**Behaviorally informed regulation of dishonest behavior: Understanding how, when and why can pledges 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 pertains to the degree of trust authorities can place in people’s self-reports. Because regulators are typically risk-averse, trust is not easily conferred, and worldwide report show that people often face excessive requirements when applying for permits, licenses, etc. However, recent research in behavioral ethics suggests that when people are asked to ex-ante pledge to behave ethically, their level of dishonesty and incompliance can be reduced. If such pledges indeed prevent people from cheating in their self-reports, regulators could then relax many bureaucratic hurdles - with their huge costs to market efficiency, voluntary compliance and trust - without major risks to the public interest. Although some evidence (including our own preliminary results) show that ex-ante pledges can sometimes curb unethical behavior, no study to date has systematically examined the conditions and factors that determine when and how can pledges reduce dishonesty, and when might they backfire. First, it is unclear whether and how pledges’ fare better than command-and-control regulation (e.g., fines or sanctions on cheating) and might pledges’ effect decay over time (which would put their efficacy in question). Moreover, it is theoretically unclear why pledges work and what are the psychological mechanisms that underlie their effects. There is also no known theoretical framework or principles for 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 at the same time provide valuable behavioral insights for regulatory policy and governance. By this, our proposed research aims to provide a clearer theoretical and useful knowledge on when, how and why could pledges prevent dishonesty, improve regulatory and enforcement choices, and help rebuild trust in governance and public policy.

# **A. Scientific background**

Much of regulators’ need and effort to control human behavior is related to the ability of governments to trust their citizens (Moyson, 2017). It is a common dilemma that 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 increased regulatory burden (Anania & Nisticò, 2004; Gilligan, 2018). Due to regulators’ common inclination for risk-aversion, states often prefer to not confer trust upon those regulated and do whatever they can to prevent risks to the public interest and ensure safety at all costs (e.g. Bews & Rossouw, 2002; Cohn, Fehr & Maréchal, 2014). For instance, the recent World Bank “Doing Business” report identifies many hurdles and obstacles people face in order to get permits to open a business, register a property or obtain a financial credit (World Bank Group, 2019). These situations create a lot of risk for regulators and they employ a whole array or burdensome requirements to reduce the likelihood of mistakes. This mistrust results in a suspicious government perception of individuals’ self-reports, because the government cannot infer ex-ante the proportion of the population who will exploit the option to self-report in order to make fraudulent claims (e.g. Feldman, 2018). The end result is often a highly bureaucratic mechanism that reduces risk to regulators, but hampers growth.

One solution to this problem could be found within the responsive regulation paradigm (Ayres and Braithwaite, 1992). This paradigm explores the value of adopting a dynamic regulatory strategy, where trust between regulators and regulatees is the first choice followed by escalation to more punitive regulation if that trust is abused (Braithwaite & Makkai, 1994). To tackle dishonesty under this paradigm, policy makers could move beyond 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 relax their monitoring by using less forceful and less coercive measures of ensuring honest and ethical conduct. An important concept within this paradigm is the “enforcement pyramid”, which suggests a careful escalation in means used by enforcers, where consequently at every compliance dilemma regulatees know that regulators could move to harsher means (Ayres & Braithwaite 1992). By focusing first on regulatory means that allow people to feel trustworthy and to engage in voluntary compliance, trust has a chance of being formed (Feldman 2018; Möllering 2006). For example, when an entrepreneur wishes to open a new business, or when citizens report their annual expenses for tax purposes, the government does not always have to insist that they provide all relevant materials beforehand and examine their documents in scrutiny before approving their application. In some cases, governments can make do with asking applicants to guarantee, in advance, that their report or application is accurate and honest, and invest more resources in auditing and sanctioning afterwards. Such *ex-ante pledges* can clearly reduce the administrative burden imposed on people, and also relieve some of the efforts imposed on regulation and licensing authorities (e.g., Kucher & Götte,1998; Torgler, 2003).

