Frequency of enforcement is more important than the severity of punishment in reducing violation behaviors

Kinneret Teodorescu[a,1], Ori Plonsky[a], Shahar Ayal[b], and Rachel Barkan[c]

[a]Industrial Engineering and Management, Technion – Israel Institute of Technology, Technion City, Haifa 3200003, Israel; [b]School of Psychology, Reichman University (IDC), Herzliya 4610101, Israel; and [c]Department of Business Administration, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved September 5, 2021 (received for review May 6, 2021)

External enforcement policies aimed to reduce violations differ on two key components: the probability of inspection and the severity of the punishment. Different lines of research offer different insights regarding the relative importance of each component. In four studies, students and Prolific crowdsourcing participants (N_{total} = 816) repeatedly faced temptations to commit violations under two enforcement policies. Controlling for expected value, we found that a policy combining a high probability of inspection with a low severity of fines (HILS) was more effective than an economically equivalent policy that combined a low probability of inspection with a high severity of fines (LIHS). The advantage of prioritizing inspection frequency over punishment severity (HILS over LIHS) was greater for participants who, in the absence of enforcement, started out with a higher violation rate. Consistent with studies of decisions from experience, frequent enforcement with small fines was more effective than rare severe fines even when we announced the severity of the fine in advance to boost deterrence. In addition, in line with the phenomenon of underweighting of rare events, the effect was stronger when the probability of inspection was rarer (as in most real-life inspection probabilities) and was eliminated under moderate inspection probabilities. We thus recommend that policymakers looking to effectively reduce recurring violations among noncriminal populations should consider increasing inspection rates rather than punishment severity.

behavioral ethics | enforcement | decisions from experience | policy making | cheating

Texting while driving, jaywalking, littering, not wearing face masks, and neglecting to keep social distancing during the COVID-19 pandemic are but a few examples of seemingly negligible violations that accumulate fast, with potentially dire social consequences. The prevalence of these violations highlights the shortcomings of reasoning with or appealing to people's civic duty and point to the need for regulation and enforcement. External enforcement is composed of the probability of inspection and the severity of the punishment delivered upon detection (1). Clearly, the combination of complete monitoring and severe punishments is the fastest way to shape behavior. However, because of the limited resources devolved to monitoring and the negative consequences of severe punishment (e.g., reactance), policymakers usually resort to two main compensatory strategies. One solution opts for less frequent inspection but greater severity of punishment and counts on its potential for deterrence. The other solution opts for close monitoring and assumes that when the inspection rate is high, minimal or even symbolic punishment will suffice. The current paper examined the effectiveness of these two enforcement strategies in reducing violations.

Several theoretical perspectives are relevant here. The economic perspective views people as rational agents, whose actions are governed by cost–benefit analyses. Accordingly, people will commit violations if their expected utility (accounting for potential gains and losses) is positive (1). Becker's model predicts that, given equivalent expected values (EVs), risk-averse people will be more sensitive to the severity of punishment than to the probability of detection. In several experimental economics studies, researchers measured risk tendencies separately from the participants' reactions to various inspection probabilities and fine sizes. In these studies, participants were generally found to exhibit risk–aversion tendencies, and, in line with Becker's model for risk-averse individuals, they were more deterred by increases in punishment severity than by equivalent increases in the probability of inspection (2–5). The economic perspective therefore implies that delivering severe punishments, even rarely, is the most effective enforcement strategy.*

Experimental findings from the economic perspective typically rely on tasks in which participants are explicitly asked to choose whether to comply or violate, and violation choices are described as a morally neutral gamble, specified by a predescribed probability of inspection and fine. However, in natural settings, violations tend to be implicit, allowing for moral gray zones.

Significance

Ramifications of seemingly small violations, such as not adhering to COVID-19 regulations, accumulate fast with dire social consequences. The high costs of close monitoring and severe sanctions often lead policymakers to prioritize either the probability of inspection or the severity of punishments. Using common one-shot, descriptive settings, findings from experimental economics support the superiority of severe punishments, whereas findings from behavioral ethics highlight the role of internal rather than external enforcement. However, these settings are estranged from real-life environments in which learning about the external enforcement policy naturally occurs via repeated experience. Using a more ecologically valid, experience-based setting, we found robust evidence for the greater effectiveness of frequent small punishments over rare severe punishments in reducing violations.

Author contributions: K.T., O.P., and S.A. designed research; K.T., S.A., and R.B. completed literature review and conceptualization; K.T. performed research; K.T. and O.P. analyzed data; and K.T., O.P., S.A., and R.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

[1]To whom correspondence may be addressed. Email: kinnerett@technion.ac.il.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2108507118/-/DCSupplemental.

Published October 13, 2021.

*Notice that prospect theory (6) predicts an even stronger effect of rare severe punishment than the standard Becker model. This is because violation decisions represent a "mixed" gamble (i.e., they involve the possibility of both gains and losses), for which the prediction of prospect theory is highly contingent on the probability weighting function. Since the weighting function implies overweighting of rare events, it predicts even stronger deterrence in the case of rare severe punishment.

Downloaded at Palestinian Territory, occupied on November 30, 2021

www.manaraa.com

Accounting for moral malleability, a psychological perspective pits the external gain associated with violations against the internal cost of damage to one's self-perception (7, 8). Self-concept maintenance theory (9) suggests that aiming to perceive oneself as highly moral creates an internal barrier to acts of violation (10, 11). Supporting this notion, findings in behavioral ethics indicate that although participants commit violations, they do so to an extent far lower than what profit maximization would predict (9, 12–14). Importantly, because of the theoretical and experimental focus on internal factors that operate in the absence of actual detection or punishment (15), the effect of external enforcement has not been rigorously addressed. Thus, the psychological perspective does not make a clear prediction as to the relative importance of frequency versus the severity of punishments (we return to this point in *Discussion*). The psychological perspective does, however, predict that internal enforcement mechanisms (16, 17) will limit violation behaviors even in the absence of external enforcement (18, 19).

The economic and psychological perspectives both rely on studies that typically involve very few choices (15) and thus do not address the long-term effects of external enforcement. Extending experimental tasks to repeated settings introduces learning as a third perspective to consider. Research on repeated decisions from experience indicates that choice behavior is more sensitive to the frequency of experienced outcomes than to their magnitude (20, 21). Importantly, this line of research shows that when people learn from experience, they tend to behave as though they underweight rare outcomes (22–26). Altogether, these findings hint that rare, large fines may not have the intended outcome and suggest instead that frequent small fines should be more effective at decreasing violations.

In four studies, we aimed to determine which enforcement strategy would be more effective in a setting that incorporates the unique characteristics of all three perspectives. We utilized the dots task (27, 28), a perceptual task that poses repeated conflicts between accuracy and profit maximization (whenever incorrect responses yield higher payoffs than correct responses). This task enables the investigation of small, morally vague violation decisions. It consists of many trials, thus permitting the examination of violation behavior over time. More importantly, it can easily be modified for the incorporation and systematic manipulation of the probability of inspection and the severity of fines while also controlling for the EV of violations. This experimental setting therefore presents a simplistic simulation of real-life settings in which people know enforcement is possible and can implicitly learn the likelihood of being caught from experience (29).

Pilot Study

The pilot study compared violation rates under two partial-enforcement conditions: rare large fines versus frequent low fines. In both conditions, the EV of violations was identical. Two control conditions administered either no external enforcement or full external enforcement.

Methods.

Participants. Forty-two undergraduate students from the Technion participated in a 1-h laboratory study. Payment was contingent on performance (M = 46.0 Israeli New Shekel [ILS] and SD = 4.0). Each participant played 768 repeated trials, half of which contained a monetary temptation to commit a violation (i.e., overall, we analyzed 32,256 observations).[†]

[†]The laboratory pilot study (n = 42) was run in 2013 under a general Technion IRB approval for decision-making experiments in Ido Erev's laboratory. All participants provided informed consent.

