Journal of Behavioral and Experimental Economics 97 (2022) 101813

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Experimental test of the effects of punishment probability and size on the decision to take a bribe

Štěpán Bahník, Marek A. Vranka *

Faculty of Business Administration, Prague University of Economics and Business, náměstí Winstona Churchilla 4, Prague 130 67, Czech Republic

ARTICLE INFO

JEL:
C91
D73
D91
K42

Keywords:
Corruption
Bribe-taking
Punishment
Laboratory experiment
HEXACO

ABSTRACT

Punishment is one of the main methods for preventing corruption. However, studies on the effect of size and probability of punishment on bribe-taking have not yielded conclusive results, possibly because studies often abstract from internal costs of wrongdoing. We introduce a punishment by a fine or termination of the task, both with varying probabilities, in a laboratory task modeling the decision to take a bribe. The punishment decreased the probability of taking higher bribes, even though the probability of taking lower bribes was unaffected. Participants took fewer bribes when the fine was larger and more probable. We did not observe any clear negative effects of small punishment crowding out intrinsic motivation to behave honestly. However, we found that the effects of punishment differ based on emotionality and honesty-humility of participants. The study shows that the prospect of punishment may deter dishonest behavior; however, personality characteristics should be taken into account when devising an effective deterrence policy.

1. Introduction

Most attempts at curbing corruption focus on increasing the probability and size of a punishment (Abbink & Serra, 2012). Despite the amount of research interest in the effects of these two factors, the results remain mixed and inconclusive (Boly & Gillanders, 2018). For example, findings in crime studies suggest that the detection probability plays a greater role in deterring criminal behavior (Nagin, 2013), while laboratory studies in general support the notion that the size of punishment has a stronger effect (Laske, Saccardo, & Gneezy, 2018). One possible reason is that laboratory studies often explore effects of punishment in settings in which punishable behavior does not clearly violate any internalized social or moral norms (Friesen, 2012). In the usually employed experimental tasks, either nobody is harmed or the harm done to other participants can be seen as an intended part of playing the experimental "game" (e.g., Abbink, Irlenbusch, & Renner, 2002; Alatas, Cameron, Chaudhuri, Erkal, & Gangadharan, 2009; Drugov, Hamman, & Serra, 2014). The tasks often present participants with a tradeoff between reward for themselves and others rather than a clear norm of conduct that should be followed. However, corruption is associated with breaking of norms in the real world and evidence from psychological studies demonstrate that internal costs of breaking norms affect people's behavior. Therefore, findings regarding punishment from studies in which norm-breaking is not a part of the experimental task may not readily apply to corrupt behavior.

In our study, we aimed to explore the effects of punishment using a laboratory task that does not suffer from these shortcomings. We used a task modeling the decision to take a bribe introduced by Vranka and Bahník (2018). Participants were asked to sort objects according to a given rule with a small fixed monetary reward for each sorted object. Randomly selected objects were associated with a much higher additional reward—simulating a "bribe"—that participants could get for breaking the sorting rule, which also caused monetary harm to a third party.[1] To the task used by Vranka and Bahník (2018) we added the possibility of punishment after taking a bribe. Unlike the previous studies (e.g., Abbink et al., 2002; Banerjee & Mitra, 2018; Schildberg-Hörisch & Strassmair, 2010; Schulze & Frank, 2003), we employed multiple forms and sizes of punishment with independently varied probabilities, allowing us to test for their main effects as well as possible interactions. We employed two sizes of monetary punishment and a punishment which leads to an end of the task, and thus of the possibility to earn additional reward. We also manipulated the probability of

* Corresponding author.
E-mail address: vranka.marek@gmail.com (M.A. Vranka).

[1] The task thus focused only on the person taking the bribe and not on the person offering the bribe. This simplification somewhat limits the applicability of the results to some real-world situations, but allows for better control of the bribe offers.

https://doi.org/10.1016/j.socec.2021.101813
Received 8 February 2021; Received in revised form 30 November 2021; Accepted 7 December 2021
Available online 12 December 2021

punishment after taking a bribe.