However, relying on pledges instead of mandatory checks also raises the concern that some people might try to take advantage of the situation and make false reports dishonestly, in order 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. Jacobsen, Fosgaard, & Pascual‐Ezama, 2018; Gerlach, Teodorescu & Hertwig, 2019), and this cheating is associated with excessive damages 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, and people are then less likely to behave dishonestly (Bazerman & Gino 2012). Namely, when participants in a study were asked to 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 is costlier and thus considered more honest; Shu, Mazar, Gino, Ariely & Bazerman, 2012).

Although it has yet to receive a systematical research, the basic effect of pledges on reducing dishonesty has been shown by a few additional studies. Beck, Bührn, Frank & Khachatryan (2018) found, using the “die under a cup” paradigm, in which participants roll a die secretly to determine their participation payment (e.g., Shalvi, Eldar, & Berbey-Meyer, 2012), 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 in order to enrich themselves. Similarly, Jacquemet, James, Luchini, Murphy & 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 their neighbors’ (Cagala, Glogowsky & Rincke, 2019).

Other studies examined the effects of pledges indirectly 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” and measured how much they would be willing to pay for an increase in taxes that would reduce carbon emission. 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 lower variance). Similarly, Kemper, Nayga, Popp & Bazzani (2016) found that an honesty oath led to significantly lower WTP values, which are regarded 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.

Apparently, pledges seem to be a potentially very useful tool for regulators in their attempts to balance regulatory efficiency and the “ease of doing business” with protecting the public interest from unethical behaviors. Indeed, reports suggest that some countries have already installed responsive regulatory approaches in some of their policies, and some of them used versions of ex-ante pledges or affidavits (Ivec & Braithwaite, 2015). However, because they were combined within larger reforms that included many other changes, it is hard to discern the actual consequences of using pledges from these policy changes. 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 would pledges prevent dishonest behavior as well as understand under which conditions might pledges backfire and should thus be avoided. Clearly, the studies conducted to date do not yet allow us to answer these important questions.

**B. Research objectives & expected significance**

While the above studies suggest that pledges can reduce dishonesty in some cases, many important insights about the actual effectiveness of pledges, its scope, scalability and possible boundaries are still missing. This makes the usage of pledges to be highly limited and decontextualized for several reasons. First, the effects of pledges have only been tested on one-shot or short-term decisions. However, the temptation or opportunity for cheating in real life can be long lasting, and also far apart from the time of the pledge. Second, previous 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, authorities will often use other, more traditional mechanisms of enforcement and deterrence from lying such as financial sanctions (fines), legal sanctions and other penalties, in conjunction with pledges or affidavits. Cheating in those situations would entail serious repercussions if caught, and the threat of financial or other sanctions can reduce cheating on its own (e.g., Laske, Saccardo, & Gneezy, 2018). Adding a pledge might not then offer a marginal benefit when such fines are present. Thus, the effect of a pledge must be tested and evaluated against such traditional sanctions, in order for it to have any ecological validity and implementation ability. Moreover, research to date has yet to provide conclusive evidence that would explain why pledges actually work (when they do), and what are the underlying mechanisms that could be driving their ability to reduce dishonesty. It is thus clear that a more systematic research program is required in order to fully understand the scope, boundaries and moderating factors that would explicate the potential effects of pledges on reducing dishonesty.

Studying the mechanism through which pledges operate could not just help reduce unethicality but also help understand how to design the pledges in ways which will maximize its efficacy in the contexts which they are expected to operate. Identifying the conditions in which pledges could be effective could also eventually contribute to building trust between the people who are interested in behaving ethically and their government. Crystalizing our understanding of how pledges function could help understand how to reach a situation where “good” people could get the relief in regulatory burden without heavy reliance on enforcement mechanisms which reduce intrinsic compliance motivation (Bowles & Polania-Reyes, 2012) while still not jeopardizing the public safety and enabling mutual trust to be built (e.g. Hardin, 2002).

Our research would thus focus on five main questions pertaining to the use of pledges to curb dishonest behavior: a) evaluate the effect of pledges in reducing dishonesty in repeated situations, and in comparison to traditional measures of fines and sanctions; b) uncover the temporal pattern of the effect of pledges over time; c) understand the psychological mechanism(s) underlying the (discovered) effects of pledges on dishonesty; d) systematically explore and identify how to optimally design pledges (e.g., in respect to language and content); and lastly, generalizing to a more institutional level, e) test how pledges can help rebuild or foster trust between the regulated parties. In the following we elaborate on each of these research directions in more detail.