Basic task. We utilized the dots task (27, 28) and presented 20 dots distributed unevenly between two adjacent rectangles. On each trial, participants were asked to decide whether there were more dots on the right or left rectangle. One side always yielded 1 point (0.01 ILS), whereas selecting the other side always yielded 10 points (0.1 ILS) irrespective of the correct answer. Thus, the payoff rule motivated "accuracy violations" such that participants could increase their profit by selecting the more rewarding side on tempting trials in which the correct answer was the less rewarding side. The more rewarding side was counterbalanced across participants and remained the same for each participant throughout the experiment. The experiment consisted of four "games" of 192 trials each. In each game, half of the trials presented more dots on the more rewarding side (nontempting trials), and half presented more dots on the less rewarding side (tempting trials). The order of trials in each game was random. Each trial started with a 0.5-s presentation of a fixation point, followed by a 1-s dots screen. Next, the participants were asked to indicate which side had more dots by clicking on one of two boxes on the screen. Last, a 1-s feedback screen presented the trial's payoff.

On top of this basic incentive structure, we modified the dots task to include conditions with and without inspection of the correctness of the answer. Without inspection, the feedback screen only presented the trial payoff. To signal inspection, the feedback screen turned red, and if an incorrect answer was detected, a fine appeared on the screen and was subtracted from the trial's payoff. Fig. 1 depicts two trials on the modified dots task: one without inspection and one with inspection.

Experimental design. The games tested the four enforcement conditions, which differed in the rate of inspection and the severity of the punishment. The no-enforcement (NE) condition corresponded to the original cheating condition on the dots task without inspection or fines. In the full-enforcement (FE) condition, each trial was inspected and each incorrect answer was fined −18 points. In the two partial-enforcement conditions, the inspection rate and punishment severity varied. In the HILS condition, 90% of the trials were inspected and each detected violation was fined −10 points. In the LIHS condition, only 10% of the trials were inspected, but each detected violation was fined −90 points. Therefore, the EV for an accuracy violation was identical in the two partial-enforcement conditions. In addition, within each of the two partial-enforcement conditions, the EV of an accuracy violation was identical to the fixed payoff that resulted from reporting the correct answer. Table 1 summarizes the experimental conditions. Utilizing a complete within-subject design, participants engaged in all the conditions (i.e., four "games" presented in random order).

Procedure. Participants signed a consent form and read the instructions explaining the perceptual task, the way points are awarded, and the exact exchange rate of points to monetary payment. Before engaging in the dots task, participants were further informed that inspection could occur and that in case of inspection, the feedback screen would turn red and that detection of an incorrect answer would be fined. However, we did not specify the enforcement policy; thus, participants had to infer the probability of inspection and the severity of punishment (i.e., the fine) based on their experience in each game. Participants were informed when one game was over and a new one began and that, in each game, the possibility of inspection, as well as the severity of the punishment, could be different.

Results. The rate of accuracy violations was calculated as the difference between the proportions of "beneficial errors" (i.e., selection of the incorrect but more rewarding side) and

Downloaded at Palestinian Territory, occupied on November 30, 2021

www.manaraa.com

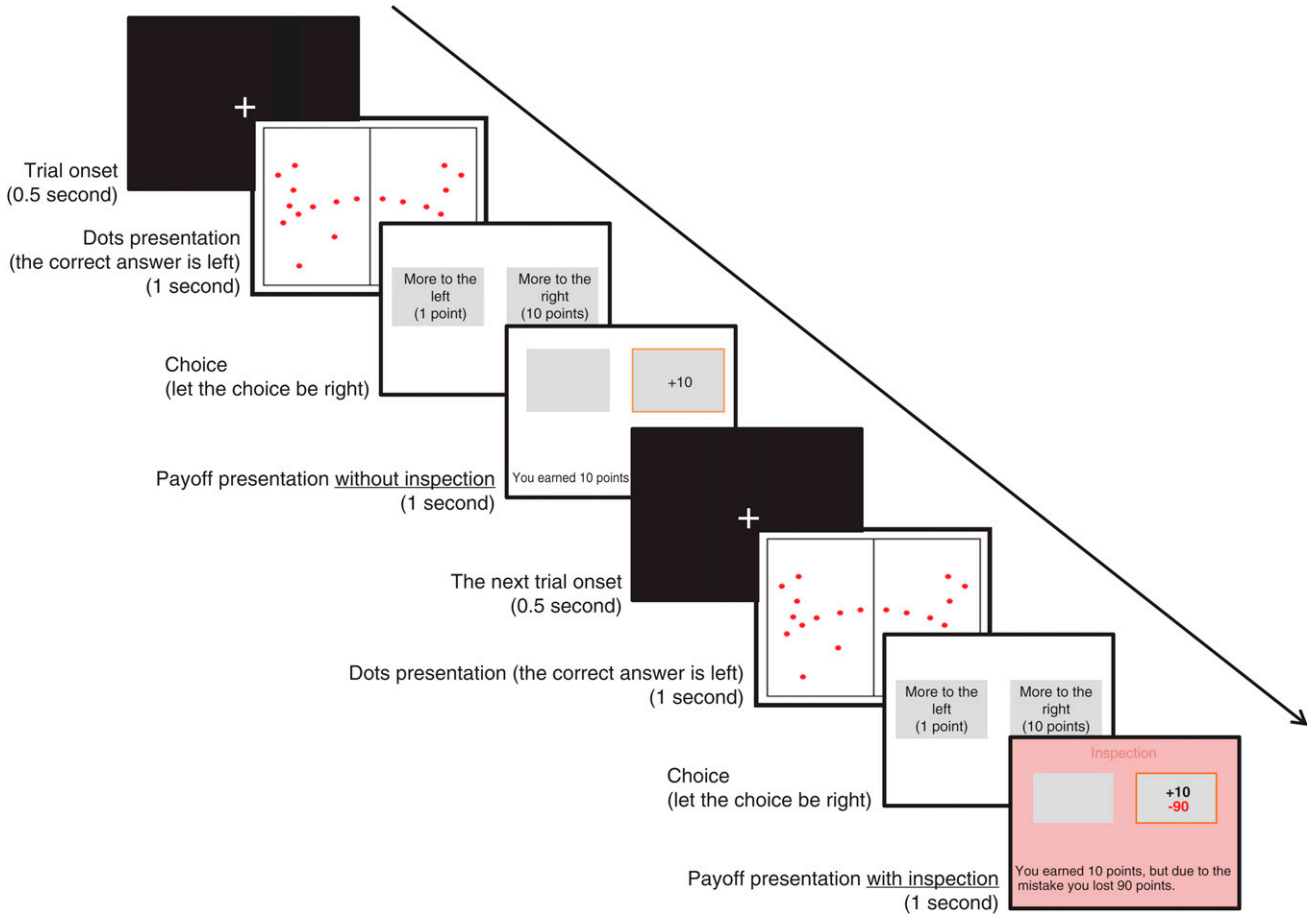

Fig. 1. Timeline example of two trials: the first trial without inspection and the second with inspection under an enforcement policy with severe punishment (fine: −90 points).

"detrimental errors" (i.e., selection of the incorrect and less rewarding side) (28). The mean rate of accuracy violations was the highest without enforcement (47.5% in the NE condition) and the lowest with full enforcement (3.0% in the FE condition). This suggests that accuracy violations were mostly intended (not random) and favored profit over accuracy. Note, however, that the rate of accuracy violations without enforcement indicates that participants took advantage of just below half of the tempting trials, lending support to the internal barrier suggested by self-maintenance theory. Not surprisingly, the two partial-enforcement conditions were less effective in reducing violations than full enforcement. Importantly, however, a clear advantage emerged for frequency of inspection over severity of punishment (despite the control for EVs). Consistent with the decisions-from-experience hypothesis, the rate of accuracy violations was significantly lower when the enforcement rule prioritized probability than when the enforcement rule prioritized severe punishment [8.53% versus 27.69% in the HILS and LIHS conditions, respectively, $t(41) = 5.9$, $P < 0.001$].

