 0.15 USD). Thus, they can earn up to 5 NIS (about 1.5 USD) on the entire task, in addition to their participation fee. In the control group, where cheating is not possible, participants are asked to provide the two numbers that add up to 10 for each problem they report as “solved.” When they complete their calculations, control participants’ responses are checked for accuracy, and they are paid only for the problems they solved correctly (there are no penalties for incorrect answers). In the experimental conditions, cheating is made possible by instructing participants that they only need to report, for each problem, whether or not they solved it; they do not have to provide the solution. They are also told that there is a 10% chance that each problem might be “audited,” in which case they would be asked to provide the solution for that problem only. In these groups, earnings were given based on self-report only, and the difference between the reported number of problems solved in the experimental groups compared to that solved in the control group was used as the measure of (group-level) cheating. To test the effect of pledges, with or without fines, we had three experimental conditions: (1) self-report (without a pledge or fine), (2) pledge only (without a fine), and (3) pledge + fine (if caught cheating in the audited problem, participants would lose all their earned bonuses).

Figure 1. Screenshot of a problem in the cheating task in the Preliminary Study.



All participants completed the 10 problems in a randomized order. Participants in Conditions 2 and 3 that included a pledge were asked to retype the following sentence to indicate their agreement with it: “I hereby promise that I will report a problem as solved only after I verified that I indeed found two numbers that add up to exactly 10. I am aware that the bonus I will receive will be based on my self-report and so I will make sure to report accurately and honestly.”

Figure 2 shows the mean percent solved (or reported as solved) in each condition, with 95% confidence intervals around the means. Participants in the control group correctly solved 32.5% (SD = 24.5) of the problems. In contrast, participants in the self-report group (Condition 1), for which there was no pledge or fine, reported solving 65.3% (SD = 25.8) of their problems. The difference between the two reporting rates, an indicator of the average cheating rate, was about 33 percent points (or 200%). However, when participants were asked to pledge beforehand (Condition 2), they reported solving only 46.7% (SD = 23.3) of the problems, reducing the cheating rate to 14 percent points (or about 130%) only. Participants who were asked to pledge and were warned about the possibility of a maximal fine (Condition 3) had a similar report rate at 50.3% (SD = 23.5). The overall differences between the conditions were statistically significant: *F* (3, 352[[1]](#footnote-2)) = 28.49, *p* < .01. However, the difference between the Conditions 2 and 3 (pledge without or with a fine) was not statistically significant: *p* = 0.79.

Figure 2. Mean percent of problems reported as solved in each condition (error bars are 95% CIs).

****

Control Self-report + Pledge + Pledge & Fine

Failure rates in the audited problem were similar (recall that each problem had a 10% chance of being audited): 58.9% of those audited in the first experimental group (no pledge or fine) failed; that is, provided an incorrect or no solution. In contrast, only 40.5% failed when asked to pledge beforehand, and 46.9% failed when the pledge was accompanied by the threat of a fine, χ2 (2) = 3.76, *p* = 0.15. We found no significant differences in the effect of pledges on participants of different religiosity levels, education levels, income, age or gender.

**D. Available Conditions for Conducting the Research**

The first PI (Pe’er) is a Senior Lecturer at the Federmann School of Public Policy and Government at the Hebrew University of Jerusalem. With a PhD in psychology, Pe’er has extensive knowledge and experience in quantitative and behavioral research in judgment and decision making in general, and in unethical behavior specifically. He has access to a fully equipped behavioral laboratory that has the resources to run experiments, as required for the proposed research program. The second PI (Feldman) is a Full Professor at the Faculty of Law at Bar-Ilan University and also holds a BA in Psychology from Bar-Ilan University. His areas of research include behavioral analysis of law, experimental law and economics, ethical decision making, regulatory impact and social norms, compliance, formal and informal enforcement strategies. Feldman has extensive experience in this domain and is a leading expert in behavioral ethics. His latest book *The Law of Good People*, on the interaction between behavioral ethics and law, was published by Cambridge University Press in 2018.

Both PIs are in close contact with top officials at relevant government offices, who took part in a recent workshop on the topic, where they expressed their interest and desire to collaborate in behavioral research on the use of pledges as regulatory tools. Currently, we are discussing potential projects with officials at the Ministry of Justice (to implement pledges in customs regulations), the Tax Authority (to explore the use of pledges in some reporting procedures), and at other ministries. As discussed earlier, the current state of knowledge on the efficacy of pledges is still far from satisfactory for policy makers, and we believe that the research outlined in this proposal could strongly enrich those efforts, allowing us to validate some of the findings in the field and test additional predictions that require real-life settings outside the experimental lab. We are thus committed (and pledge) to engage in this research topic with regulators in the field to ensure that our findings yield both theoretical and applicable conclusions.