**C. Detailed description of the proposed research**

1. **Pledges vs. fines**

The first part would focus on empirically testing and determining the effect of pledges on reducing dishonesty in repeated decisions (instead of one-shot choices, which were the focus of previous studies), and compared to traditional measures of preventing dishonesty such as fines and sanctions. Even those who believe that many people could be trusted to follow their pledges will admit to the need to still maintain some monitoring mechanisms with penalties for inaccuracies. Thus, understanding the interaction between sanctions and pledges is hence crucial. In the current research on sanctions, there is 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 (Tenbrunsel & Messick, 1999). On the other hand, there are studies which suggest that in some contexts, incentives could create a process of “crowding-in” by which external monetary interventions actually increase voluntary compliance and pro-social behavior (e.g. Galbiati & Vertova 2014). Thus far, none of the studies on crowding-out and crowding-in focused on honesty, but rather on helping others and donation in public good experiments. Studying the joint effect of incentives and pledges on honesty is hence an important step in the realistic understanding of how pledges operate.

To examine the effect of pledges compared to traditional fines and sanctions, we will build upon the basic design that is detailed in our preliminary study’s description. Namely, we will have participants perform a task that holds an opportunity to cheat in order to gain more money by over-reporting their performance on the task (Mazar, Amir, & Ariely, 2008). The task involves solving simple calculus problems sequentially, for a small bonus payment per problem, while the bonuses are earned differently between 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. While this design does not enable measuring how much (if at all) each participant cheated individually, it does allow to compare 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 recent decade or more (Gerlach et al., 2019). Recently, using a somewhat different, but conceptually similar task of the “die under a cup”, researchers were also able to prove that such behavioral abstract tasks are strongly correlated with actual behaviors of dishonesty by public sector employees (Olsen, Hjorth, Harmon, Barfort, 2019) as well as predict degrees of rule violations between countries (Gacher & Schultz, 2016).

Building upon this basic cheating task allows us 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 in the beginning (as in Shu et al., 2012) and we will measure how the addition of the pledge curbed the dishonesty gap between the cheating and the control groups. Then, we will vary the explicit probability of participants’ responses being audited for correctness (as in Laske et al., 2018), and also vary the threat and level of the sanction (e.g., a fine) that would be imposed on those caught cheating in the audit. For example, a participant in one of these conditions could be told that a random 10% of their answers will be audited and that if, on these audits, they will fail to provide the actual solution they will suffer a fine of, say, half of their earnings or a fixed pre-defined sum.

Our preliminary results show that a pledge can indeed reduce dishonest over-reporting in a considerable and consistent manner. This was found when the chance of an audit was relatively small (10%) and the fine used was minimal (only lose the earning of the audited problem). Actually, increasing the fine to its maximum level (losing all earnings) did not, in our preliminary study, change the effect of the pledge. However, other levels of probabilities and fines, including other sanctions (such as bans from future participation) must also be examined. Thus, our first set of studies would follow the basic design of the preliminary study and vary a) the gains to be earned from cheating, b) the probability of being audited, and c) the size of the fine. This will enable us a clearer understanding of the boundaries within which pledges can effectively reduce dishonest behavior and when, if ever, pledges might not be effective or even backfire.

1. **Temporal decay**

The second part would focus on the temporal aspects of the effect of the pledge and examine whether and how it might decay over time. Opposed to the common version of asking people to confirm, post-hoc, that their report was truthful, pledges are solicited before the action is taken. For example, students may be asked to pledge they would not cheat in the test before it begins; managers and public servants could be asked to pledge their ethicality before taking office; witnesses are asked to pledge their truthfulness before providing their testimony, etc. Sometimes the pledge is solicited in close temporal proximity, right before the action, and sometimes a large span of time might pass after the pledge until the person is confronted with the temptation to behave dishonestly. Studies that examined the effects of pledges to date focused only on the former case, when the pledge is taken right before the action (e.g., Shu et al., 2012). However, in real-life there could often be 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 as well as from a prescriptive point of view of when should pledges be taken, and when should they be renewed or reminded.