Main Studies

The main studies compared the two partial-enforcement policies directly in a between-subject design. Because of COVID-19 restrictions, these studies were conducted online.[‡] To increase power and reduce noise, we increased the number of tempting trials. In each study, participants engaged in two games. They first completed the original dots task without enforcement and then repeated the task under either the HILS or LIHS enforcement conditions. This design served to 1) measure the effectiveness of each enforcement policy in reducing accuracy violations and 2) categorize participants according to their baseline tendencies without enforcement. Several studies that have examined individual differences suggest that certain people react to incentives to lie, whereas others do not (14, 30, 31). Establishing individuals' baselines thus made it possible to identify participants who rarely misreport, even without enforcement, and therefore examine the effectiveness of partial enforcement on those participants for whom enforcement mattered. The preregistered hypotheses were 1) HILS enforcement will reduce violations more than LIHS (in line with the decisions-from-experience prediction and the pilot results) and 2) this effect will be most pronounced in participants with high baseline violation rates.

Study 1

Methods. All main studies were run during 2020 and 2021 and were approved by the Technion Institutional Review Board (IRB), approval number 2020-030. All participants provided informed consent.

Participants. A total of 202 participants were recruited via the crowdsourcing platform Prolific to the study, which lasted about 17 min. Participants received a show-up payment of £1.45 and a bonus based on the number of points they earned during the

[‡]The transition to online participants was not straightforward. We ran three online studies to calibrate the online paradigm. The results of these studies are detailed in *SI Appendix*.

Downloaded at Palestinian Territory, occupied on November 30, 2021

www.manaraa.com

Table 1. Experimental conditions in the pilot study

Condition	P(inspection)	Fine	EV (incorrect) when reporting the incorrect answer is tempting*
No enforcement (NE)	0	—	10
Low inspection high severity (LIHS)	0.1	90	1
High inspection low severity (HILS)	0.9	10	1
Full enforcement (FE)	1	18	−8

*Compared with EV (correct) = 1 for tempting trials in all conditions.

task (M = £1, SD = 0.29). The number of participants was determined a priori and preregistered (at least 94 participants in each group) according to a power analysis based on a preliminary study (available in *SI Appendix*). The preregistration (available at https://aspredicted.org/blind.php?x=4ux6ja) also describes a predetermined exclusion criterion omitting participants who made more than 35% detrimental errors in the baseline NE game from the analysis.[§] Six participants were excluded based on this criterion. The final sample was composed of 196 participants (104 in HILS and 94 in LIHS).

Experimental task. We employed the dots task from the pilot study, with several changes aimed to increase power and make accuracy violations at baseline more tempting to avoid a floor effect. First, we replaced the vertical line with a diagonal and always presented 17 versus 13 dots on each side on each trial (the location of the dots within each triangle was random, thus naturally creating varied difficulty levels but with reduced variance). Second, the payoff for the less rewarding side was reduced to zero points. Third, each game was composed of 120 trials, 80 of which were tempting (the incorrect answer was rewarding) and 40 nontempting (the correct answer was rewarding). Last, we adjusted the inspection feedback such that under inspection, if the answer was incorrect, the feedback screen turned red (and a fine was subtracted from the trial payoff), but if the answer was correct, the inspection screen turned green.

Experimental design. All participants started with an NE game and were then randomly assigned to either the HILS or the LIHS condition. The HILS condition had a 0.90 probability of inspection and a fine of −11 points. The LIHS condition had a 0.10 probability of inspection and a fine of −99 points. We adjusted the fines to keep the EV of accuracy violation on the tempting trials identical in the two conditions (i.e., EV = 0.1, which is slightly larger than the zero points obtained from reporting the correct answer).

Procedure. The procedure was identical to the pilot study with the following minor changes to adapt to the online platform. Participants were explicitly instructed to answer as accurately as possible but were also told that "because this task is not easy, your submission will not be rejected even if you make many mistakes." All the participants started with an NE game. Next, the possibility of enforcement was explicitly introduced, and participants played a second game according to the experimental condition. Upon completion, we asked participants to guess the goal of the study and whether they would have acted differently if they had to do the task again. Participants were then notified about their bonus payment and received their task–completion code.

[§]This exclusion criterion was based on the online pilot studies (*SI Appendix, Part 1*) and was aimed to exclude extremely inattentive participants and/or participants with extremely low perceptual ability, for whom the hypothesis would not apply because their responses would mostly be random. Importantly, based on this criterion, we excluded fewer than 3% of participants in all main studies.

Data analysis. As in the pilot study, accuracy–violation rates were calculated as the difference between "beneficial errors" (percentage of incorrect answers in tempting trials out of all tempting trials) and "detrimental errors" (percentage of incorrect answers in nontempting trials out of all nontempting trials). Given the within-subject nature of the experimental design, the dependent variable for the main analysis was the difference between accuracy–violation rates in the enforcement and baseline games. A larger difference indicated greater effectiveness of enforcement, since we expected both enforcement conditions to reduce accuracy violations as compared to their respective baselines. Based on the results of the pilot and consistent with the decisions-from-experience hypothesis, in our main analysis, we expected a larger difference in the HILS enforcement condition and a smaller difference in the LIHS enforcement. A secondary analysis further investigated the expected effect of enforcement on participants with different accuracy–violation baseline rates in the first game without enforcement.

Results.

Testing the main hypothesis. Fig. 2*A* presents the main finding. On average, accuracy–violation rates decreased by 21.3% under the HILS enforcement that prioritized probability (SD = 25.7; from 33.6% in the NE game to 12.3% in the with-enforcement game). In contrast, accuracy–violation rates decreased by only 10.3% under the LIHS enforcement that prioritized punishment severity (SD = 21.1; from 32.7% in the NE game to 22.4%). Because a Shapiro–Wilk test rejected the assumption of normality for the accuracy–violation rates ($P < 0.001$), and consistent with our preregistration plan, we used a Mann–Whitney U test to test our hypothesis. Supporting our hypothesis, the test indicated that, compared to baseline, the reduction in accuracy–violation rates in the HILS condition (median = 15% and range [−18, 133]) was greater than in the LIHS condition (median = 7.5% and range [−25, 90]), U = 6,304, $P < 0.001$, one-tailed test, and effect size $r = 0.25$.

One possible concern was that the inspection screens could have confounded the experience of enforcement with simple accuracy feedback. That is, frequent inspection (HILS) could have improved participants' perceptual ability, which in turn could have led to fewer accuracy violations in both the tempting and nontempting trials. To test for this possibility, we examined participants' accuracy in the nontempting trials. If the feedback improvement explanation holds, there should be a greater improvement in accuracy between the first and the second block in the HILS condition compared to the LIHS condition. There was no evidence for this. Surprisingly, a Welch's two-sample t test for the nontempting trials revealed a significant difference between the two groups but in the opposite direction (M = 0.9%, SD = 6.0 versus M = −2.4%, SD = 9.9 in the LIHS and HILS conditions, respectively), $t(173.5) = -2.87$, and P = 0.005. These results rule out the alternative explanation of simple feedback improvement.