**References**

Allcott, H. (2011). “Social Norms and Energy Conservation. *Journal of Public Economics* , 95, 1082–1095.

Anania, G., & Nisticò, R. (2004). Public regulation as a substitute for trust in quality food markets: what if the trust substitute cannot be fully trusted?. *Journal of Institutional and Theoretical Economics JITE*, *160*(4), 681-701

Ayal, S., Gino, F., Barkan, R., & Ariely, D. (2015). Three principles to REVISE people’s unethical behavior. *Perspectives on Psychological Science*, *10*(6), 738-741.

Ayres, I., & Braithwaite, J. (1992). *Responsive Regulation: Transcending the Deregulation Debate*. Oxford University Press.

Baca-Motes K., Brown A., Gneezy A., Keenan E. A., Nelson L.D. (2013). Commitment and behavior change: Evidence from the field. *Journal of Consumer Research* 39, no. 5 (2012): 1070-1084.

Bazerman, M. H., & Gino, F. (2012). Behavioral ethics: Toward a deeper understanding of moral *judgment* and dishonesty. *Annual Review of Law and Social Science*, 8, 85–104.

Beck, T., Bührn, C., Frank, B., and Khachatryan, E. (2018). Can honesty oaths, peer interaction, or monitoring mitigate lying? *Journal of Business Ethics*, 1-18.

Bews, N. F., & Rossouw, G. J. (2002). A role for business ethics in facilitating trustworthiness. *Journal of Business Ethics*, 39: 377-390. doi:10.1023/A:1019700704414

Boussalis, C., Feldman, Y., & Smith, H. E. (2018). Experimental analysis of the effect of standards on compliance and performance. *Regulation & Governance*, *12*(2), 277-298.

Bowles, S., & Polania-Reyes, S. (2012). Economic incentives and social preferences: substitutes or complements?. *Journal of Economic Literature*, *50*(2), 368-425.

Braithwaite, J., & Makkai, T. (1994). Trust and compliance. *Policing and Society: An International Journal*, 4(1), 1-12.

Cagala, T., Glogowsky, U., & Rincke, J. (2019). Content Matters: The Effects of Commitment Requests on Truth-Telling. *Available at SSRN 3432445*.

Carlsson, F., Kataria, M., Krupnick, A., Lampi, E., Löfgren, Å., Qin, P., & Sterner, T. (2013). The truth, the whole truth, and nothing but the truth—A multiple country test of an oath script. *Journal of Economic Behavior & Organization*, *89*, 105-121.

Cohn, A., Fehr, E., & Maréchal, M., A., (2014). Business culture and dishonesty in the

banking industry, *Nature* 516, 86–89.

Cramton, R. C. (1969). Driver behavior and legal sanctions: a study of deterrence. *Michigan Law Review*, *67*(3), 421-454.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67-80.

Erat S, Gneezy U. (2012) White lies. *Management Science*. 58 (4): 723–33.

Feld, L. P., & Frey, B. S. (2018). Illegal, immoral, fattening or what?: How deterrence and responsive regulation shape tax morale. In *Size, causes and consequences of the underground economy* (pp. 15-37). Routledge.

Feldman, Y. (2018) The Law of Good People: Challenging States’ Ability to Regulate Human Behavior. *Cambridge (UK): Cambridge University Press.*

Feldman, Y., & Smith, H. E. (2014). Behavioral equity. *Journal of Institutional and Theoretical Economics JITE*, *170*(1), 137-159.

Gächter, S. and Schulz, J.F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature* 531: 496–499. http://doi.org/10.1038/nature17160

Galbiati, R., & Vertova, P. (2014). How laws affect behavior: Obligations, incentives and cooperative behavior. *International review of law and economics*, *38*, 48-57.

Gerlach, P., Teodorescu, K., & Hertwig, R., (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological bulletin* 145, no. 1: 1.

Gilligan, G. (2018). The Hayne royal commission and trust issues in the regulation of the Australian financial sector. *Law and Financial Markets Review,* 12:4, 175-185, DOI: [10.1080/17521440.2018.1548209](https://doi.org/10.1080/17521440.2018.1548209)

Hardin, R. (2002) Trust and Trustworthiness *(New York: Russell Sage, 2002).*

Ivec, M., & Braithwaite, V. (2015). Applications of responsive regulatory theory in Australia and overseas: Update. *Regulatory Institutions Network Occasional Paper 23. Canberra: Australian National University.*

Jacobsen, C., Fosgaard, T. R., & Pascual‐Ezama, D. (2018). Why do we lie? A practical guide to the dishonesty literature. *Journal of Economic Surveys*, *32*(2), 357-387.