The research on ethical decision making provides important insights into the potential fading of ethical nudges over time. On one hand, being aware to the ethical meaning of one’s behavior can increase honesty (Ruedy & Schweitzer 2010; Ayal, Gino, Barkan & Ariely, 2015). Thus, if one assumes that awareness reduces over time, so should the effect of the pledge on curbing dishonesty diminish over time. In contrast, research on preference change (e.g. Vanberg, 2008), as well as on personal commitment and internalization of morality (O'Reilly & Chatman, 1986) support a view where the process through commitment to ethicality will have a sustainable effect on people’s ethical decision making, unrelated to their level of activated awareness to morality. This suggests that pledges’ effects will not reduce with 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.

Our goal in this part would thus be to systematically examine if and how pledges’ effects fluctuate over time. Specifically, we will examine whether the effect of pledges decay over short vs. long periods of time, and to what degree. Moreover, we will examine how does the interval between the pledge and the time of action affect the efficacy of the pledge. In this context, we will also explore whether reminders, which have been often 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, as well as b) add varying time intervals between the time when the pledge is taken and when the opportunity to cheat occurs. The post-pledge time interval will be filled with either a non-related 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 which could occur that might have a positive or a negative effect on the potency of the pledge, in order to estimate these moderators’ role on the effects of pledges.

1. **Why pledges work**

Next, we aim to examine and 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 or should not work in reality (cf. Jacobsen et al., 2018). Current research suggest that pledges’ effects are attributed to reminding people about their morality values. Shu et al., (2012), who found that signing at the beginning leads to less dishonest reporting or over-claiming, also found in another study that signing at the beginning increased the frequency of ethicality-related words in a word-completion task, suggesting 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, while 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 different labs worldwide, showed disconfirming results: in none of the cases, did this type of moral reminder reduced cheating significantly, and it even increased cheating in one (Verschuere et al., 2018).

Following the literature on behavioral ethics, we can identify several behavioral and psychological accounts that might explain pledges’ effects. First, despite the failure to replicate the Ten Commandments nudge, pledges could indeed simply act as morality reminders, highlighting the importance and desirability of behaving in an ethical manner (Shu et al., 2011). However, a different possibility could be that pledges might act as signals of social norms, encouraging people to not deviate from others’ behavior and 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 receives more saliency than it had before. Another, 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 in self-consistency (Swann, Rentfrow & Guinn, 2003) . A fourth reason why pledges work might lie in 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 might make some aspects of the situation, such as the perceived risk of getting caught and/or the perceived penalty for dishonest behavior, more salient and clearer, compared to when people are not required to actively sign or affirm their honest behavior ex-ante.

Theoretically, any one or more of these accounts might explain pledges’ effects. Thus, our third goal in the proposed research would be to conduct a series of experimental studies that could isolate and test each of the above accounts. For instance, to test whether pledges affect the perceived social norm, we will ask participants in the study to indicate what is their personal prediction about how many people would cheat under this situation. We will predict that the more people will believe others would cheat, the less of an effect the pledge would have. Similarly, we would 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 either highlighting the percent of people who cheated – to invoke a dishonest social norm - or the percent of people who did not cheat – to stress a social norm of honesty). We would design additional experimental manipulations to test the role of the other factors that could explain the effects of pledges, including their effects on moral values saliency, self-consistency, or perceived risks and penalties.

1. **Pledges’ optimal design**

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

Another important question in the design of pledges is whether it would be better to use a legal or formal language or is it better to use a more accessible language that lay people could more easily comprehend and relate to. 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 formal language over softer and less formal reference, for example 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 stay committed to their pledges. To confirm this hypothesis, we will devise several different, yet equivalent phrasing for the same pledge and randomly assign participants to the different phrasings, to examine which type of language (i.e., formal vs. informal, specific vs. general) would be most effective in curbing dishonesty.