Downloaded at Palestinian Territory, occupied on November 30, 2021

www.manaraa.com

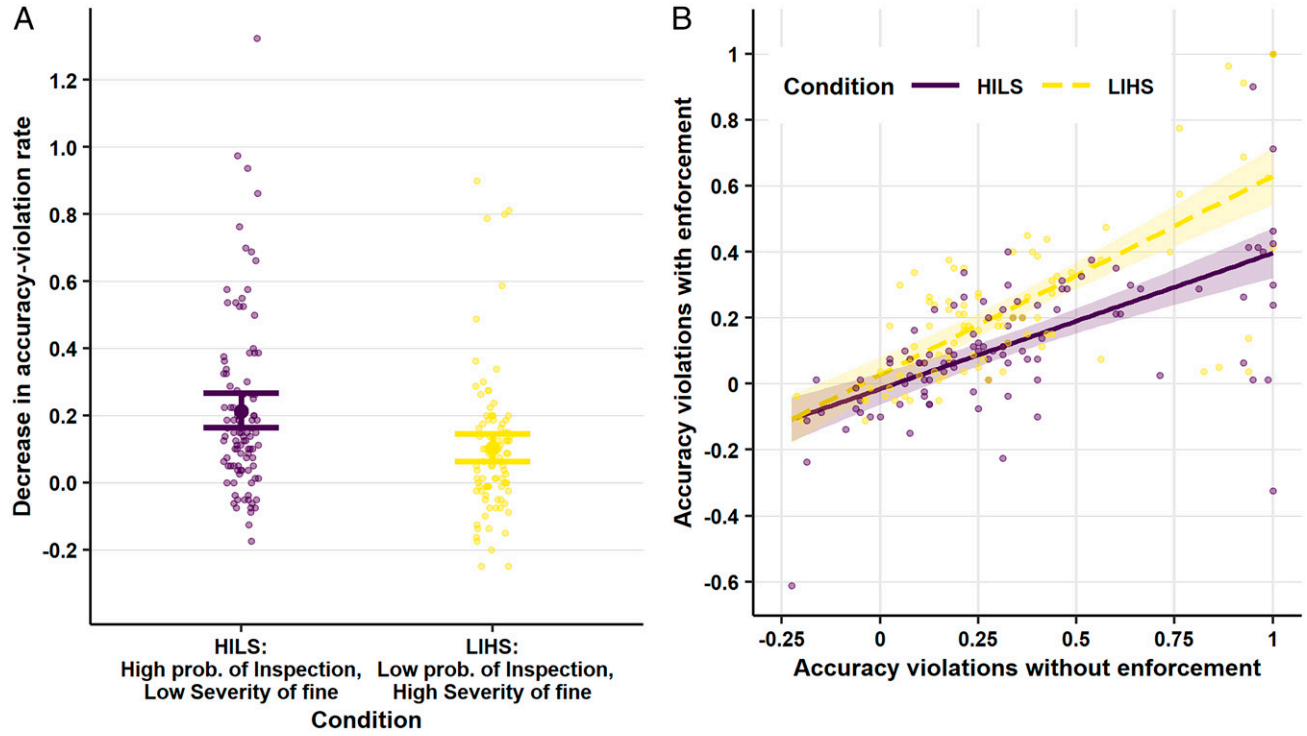

Fig. 2. Results of Study 1. Small dots depict individual participants. (*A*) Decrease in accuracy–violation rates from baseline to the enforcement block by condition. Error bars represent the 95% bootstrapped CI for the mean. (*B*) Regression lines of the accuracy–violation rates in the enforcement blocks on the accuracy–violation rates in the NE block by condition. Shading around the lines shows the 95% CIs.

Testing the secondary hypothesis. A modification to the preregistered regression analysis was made.[¶] The final regression analysis predicted the accuracy–violation rates in the second game by the enforcement condition (HILS/LIHS), the continuous accuracy–violation rates in the NE game, and their interaction. The interaction was significant: $F(1,194) = 5.79$, $P = 0.017$, $b = 0.19$, 95% CI (0.03, 0.35), and partial $f^2 = 0.03$. The fitted regression lines for the two conditions (Fig. 2*B*) showed that the difference between the two enforcement conditions was more pronounced when the accuracy–violation rates in the NE game were higher. Specifically, the simple slopes for the accuracy–violation rates with enforcement on the accuracy–violation rates without enforcement were 0.41, 95% CI (0.31, 0.51) for HILS, and 0.60, 95% CI (0.48, 0.72) for LIHS.

Study 2

In many real-life situations, the severity of sanctions is known ahead of time. That is, although the probability of inspection is unknown (29), the magnitude of the fine is sometimes public knowledge (e.g., via traditional or social media platforms). Accordingly, arguments in favor of severe sanctions are rather common and usually emphasize the deterrence effect. Moreover, studies in behavioral economics suggest that the presentation of small sanctions might constitute too low a price, which could legitimize violations (32). Thus, knowing the severity of the punishment in advance might undermine the effectiveness of enforcement that prioritizes probability. Nevertheless, findings from the literature on decisions from experience consistently indicate that repeated ongoing experience with feedback eliminates the initial effects of descriptive information (25, 33–35). We therefore hypothesized that the observed advantage of prioritizing probability (HILS) over severity of punishment (LIHS) would hold in the long run even with a priori information about the magnitude of the fine.

Methods.

Participants. A total of 196 participants who did not participate in the previous study were recruited via Prolific to complete this follow-up study. They were paid £1.45 for participating and were given a bonus contingent on their cumulative points earned during the experiment (M = £1.02 and $SD = 0.29$). Four participants were excluded based on the exclusion criterion of 35% "detrimental errors" or more, resulting in 94 versus 98 participants in the HILS and LIHS conditions, respectively.

Task and procedure. We used the same task, design, and procedure as in Study 1 with one exception: the addition of information about the size of the fine (the probability of inspection remained unspecified). Participants now expected fines of −11 or −99 points in the HILS and LIHS conditions, respectively.

Results.

Testing the main hypothesis. The findings replicated the results of Study 1, although the effects were smaller. Fig. 3*A* presents the main result. In the HILS condition, the mean accuracy–violation rate decreased by 21.8% ($SD = 28.6$; from 35.4% in the NE game to 13.6% in the with-enforcement game). In the LIHS condition, the accuracy–violation rates decreased by 14.6% ($SD = 24.0$; from 36.9% in the NE game to 22.3% in the with-enforcement game). Again, the Shapiro–Wilk test rejected the assumption of normality for the accuracy–violation rates ($P < 0.001$). A Mann–Whitney U test indicated that the reduction of accuracy–violation rates in the

[¶]The preregistered regression analysis included a categorial division of participants into three "types" depending on their baseline accuracy–violation rates. The interaction was not significant, probably due to the arbitrariness of the types and the loss of power associated with that categorization. This analysis can be found in *SI Appendix, Part 2*. The modified and more appropriate analysis replaced the intended three-level type categorization with a continuous variable consisting of the individual accuracy–violation rate scores.

Downloaded at Palestinian Territory, occupied on November 30, 2021

www.manaraa.com

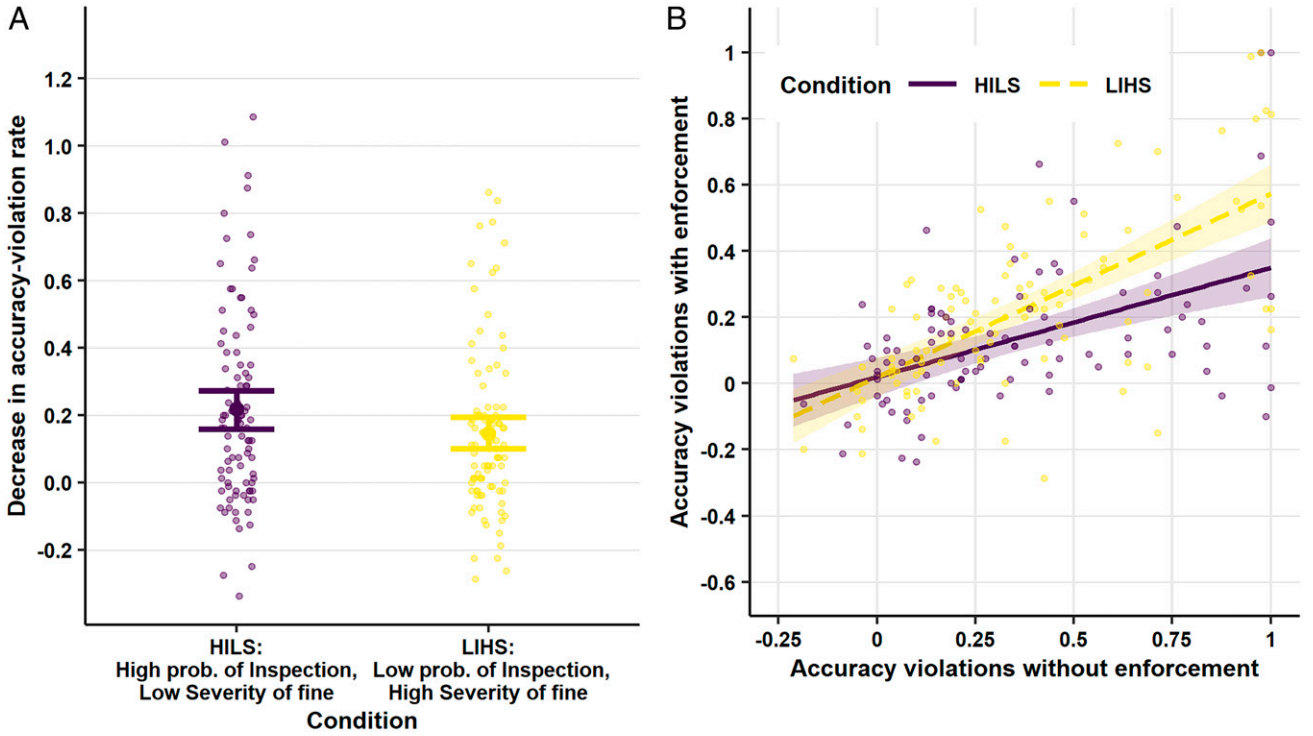

Fig. 3. Results of Study 2. Small dots depict individual participants. (*A*) Decrease in accuracy–violation rates from baseline to the enforcement block by condition. Error bars represent the 95% bootstrapped CI for the mean. (*B*) Regression lines of the accuracy–violation rates in the enforcement blocks on the accuracy–violation rates in the NE block by condition. Shading around the lines shows the 95% CIs.