We were thus able to disentangle the effects of probability and size of punishment, which was not previously done in the context of bribe taking. In addition, corrupt behavior in the experiment was associated with breaking rules and causing harm to a party not involved in the task itself. We explored how punishment affects perceived morality of the corrupt behavior and thus indirectly assessed punishment's effect on the internal costs of wrongdoing. By measuring personality characteristics, we also explored interindividual differences in the effects of punishment. Finally, thanks to the repeated nature of the opportunity to take bribes, we also explored the effects of administered fines on subsequent bribe taking.

The rest of the paper is organized as follows. Section 2 presents an overview of existing studies concerning effects of punishment on dishonest and corrupt behavior. In Section 3, we describe hypotheses of the study. Section 4 presents our sample and experimental design, followed by presentation of results in Section 5. Lastly, we discuss the results and conclude in Section 6 and 7.

2. Literature review

Because most of the forms of corruption are illegal or at least socially unacceptable, people try to hide their corrupt activities. The clandestine nature of corruption makes its study challenging as observing corrupt behavior is difficult (Sequeira, 2012). Corrupt behavior has been therefore usually studied indirectly and on a country or cross-country level. Public perception of the prevalence of corruption or conviction rates are, for example, used as common measures of corruption (Goel & Nelson, 2011; Sequeira, 2012). Studies using these measures may indicate factors influencing corrupt behavior and its consequences, but because of endogeneity issues, they are ill-suited for the study of processes that lead people to act corruptly or for designing anti-corruption policies (Abbink & Serra, 2012). Promising to overcome these limitations, experimental tasks modeling the decision behind taking or offering a bribe in a laboratory setting have gained popularity in recent years (Serra & Wantchekon, 2012). Results of these laboratory studies largely corroborate findings of studies of corruption conducted in real settings, which supports their external validity (Armantier & Boly, 2012). Even though they cannot capture the whole scope of factors that play a role in corrupt behavior in the real world, laboratory experiments offer a unique possibility to manipulate specific features of the decision process and factors influencing it and allow causal inferences which are necessary to create effective interventions aimed at reducing corruption (da Hora & Sampaio, 2019).

The most straightforward tool against corruption is punishment. From a purely economic perspective, people's behavior is motivated by their rational self-interest. They should therefore cheat and act corruptly whenever the expected gains from such actions exceed the expected punishment (Becker, 1968). In this framework, the expected punishment is determined by the monetary value of a penalty[2] and the probability of being caught and punished. However, a large number of psychological studies show that people behave much less dishonestly than would be predicted by the *homo economicus* perspective (Mazar & Ariely, 2006; Van Winden & Ash, 2009). A possible explanation of the discrepancy lies in the existence of internalized moral norms that limit people's self-interest by imposing additional internal costs of wrongdoing: because people anticipate they would feel discomfort after behaving immorally, they forgo opportunities to cheat even when the probability of getting caught and/or the size of external punishment are relatively small (Mazar, Amir, & Ariely, 2008). From this extended perspective, not only the monetary value of punishment is important, but also non-monetary penalties, such as social condemnation or loss of positive moral self-image, affect the decision to act corruptly (Salmon & Serra, 2017).

Regardless of its specific nature, with increasing probability and size of punishment, the corrupt behavior is supposed to decline. While the general deterrence effect of punishment seems to be well established with observational (Goel & Rich, 1989) as well as with experimental data (Boly & Gillanders, 2018; Hanna, Bishop, Nadel, Scheffler, & Durlacher, 2011; Nagin, 2013), the interplay between the effects of increasing its severity and probability is less clear. Although Nagin (2013)) argues that the probability of punishment[3] and not the severity of punishment serves as a deterrent, findings from laboratory studies in general suggest that the severity of punishment has the stronger deterrent effect (Friesen, 2012; Laske et al., 2018). One reason for the stronger effect of severity of punishment is that it is easier to evaluate fines than absolute probabilities. Laske et al. (2018) showed that when participants faced a single decision, they were completely unaffected by the detection probability and its effect emerged only when different probabilities could be compared. Alternatively, people might simply imagine how they would feel in the worst-case scenario, regardless of its probability, and when the possible loss looms large they avoid even a small risk (Qin & Wang, 2013). This explanation is supported by Banerjee and Mitra (2018) who experimentally demonstrated that despite the same expected value of punishment, high fines with a low probability decreased the corrupt behavior, unlike lower fines with higher probability. In a similar vein, it was shown that even a small probability of severe punishment in the form of a *sudden death* (i.e., after being discovered, the players would lose all of their earnings from the experiment above the show-up fee) led to a significant reduction in offering and accepting bribes (Abbink et al., 2002). The corruption behavior was reduced despite the fact that participants still tended to *underestimate* the overall probability of punishment.