Jacquemet, N., James, A., Luchini, S., Murphy, J., & Shogren, J. F. (2019). Lying and shirking under oath. *ESI Working Paper* 19-19. Retrieved from https://digitalcommons.chapman.edu/esi\_working\_papers/278/

Kaplow, L. (1992). Rules versus standards: An economic analysis. *Duke Lj*, *42*, 557.

Kemper, N., Nayga, R. M. Jr., Popp, J., & Bazzani, C. (2016). The effects of honesty oath and consequentiality in choice experiments. In *Selected Paper prepared for presentation at the Agricultural & Applied Economics Association’s 2016 AAEA Annual Meeting, Boston, Massachusetts, July 31–August 2.*

Korobkin, R. (1997). Status quo bias and contract default rules. *Cornell Law. Review*, *83*, 608.

Kouchaki, M., Gino, F., & Feldman, Y. (In press). The Ethical Perils of Personal, Communal Relations: A Language Perspective. *Psychological Science*. Available at <https://www.researchgate.net/publication/336021711_The_Ethical_Perils_of_Personal_Communal_Relations_A_Language_Perspective>

Kucher, M. & L. Götte (1998). Trust Me. An Empirical Analysis of Taxpayer Honesty. *finanzarchiv* 55: 429-444.

Laske, K., Saccardo, S., & Gneezy, U., (2018). Do Fines Deter Unethical Behavior? The Effect of Systematically Varying the Size and Probability of Punishment. (April 5, 2018). Available at SSRN: https://ssrn.com/abstract=3157387 or http://dx.doi.org/10.2139/ssrn.3157387

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

Möllering, G. (2006). *Trust: Reason, Routine, Reflexivity*. Emerald Group Publishing.

Moyson, S. (2017). Trust in regulatory regimes, by F. Six and K. Verhoest. *Journal of Trust Research*, 7(2), 226–229. doi:10.1080/21515581.2017.1364028

Nickerson, D. W., & Rogers T. )2010(. Do you have a voting plan? Implementation, intentions, voter turnout, and organic plan making*. Psychol. Sci*. 21:194–99

Olsen, A. L., Hjorth, F., Harmon, N., & Barfort, S. (2018). Behavioral dishonesty in the public sector. *Journal of Public Administration Research and Theory*. doi:10.1093/jopart/muy058

O'Reilly, C. A., & Chatman, J. (1986). Organizational commitment and psychological attachment: The effects of compliance, identification, and internalization on prosocial behavior. *Journal of applied psychology*, *71*(3), 492.

Pruckner, G. J. & Sausgruber, R. (2013). Honesty on the streets: a field study on newspaper purchasing. *Journal of the European Economic Association*. 11, 661–679.

Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The transsituational influence of social norms. *Journal of personality and social psychology*, *64*(1), 104.

Ruedy, N. E., & Schweitzer M. E. (2010). In the moment: The effect of mindfulness on ethical decision making. *Journal of Business Ethics* 95, no. 1: 73-87.

Shalvi, S., Eldar, O., & Bereby-Meyer, Y. (2012). Honesty requires time (and lack of justifications). *Psychological science*, *23*(10), 1264-1270.

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and social psychology bulletin*, *37*(3), 330-349.

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, USA*, 109, 15197–15200

Stevens, B. (1994). An analysis of corporate ethical code studies:“Where do we go from here?”. *Journal of business ethics*, *13*(1), 63-69.

Sunstein, C. R. (2013). Nudges vs. shoves. *Harv. L. Rev. F.*, *127*, 210.

Swann Jr, W. B., Rentfrow, P. J., & Guinn, J. S. (2003). Self-verification: The search for coherence. *Handbook of self and identity*, 367-383.

Tenbrunsel, A. E., & Messick, D. M. (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly*, *44*(4), 684-707.

Torgler, B., (2003). Tax Morale, Rule-Governed Behaviour, and Trust. *Constitutional Political Economy*, June, 14(2): 119–140.

Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations 1. *Econometrica*, *76*(6), 1467-1480.

Verschuere, B., Meijer, E. H., Jim, A., Hoogesteyn, K., Orthey, R., McCarthy, R. J., Yıldız, E. (2018). Registered replication report on Mazar, Amir, and Ariely (2008). *Advances in Methods and Practices in Psychological Science*, 1, 299–317

World Bank Group (2019). *Doing Business 2019 16th Edition: Comparing Business Regulations for Domestic Firms in 190 Economies.* Retrieved October 1st 2019 from https://www.doingbusiness.org/content/dam/doingBusiness/media/Annual-Reports/English/DB2019-report\_web-version.pdf

1. We had to exclude 124 participants who failed a reading comprehension question, as well as 11 others who had duplicate IP addresses. However, the pattern and significance of the results do not change when all responses are included. [↑](#footnote-ref-2)