HILS condition (median = 18.1% and range [−34, 109]) was greater than in the LIHS condition (median = 10% and range [−29, 86]), U = 5,291, $P = 0.038$, one-tailed test, and effect size $r = 0.128$. This finding thus further supports our hypothesis that even when the severity of the fine is provided explicitly as a means of deterrence, enforcement with a high probability of inspection and low fines is still more effective in reducing violation behavior than enforcement with severe fines but a low probability of inspection.

Again, we found no evidence that the frequent inspections caused participants in the HILS condition to make fewer detrimental errors than in the LIHS condition. A Welch's two-sample t test revealed no difference between the two groups ($M = -1.8\%$ and $SD = 7.2$ versus $M = -1.3\%$, $SD = 9.3$ in the HILS and LIHS conditions, respectively), $t(182.5) = -0.36$, and $P = 0.72$. Thus, the increased effectiveness of the HILS over the LIHS enforcement cannot be attributed to feedback improving perceptual accuracy.

Testing the secondary hypothesis. A regression analysis indicated that the difference between the two enforcement conditions was more pronounced when the accuracy–violation rates in the NE game were higher (Fig. 3*B*). The interaction with condition was significant: $F(1,188) = 6.49$, $P = 0.012$, $b = 0.22$, 95% CI (0.05, 0.40), and partial $f^2 = 0.03$. Specifically, the simple slopes of the accuracy–violation rate with enforcement on the accuracy–violation without enforcement were 0.33, 95% CI (0.21, 0.45) for HILS and 0.55, 95% CI (0.43, 0.68) for LIHS.

Study 3

The results of the lab pilot study, Study 1 and Study 2 suggest that the combination of frequent inspections with small fines is more effective in reducing violation behaviors than rare but severe fines. This is in line with the decisions-from-experience hypothesis, which predicts that rare events (such as rare, severe punishments) are underweighted, whereas frequent events (such as frequent mild punishments) are not. Theoretically, one factor that can impose boundary conditions on the effect is the probability of the rare event, since the degree of underweighting is predicted to be negatively related to the probability of the event (20). Accordingly, making inspection even more rare (by further decreasing the probability of inspection in LIHS) should lead to a greater effect and vice versa. Importantly, in previous studies, underweighting has often been observed for events with probabilities below 0.2, whereas more frequent events ($P > 0.2$) were not typically underweighted (20, 36, 37). Hence, increasing the probability of inspection above 0.2 could be enough to serve as a boundary condition and eliminate the difference between the two enforcement policies.

To examine these predictions, we conducted two additional studies. In Study 3a, we used more extreme probabilities of inspection (0.06 versus 0.94) and expected to replicate the observed effect from our previous studies but with an even stronger advantage of prioritizing frequency (HILS) over severity (LIHS) in reducing accuracy violations. In Study 3b, we aimed to examine the predicted boundary condition of the effect. We therefore used moderate probabilities of inspection (0.33 versus 0.66) and tested the prediction that in the absence of rare events, the advantage of prioritizing frequency (HILS) overprioritizing severity (LIHS) would be reduced. In both studies, we followed the methods of Study 2 and also kept the same EV for committing accuracy violations.

Study 3a

Methods.

Participants. A total of 199 Prolific users who did not participate in the previous studies were paid £1.45 for participating and were given a bonus contingent on their cumulative points earned (M = £1 and $SD = 0.31$). Three participants were excluded based on the exclusion criterion of 35% "detrimental

Downloaded at Palestinian Territory, occupied on November 30, 2021
www.manaraa.com

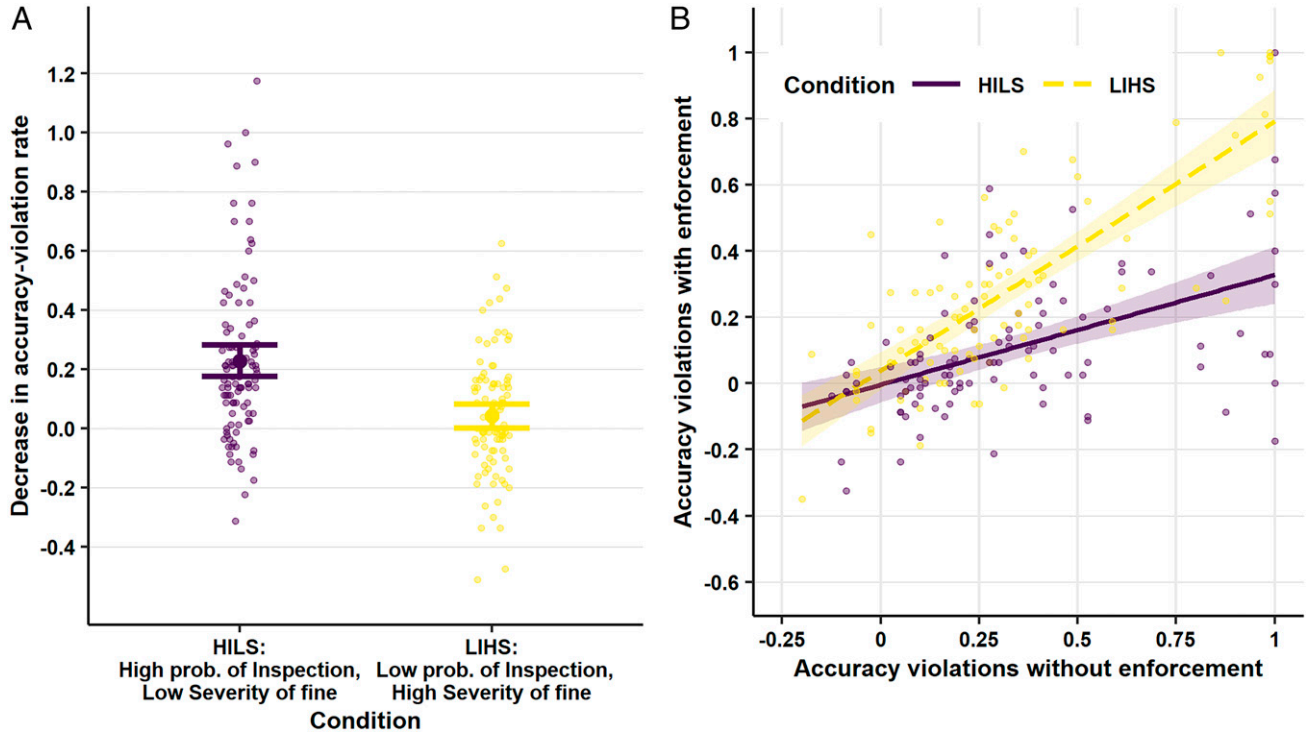

Fig. 4. Results of Study 3a. Small dots depict individual participants. (*A*) Decrease in accuracy–violation rates from baseline to the enforcement block by condition. Error bars represent the 95% bootstrapped CI for the mean. (*B*) Regression lines of the accuracy–violation rates in the enforcement blocks on the accuracy–violation rates in the NE block by condition. Shading around the lines shows the 95% CIs.

errors" or more, resulting in 102 versus 94 participants in the HILS and LIHS conditions, respectively.