However, the interplay between punishment and internal costs of wrongdoing is much less explored in the current literature. In some circumstances, introducing a small penalty for undesirable behavior may paradoxically lead to an increase of the penalized behavior (Gneezy & Rustichini, 2000; Khadjavi, 2014; Schildberg-Hörisch & Strassmair, 2010). For example, Schulze and Frank (2003) found that a relatively small risk of punishment increased the overall number of participants who accepted bribes in the condition with punishment in comparison to the condition without punishment. Still, when the risk of punishment increased with the size of a taken bribe, the punishment worked as a deterrent and the proportion of participants taking the highest bribes was lower in the punishment condition than in the control condition. The reversed effect of small punishment is supposedly caused by "crowding out" of the internalized moral norms by the monetary punishment, which, however, is itself insufficient to deter dishonest behavior (Frey & Jegen, 2001). Although the crowding out effect of small punishment has been observed in a number of economic experiments, apart from the above-mentioned study by Schulze and Frank (2003), no other study demonstrated the effect in a task modeling corruption (Bowles & Polanía-Reyes, 2012).

Moreover, there remain many unanswered questions regarding external as well as internal factors possibly modifying effects of punishment. For example, Boly, Gillanders, and Miettinen (2019) showed that legitimacy of punishments moderates their effect and Hilbig, Zettler, and Heydasch (2012) demonstrated that punishments affect only participants with a low honesty-humility personality trait. It is therefore possible that the deterrent effect of punishment might differ among participants with different internalized norms (Zettler & Hilbig, 2010) and the crowding out effect might occur only for a subset of participants (Schildberg-Hörisch & Strassmair, 2010). In addition, no previous study on corruption has explored the effect of administration of bribe on

[2] For example, the value of a monetary fine, disutility from lost income after one is fired, and/or disutility from a prison or community service sentence after one is sentenced.

[3] Specifically, the probability of apprehension.

subsequent bribe taking and the available studies on effects of being fined present mixed results (Dušek & Traxler, 2017; Lawpoolsri, Li, & Braver, 2007).

In the current study, we contribute to the above reviewed literature by exploring how probability and size of punishment, manipulated between subjects, affect self-reported perception of bribe taking, and thus internal costs of wrongdoing, how bribe taking itself is affected, and whether the effects of punishment are moderated by personality characteristics covered by the HEXACO model (Ashton & Lee, 2009; Ashton, Lee, & de Vries, 2014).

Although other experimental studies tested effects of punishment and its probability (Friesen, 2012; Laske et al., 2018), they do so in settings without any clearly stated norms of conduct. Because the usually used tasks do not contain violation of any explicit rules or norms and often use abstract and neutral language, it is unclear whether participants view the behavior in question as corruption rather than a simple tradeoff between their and others' rewards. To overcome this issue, we use a task developed by Vranka and Bahník (2018). The task models the decision to disregard one's duties to enrich oneself at the expense of others. It thus shares features with corruption in a bureaucratic setting, in which a public official can gain a reward for providing a favor for another person (Jain, 2001). The task may be therefore better suited for the study of the interplay between effects of external punishment and internal costs of wrongdoing.

3. Hypotheses

We extend the existing studies on punishment in general and on punishment for corruption in particular by systematically exploring how different probabilities and forms of punishment for taking a bribe decrease the probability of taking a bribe. Specifically, following the economic analysis of crimes (Becker, 1968), we expect that the likelihood of taking a bribe will decrease with higher punishment size (H1) and punishment probability (H2) and that the effect of punishment probability on the likelihood of taking a bribe will be larger for higher punishment size (H3).[4]

We also study personality characteristics that might influence the decision to take a bribe and interact with the effect of punishment. Based on the related literature (Hilbig & Zettler, 2015) we expect that people higher in honesty-humility will be less likely to take a bribe (H4). However, we also explore possible association between the bribe-taking and the remaining five HEXACO personality dimensions (emotionality, extraversion, agreeableness, conscientiousness, openness to experience).