1. **Can pledges enhance trust**

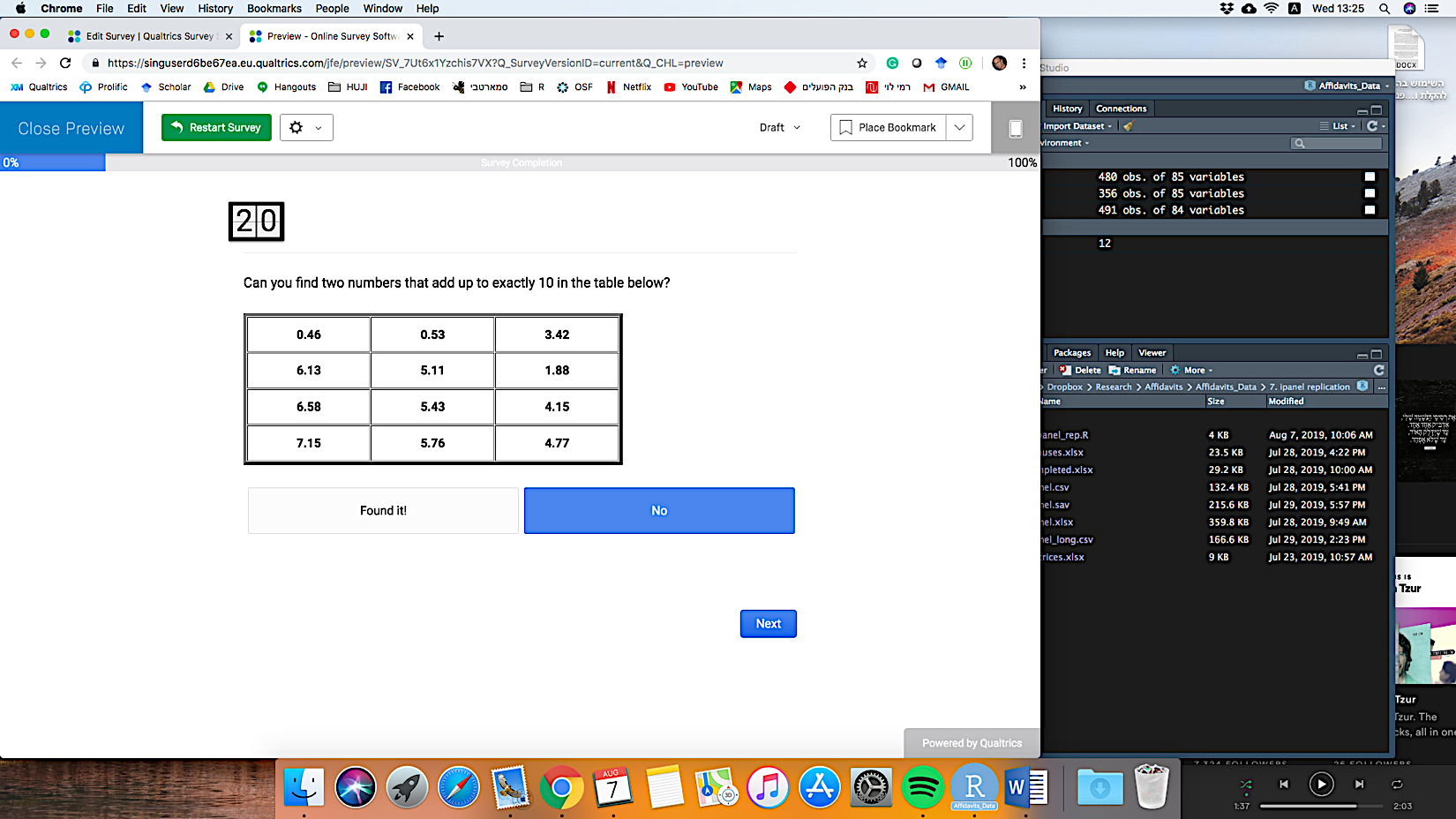
Lastly, we wish to explore the consequences of transitioning to a regulation model that uses pledges as a common and frequent tool. In this, we wish to explore how would pledges affect trust between the people and regulators over time: could pledges, being a trust-giving mechanism, invoke more trust from people and would that increased trust lead to more uses of trust-giving regulation, creating a repeating 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 would 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 the conditions that were either asked or were not asked to pledge to their honesty, controlling for actual performance (and payment) in the experimental task.