Task and procedure. We used the same task, design, and procedure as in Study 2 (including the a priori description of the magnitude of the fine) but with more extreme probabilities and fines. The HILS enforcement game employed a 0.94 probability of inspection and a fine of 10.5 points, and the LIHS enforcement game employed a 0.06 probability of inspection and a fine of 165 points.[#] In addition, we added a block of three true/false questions before the onset of each game to verify that participants understand the instructions (*SI Appendix*, *Part 3*). After each true/false response, feedback was provided, and the relevant part of the instructions was further highlighted.

Results. The findings replicated the results of Study 1 and 2 with larger effects. Fig. 4*A* presents the main result. In the HILS condition, the mean accuracy–violation rate decreased by 22.7% (SD = 27.4; from 33.4% in the NE game to 10.6% in the with-enforcement game). In the LIHS condition, the accuracy–violation rates decreased by 4.2% (SD = 20.4; from 32.2% in the NE game to 28.0% in the with-enforcement game). The Shapiro–Wilk test rejected the assumption of normality for the accuracy–violation rates ($P < 0.001$). A Mann–Whitney U test indicated that the reduction of accuracy–violation rates in the HILS condition (median = 17.5% and range [−31, 118]) was greater than in the LIHS condition (median = 0.6% and range [−51, 63]), U = 6,774, $P > 0.001$, one-tailed test, and effect size $r = 0.356$.

Again, we found no evidence that the frequent inspections caused participants in the HILS condition to make fewer detrimental errors than in the LIHS condition. In fact, participants in the HILS condition made more such mistakes than participants in the LIHS condition (M = 1.8% and SD = 6.3 versus $M = -3.1\%$ and SD = 8.9 in the LIHS and HILS conditions, respectively), $t(183.1) = -4.42$, and $P < 0.001$, indicating once again that the increased effectiveness of the HILS over the LIHS enforcement cannot be attributed to feedback improving perceptual accuracy.

As in the previous studies, here again, a regression analysis indicated that the difference between the two enforcement conditions was more pronounced when the accuracy–violation rates in the NE game were higher (Fig. 4*B*). The interaction with condition was significant: $F(1,192) = 22.85$, $P < 0.001$, $b = 0.42$, 95% CI (0.25, 0.60), and partial $f^2 = 0.12$. Specifically, the simple slopes of the accuracy–violation rate with enforcement on the accuracy–violation without enforcement were 0.33, 95% CI (0.21, 0.45) for HILS and 0.75, 95% CI (0.63, 0.88) for LIHS.

Study 3b

Methods.

Participants. A total of 193 Prolific users who did not participate in the previous studies were paid £1.45 for participating and were given a bonus contingent on their cumulative points earned (M = £1.03 and SD = 0.30). Three participants were excluded based on the exclusion criterion of 35% "detrimental errors" or more, resulting in 94 versus 96 participants in the HILS and LIHS conditions, respectively.

Task and procedure. The task, design, and procedure were identical to Study 3a but with moderate instead of extreme probabilities and fines. The HILS enforcement game included a 0.66 probability of inspection and a fine of 15 points and the LIHS enforcement game included a 0.33 probability of inspection and a fine of 30 points (EV = 0.1 in both).

[#]The probabilities were chosen under the constraints of p~0.05/0.95, relatively round fine sizes, and EV~0.1. Note that the final parameters suggest that the EV from committing accuracy violations was slightly higher in the HILS condition (0.13 versus 0.1), which runs counter to the predicted effect (our prediction was that HILS would reduce accuracy–violation rates more, despite of the slight increase in EV to violate).

SOCIAL SCIENCES

Downloaded at Palestinian Territory, occupied on November 30, 2021

www.manaraa.com

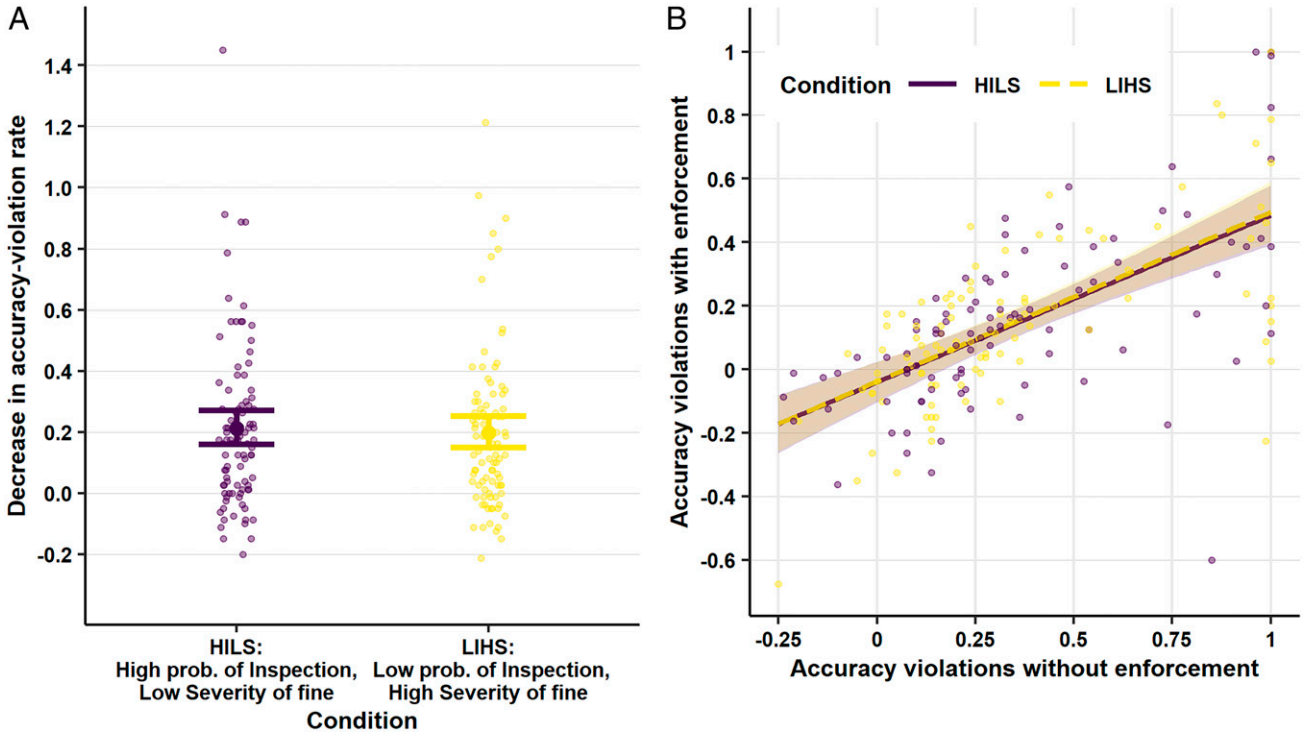

Fig. 5. Results of Study 3b. Small dots depict individual participants. (*A*) Decrease in accuracy–violation rates from baseline to the enforcement block by condition. Error bars represent the 95% bootstrapped CI for the mean. (*B*) Regression lines of the accuracy–violation rates in the enforcement blocks on the accuracy–violation rates in the NE block by condition. Shading around the lines shows the 95% CIs.

Results. As expected, and in contrast to the previous studies, the results of this boundary condition study revealed that the HILS and the LIHS enforcement policies exhibited similar effectiveness in reducing accuracy violations. Fig. 5*A* presents the main result. In the HILS condition, the mean accuracy–violation rate decreased by 21.2% (SD = 27.1; from 36.1% in the NE game to 14.9% in the with-enforcement game). In the LIHS condition, the accuracy–violation rates decreased by 19.7% (SD = 25.5; from 34.0% in the NE game to 14.3% in the with-enforcement game). The Shapiro–Wilk test rejected the assumption of normality for the accuracy–violation rates ($P < 0.001$). A Mann–Whitney U test found no difference between the accuracy–violation rates in the HILS condition (median = 17.5% and range [−20, 145]) and in the LIHS condition (median = 15.6% and range [−21, 121]), U = 4620, $P = 0.77$, two-tailed test (since, unlike in the previous studies, we did not a priori expect a difference between the conditions), and with an effect size of $r = 0.021$. We also did not find any difference between the two enforcement conditions, as they pertained to different types of individuals (Fig. 5*B*).

Discussion

What is the best way to reduce violations? The high cost of close monitoring and negative consequences of severe sanctions (e.g., more officers and negative mass reactance) render full severe enforcement impractical. Instead, two common enforcement policies trade off the probability of inspection with the severity of punishment. Our findings suggest that in repeated settings, frequent inspection with mild punishment is more effective in reducing violation behaviors than rare severe punishment. In all four studies, which included frequent versus rare inspections, the HILS policy was more effective at reducing violation rates than the LIHS policy, although these two polices had an equal EV. This finding was independent of potential improvement in perceptual skill and held even when we increased deterrence by providing information about the magnitude of the fine in advance. Furthermore, the advantage of the HILS policy was more pronounced among participants who tended to commit more violations at baseline. The results of the last study demonstrate the effects of changing the probability of inspection on enforcement effectiveness. Study 3a showed that the advantage of HILS over LIHS increased for more extreme probabilities of inspection (0.06 versus 0.94). Study 3b showed a boundary condition for the greater effectiveness of the probability of inspection over the severity of punishment in reducing violations. When the enforcement policy presented moderate probabilities (0.33 versus 0.66) and penalties, there was no significant difference between a policy with higher probability of inspection and smaller fines and a policy with lower probability of inspection and larger fines. Importantly, LIHS was not found to be more effective than HILS in any of the studies.

These findings are consistent with the decisions-from-experience hypothesis and specifically with the phenomenon of underweighting of rare events (23, 38, 39). Consistent with this phenomenon, in repeated settings, rare severe punishments are underweighted, causing deterrence to lose its initial effect over time. Underweighting of rare events is often argued to be the outcome of a decision mechanism that implies reliance on small samples of past experiences (39, 40). According to this mechanism, on each trial, decision makers only recall a small sample of previous experiences with each of the options and choose the option with the higher sample mean. The rarer an experience is, the less likely it is to be included in the recalled sample, and therefore the more likely it is to be given less weight than it deserves. This explains why, in the current context, rare severe punishments were less deterring than frequent but smaller punishments. The reliance on a small samples mechanism also helps to explain the boundary conditions found in Study 3b, in which the probabilities of inspection were moderate (0.33

Downloaded at Palestinian Territory, occupied on November 30, 2021

www.manaraa.com

versus 0.66). Here, the recalled sample is less likely to underrepresent inspection experiences, thus substantially reducing the advantage of HILS over LIHS.

Interestingly, our findings are also consistent with empirical studies showing that the severity of criminal sanctions is not correlated with the level of crime in society (41), whereas enforcement prevalence has consistently been found to be related to crime rates (42). Our controlled experimental setting demonstrates a causal link between enforcement prevalence and violation rates for a general (noncriminal) population and for small-scale violations and punishments, therefore supporting and extending the recommendation often made in the crime literature to increase punishment prevalence rather than severity (43–45).

Our results run counter to the economic hypothesis, which posits that severe punishments are more effective than high monitoring in reducing violations in the general population (1–3, 5). One plausible reason for this discrepancy is that experimental studies supporting the economic prediction tend to employ an explicit descriptive setting with very few opportunities for violation. The current studies employed an arguably more ecologically valid setting, in which implicit opportunities to violate were repeated many times and the probability of inspection was not described a priori. Hence, the difference could be the result of a description-experience gap in violation decisions, similar to the gap found in risky choice (46).

In line with self-concept maintenance theory, under no external enforcement, most of our participants did not maximize profits through violations, supporting the suggested role of internal cost as a gatekeeper of morality. Note that self-concept maintenance theory does not offer a clear prediction regarding external enforcement. Our findings demonstrate the sensitivity to external enforcement and pave the way for further research in this direction. In real life, small violations tend to be recurring. Incorporating the effect of feedback, inspection, and fine severity in the research of behavioral ethics may provide insights and shed light on how enforcement can shape the moral self and the internal barrier over time. To give but one example, applying external enforcement for a certain violation may either trigger an internal signal and direct the moral self to realize that this violation is "morally wrong" (47), or it may signal that this is something "you can buy your way out of," as in "a fine is a price" (32). Future research is thus needed to explore this and other long-term dynamics of moral considerations in the presence of enforcement.

From a practical standpoint, our findings suggest that when the inspection rate is low, policymakers should prioritize increasing the frequency of inspections over the severity of punishments. Since, in the real world, inspection rates are commonly very low,[||] our findings suggest that increasing inspection rates even by as little as a few percentage points could be highly effective in reducing violations. For example, in our studies, an inspection rate of 6% reduced violation rates by 12% (from 32 to 28% in Study 3a), while an inspection rate of 10% with smaller fines reduced violation rates by more than 38% (from 36 to 22% in Study 2). Although increasing the magnitude of fines was often considered as less costly (1), recent advances in technology and the increasing usage of artificial intelligence algorithms enable more effective monitoring at significantly lower costs (49–51). Moreover, large fines could result in a perception of unfairness and consequently reduce the probability of detection (52, 53), which, according to our results, is the key factor. In a similar vein, empirical crime researchers have argued that programs focusing on increasing punishment severity entail greater costs in their implementation (43–45). Thus, while many regulators in the current COVID-19 pandemic have publicly called to increase the magnitude of fines, our findings strongly suggest that "gentle rule enforcement," which includes smaller punishments with a higher probability (54–57), would be more effective in reducing violation rates, especially for high offenders, the target population of any enforcement policy.

Data Availability. Raw data have been deposited in the Open Science Framework (https://osf.io/7f26g/). All other study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. This research was supported by the Israel Science Foundation (Grant No. 2740/20).

[||]For example, according to the 2020 Internal Revenue Service (IRS) Data Book (Table 17) "for Tax Years 2010 through 2018, the IRS has examined 0.63 percent of individual returns filed."

1. G. S. Becker, Crime and punishment: An economic approach. *J. Polit. Econ.* **76**, 169–217 (1968).
2. L. R. Anderson, S. L. Stafford, Punishment in a regulatory setting: Experimental evidence from the VCM. *J. Regul. Econ.* **24**, 91–110 (2003).
3. M. K. Block, V. E. Gerety, Some experimental evidence on differences between student and prisoner reactions to monetary penalties and risk. *J. Legal Stud.* **24**, 123–138 (1995).
4. C. Engel, D. Nagin, Who is afraid of the stick? Experimentally testing the deterrent effect of sanction certainty. *Rev. Behav. Econ.* **2**, 405–434 (2015).
5. L. Friesen, Certainty of punishment versus severity of punishment: An experimental investigation. *South. Econ. J.* **79**, 399–421 (2012).
6. A. Tversky, D. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertain.* **5**, 297–323 (1992).
7. S. Shalvi, F. Gino, R. Barkan, S. Ayal, Self-serving justifications: Doing wrong and feeling moral. *Curr. Dir. Psychol. Sci.* **24**, 125–130 (2015).
8. J. Abeler, A. Becker, A. Falk, Representative evidence on lying costs. *J. Public Econ.* **113**, 96–104 (2014).
9. N. Mazar, O. Amir, D. Ariely, The dishonesty of honest people: A theory of self-concept maintenance. *J. Mark. Res.* **45**, 633–644 (2008).
10. S. Ayal, F. Gino, "Honest rationales for dishonest behavior" in *The Social Psychology of Morality: Exploring the Causes of Good and Evil*, M. Mikulincer, P. R. Shaver, Eds. (American Psychological Association, Washington, DC, 2011), pp. 149–166.
11. R. Barkan, S. Ayal, D. Ariely, Ethical dissonance, justifications, and moral behavior. *Curr. Opin. Psychol.* **6**, 157–161 (2015).
12. F. Gino, S. Ayal, D. Ariely, Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychol. Sci.* **20**, 393–398 (2009).
13. F. Gino, S. Ayal, D. Ariely, Self-serving altruism? The lure of unethical actions that benefit others. *J. Econ. Behav. Organ.* **93**, 285–292 (2013).
14. U. Gneezy, B. Rockenbach, M. Serra-Garcia, Measuring lying aversion. *J. Econ. Behav. Organ.* **93**, 293–300 (2013).
15. P. Gerlach, K. Teodorescu, R. Hertwig, The truth about lies: A meta-analysis on dishonest behavior. *Psychol. Bull.* **145**, 1–44 (2019).
16. S. Ayal, F. Gino, R. Barkan, D. Ariely, Three principles to REVISE people's unethical behavior. *Perspect. Psychol. Sci.* **10**, 738–741 (2015).
17. C. Schild, D. W. Heck, K. A. Ścigała, I. Zettler, Revisiting REVISE: (Re)Testing unique and combined effects of REminding, VIsibility, and SElf-engagement manipulations on cheating behavior. *J. Econ. Psychol.* **75**, 102161 (2019).
18. D. Ariely, S. Jones, *The (Honest) Truth About Dishonesty* (Harper Collins Publishers, New York, 2012).
19. M. H. Bazerman, A. E. Tenbrunsel, *Blind Spots* (Princeton University Press, 2011).
20. K. Teodorescu, I. Erev, On the decision to explore new alternatives: The coexistence of under-and over-exploration. *J. Behav. Decis. Making* **27**, 109–123 (2014).
21. K. Teodorescu, I. Erev, Learned helplessness and learned prevalence: Exploring the causal relations among perceived controllability, reward prevalence, and exploration. *Psychol. Sci.* **25**, 1861–1869 (2014).
22. R. Barkan, D. Zohar, I. Erev, Accidents and decision making under uncertainty: A comparison of four models. *Organ. Behav. Hum. Decis. Process.* **74**, 118–144 (1998).
23. R. Hertwig, G. Barron, E. U. Weber, I. Erev, Decisions from experience and the effect of rare events in risky choice. *Psychol. Sci.* **15**, 534–539 (2004).
24. A. R. Camilleri, B. R. Newell, When and why rare events are underweighted: A direct comparison of the sampling, partial feedback, full feedback and description choice paradigms. *Psychon. Bull. Rev.* **18**, 377–384 (2011).
25. K. Teoderescu, M. Amir, I. Erev, The experience-description gap and the role of the inter decision interval. *Prog. Brain Res.* **202**, 99–115 (2013).
26. O. Plonsky, K. Teodorescu, The influence of biased exposure to forgone outcomes. *J. Behav. Decis. Making* **33**, 393–407 (2020).
27. F. Gino, M. I. Norton, D. Ariely, The counterfeit self: The deceptive costs of faking it. *Psychol. Sci.* **21**, 712–720 (2010).