Additionally, to explore this in a more institutional level, we are in close contact with several government branches in Israel that plan to introduce pledges in some of their regulatory reforms. These include currently branches and agencies within the Israeli Ministries of Justice, Interior and Finance. With the cooperation of at least some of them, we plan to examine how people trust agencies before vs. after they introduce the use of pledges, instead of requiring ex-ante proofs, for providing permits, licenses, etc. We predict that as long as the pledges would indeed reduce bureaucratic burden, requiring pledges will be accompanied with increased levels of trust for the relevant agency.

**Preliminary Results**

To test the basic 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 a series of simple calculus problems, adopted from Mazar et al.’s (2008) “Martix Task”. In each problem in the task, 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 numbers that, when added up, accumulate to exactly 10. For each problem they solve they are awarded with 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 numbers of the solution after each problem they report as “solved”. When the study is done, control participants’ responses are checked for accuracy and they are paid only for the problems they solved correctly (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 they solved it or not. They are also told that there is a 10% chance for each problem that they might be “audited” and asked to provide the solution numbers 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 found in the control group is used as the measure of (group-level) cheating. To test the effect of pledges, with or without fines, we had three experimental conditions: a) self-report (without a pledge or fine), b) pledge only (without a fine), and c) pledge + fine (if caught cheating in the audited problem, participants lose all their earned bonuses).

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



All participants completed 10 problems consecutively in a randomized order. Participants in the conditions that include a pledge were asked to re-type the following sentence to indicate they agree 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) between conditions with 95% confidence intervals around the means. Participants in the control group correctly solved 32.5% (SD=24.5) of the problems correctly. In contrast, participants in the self-report group, that had no pledge or fine, reported solving 65.3% (SD=25.8) of their problems. That is, the average cheating rate in this group was about 33 percent points (or 200%). However, when participants were asked to pledge beforehand, they reported solving only 46.7% (SD=23.3) of the problems, reducing the cheating rate to 14 percent points (or about 130%) only. When participants were asked to pledge, and were warned of a maximal fine, report rate was similar at 50.3% (SD=23.5). The overall differences between conditions were statistically significant, *F* (3, 352[[1]](#footnote-2)) = 28.49, *p* < .01. However, the difference between the last two conditions (pledge without or with a fine) was not statistically significant, *p* = 0.79.

Figure 2. Mean percent of problems reported as solved between the conditions (error bars are 95% CIs).

****

Control Self-report + Pledge + Pledge & Fine

Similar findings were found for the failure rates in the audited problem (recall that each problem had a 10% chance to be audited): 58.9% of those audited in the first experimental group (no pledge or fine) failed it (that is, provided a wrong 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 between 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 Ph.D. 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. Pe’er has access to a fully equipped behavioral laboratory that enables running 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 B.A. 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 Non-Formal Enforcement Strategies. Feldman has extensive experience in this domain and is aleading expert in behavioral ethics in the scientific world. His book – The Law of Good People - on the interaction between behavioral ethics and law, was published in CUP last year.

Both PIs are in close contact with top officials at relevant government offices, who took part in a recent workshop organized on the topic, where they expressed their interest and desire to collaborate in behavioral research on pledges as regulatory tools. Currently, we are discussing potential projects with selected officials at the Ministry of Justice (to implement pledges in custom regulations), the Tax Authority (to explore pledges in some reporting procedures) and others. As discussed above, the current state of the knowledge on the efficacy of pledges is still far from being satisfactory for policy makers and we believe that the research outlined in this proposal could strongly interact with those efforts, allowing us to validate some of the findings in the field and to also 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 actual regulators in the field in order to ensure that our findings provide both theoretical and applicable conclusions.

**Bibliography**

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.

Gachter, 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.

Mollering, 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 that failed a comprehension check question, as well as 11 others that had duplicate IP addresses. However, the pattern and significance of the results does not change when including all responses. [↑](#footnote-ref-2)