Downloaded at Palestinian Territory, occupied on November 30, 2021
www.manaraa.com

28. G. Hochman, A. Glöckner, S. Fiedler, S. Ayal, "I can see it in your eyes": Biased processing and increased arousal in dishonest responses. *J. Behav. Decis. Making* **29**, 322–335 (2016).
29. A. Harel, U. Segal, Criminal law and behavioral law and economics: Observations on the neglected role of uncertainty in deterring crime. *Am. Law Econ. Rev.* **1**, 276–312 (1999).
30. B. E. Hilbig, I. Zettler, When the cat's away, some mice will play: A basic trait account of dishonest behavior. *J. Res. Pers.* **57**, 72–88 (2015).
31. Y. Feldman, *The Law of Good People: Challenging States' Ability to Regulate Human Behavior* (Cambridge University Press, 2018).
32. U. Gneezy, A. Rustichini, A fine is a price. *J. Legal Stud.* **29**, 1–17 (2000).
33. R. K. Jessup, A. J. Bishara, J. R. Busemeyer, Feedback produces divergence from prospect theory in descriptive choice. *Psychol. Sci.* **19**, 1015–1022 (2008).
34. D. Marchiori, S. Di Guida, I. Erev, Noisy retrieval models of over-and undersensitivity to rare events. *Decision (Wash. D.C.)* **2**, 82–106 (2015).
35. I. Erev, E. Ert, O. Plonsky, D. Cohen, O. Cohen, From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. *Psychol. Rev.* **124**, 369–409 (2017).
36. O. Plonsky, K. Teodorescu, Perceived patterns in decisions from experience and their influence on choice variability and policy diversification: A response to Ashby, Konstantinidis, & Yechiam, 2017. *Acta Psychol. (Amst.)* **202**, 102953 (2020).
37. A. Luria, I. Erev, E. Haruvy, The reinforcing value of lottery tickets, and the synergetic effect of distinct reinforcements. *J. Behav. Decis. Making* **30**, 533–540 (2017).
38. B. R. Newell, T. Rakow, The role of experience in decisions from description. *Psychon. Bull. Rev.* **14**, 1133–1139 (2007).
39. I. Erev, A. E. Roth, Maximization, learning, and economic behavior. *Proc. Natl. Acad. Sci. U.S.A.* **111** (suppl. 3), 10818–10825 (2014).
40. O. Plonsky, K. Teodorescu, I. Erev, Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychol. Rev.* **122**, 621–647 (2015).
41. A. N. Doob, C. M. Webster, Sentence severity and crime: Accepting the null hypothesis. *Crime Justice* **30**, 143–195 (2003).
42. A. Chalfin, J. McCrary, Criminal deterrence: A review of the literature. *J. Econ. Lit.* **55**, 5–48 (2017).
43. J. L. Nichols, H. L. Ross, The effectiveness of legal sanctions in dealing with drinking drivers. *Alcohol Drugs Driving* **6**, 33–60 (1990).
44. A. Hawken, M. Kleiman, *Managing Drug Involved Probationers With Swift and Certain Sanctions: Evaluating Hawaii's HOPE: Executive Summary* (National Criminal Justice Reference Services, Washington, DC, 2009).
45. M. Tonry, "An honest politician's guide to deterrence: Certainty, severity, celerity, and parsimony" in *Deterrence, Choice, and Crime*, D. S. Nagin, F. T. Cullen, C. L. Jonson, Eds. (Routledge, New York, 2018), pp. 365–391.
46. R. Hertwig, I. Erev, The description-experience gap in risky choice. *Trends Cogn. Sci.* **13**, 517–523 (2009).
47. M. Bateson, D. Nettle, G. Roberts, Cues of being watched enhance cooperation in a real-world setting. *Biol. Lett.* **2**, 412–414 (2006).
48. Internal Revenue Service, Internal Revenue Service Data Book, 2020: Publication 55-B. https://www.irs.gov/pub/irs-pdf/p55b.pdf. Accessed 5 October 2021.
49. W. F. Abaya, J. Basa, M. Sy, A. C. Abad, E. P. Dadios, "Low cost smart security camera with night vision capability using Raspberry Pi and OpenCV" in *2014 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)* (IEEE, 2014), pp. 1–6.
50. E. L. Piza, B. C. Welsh, D. P. Farrington, A. L. Thomas, CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis. *Criminol. Public Policy* **18**, 135–159 (2019).
51. S. Raaijmakers, Artificial intelligence for law enforcement: Challenges and opportunities. *IEEE Secur. Priv.* **17**, 74–77 (2019).
52. E. Feess, H. Schildberg-Hörisch, M. Schramm, A. Wohlschlegel, The impact of fine size and uncertainty on punishment and deterrence: Theory and evidence from the laboratory. *J. Econ. Behav. Organ.* **149**, 58–73 (2018).
53. A. M. Polinsky, S. Shavell, The fairness of sanctions: Some implications for optimal enforcement policy. *Am. Law Econ. Rev.* **2**, 223–237 (2000).
54. I. Erev, P. Ingram, O. Raz, D. Shany, Continuous punishment and the potential of gentle rule enforcement. *Behav. Processes* **84**, 366–371 (2010).
55. A. Schurr, D. Rodensky, I. Erev, The effect of unpleasant experiences on evaluation and behavior. *J. Econ. Behav. Organ.* **106**, 1–9 (2014).
56. I. Erev, O. Plonsky, Y. Roth, Complacency, panic, and the value of gentle rule enforcement in addressing pandemics. *Nat. Hum. Behav.* **4**, 1095–1097 (2020).
57. Y. Roth, O. Plonsky, E. Shalev, I. Erev, On the value of alert systems and gentle rule enforcement in addressing pandemics. *Front. Psychol.* **11**, 577743 (2020).

Downloaded at Palestinian Territory, occupied on November 30, 2021

www.manaraa.com