As the main proposed mechanism for the crowding out of the internal motivation for honesty is that the punishment frames the task differently, thus changing its perception by participants (Bowles & Polanía-Reyes, 2012), our study also assesses the effect of punishment on perceived morality of decisions in the task. Based on the prediction of a crowding-out effect, we expect that participants will perceive accepting and refusing an offered bribe in the task in less moral terms with increasing punishment size (H5) and punishment probability (H6).

4. Methods

The materials used in the study, data, analysis scripts, as well as pre-registration of the study are available at: https://osf.io/szrtd

4.1. Participants

Five hundred fifteen participants were recruited from a laboratory participant pool for participation in a batch of studies conducted in a lab on computers, the first of which was the present study. Three data files were incomplete for an unknown reason, we therefore conducted analysis with the data from the remaining 512 participants. Majority of the participants were students ($n = 383$) of humanities or social sciences ($n = 139$) and economics or management ($n = 111$). Participants were predominantly young ($Mdn_{age} = 23$, $IQR_{age} = 6$) and a majority were female ($n = 333$). The experiment was administered in groups of up to 17. The sample size is sufficient to detect a correlation $r = 0.12$ with power 0.80. A more precise power analysis for other analyses is complicated by the nature of the repeated measures design of the experiment and by complexity of mixed-effect regression, which was used for most of the analysis.

4.2. Procedure

Before beginning the experiment, participants chose one out of four well-known Czech charitable organizations for which they would be able to win money in the subsequent task. The choice was included to ensure that participants value positively the money won for the charity. Afterwards, they were explained the task and they were told that they have a 20% probability[5] that the points they would earn during the task would be converted to a monetary reward using the conversion rate 10 points = 1 CZK (~0.044 USD).

Participants were told to sort objects running on a computer screen according to their color (see Fig. 1). The objects had three possible shapes (triangle, square, and circle), each of which could be one of three possible colors (yellow, blue, and orange). The sorting was done by pressing one of three keys ("1", "2", and "3"), each of which was associated with a single color and shape. For example, in Fig. 1, the participant is shown a yellow square and they can match the object to a yellow circle (by pressing "1"), orange square (by pressing "2"), or blue triangle (by pressing "3"). Colors associated with the three keys were randomly determined for each trial. The task simulated work an employee does as a part of their job.

At the beginning of the task, 2000 points (corresponding to 200 CZK, ~8.7 USD) were allotted to a charity. If participants sorted the object to an incorrect color, the charity lost 200 points (in the trial depicted in Fig. 1, if the participant pressed "2" or 3" instead of "1"). The loss simulated negative societal effects of not performing given work according to the given rule (see van Veldhuizen, 2013, for a similar method). Regardless of whether the object was sorted according to the rule or not, participants got a fixed reward of 3 points for each sorted object, which represented the salary given to a worker for performing their job. Finally, only in trials where the two sorting criteria were mismatched (i.e., in about two thirds of all trials where sorting according to the color was performed by pressing a different key than sorting according to the shape), there was a 22.5% probability that a given object was associated with a "bribe" (i.e., on average in 15% of all trials). These objects were shown with a number corresponding to the value of the bribe, which a participant received if they sorted the object according to its shape. The bribe size was randomly selected from the range 40 to 190 points (in 10-point increments). In the trials with an offered bribe, participants could have therefore disregarded the sorting rule they were instructed to use to earn additional reward for themselves

[4] While our pre-registration (https://osf.io/szrtd/wiki/home/) includes only information about the tested effects, we describe expected directional relationships derived from existing literature in this section. Apart from the pre-registered hypotheses we also conducted additional exploratory analyses. In particular, the analysis related to the perception of taking a bribe, emotionality, interaction of personality characteristics with the presence of punishment, and the effect of administration of fines on subsequent bribe-taking were not mentioned in the pre-registration.

[5] For budgetary reasons, we were not able to pay all the participants. Existing research suggests that participants behave similarly if their reward is paid only with a certain probability and when the pay off is certain (Cubitt, Starmer, & Sugden, 1998; Starmer & Sugden, 1991).

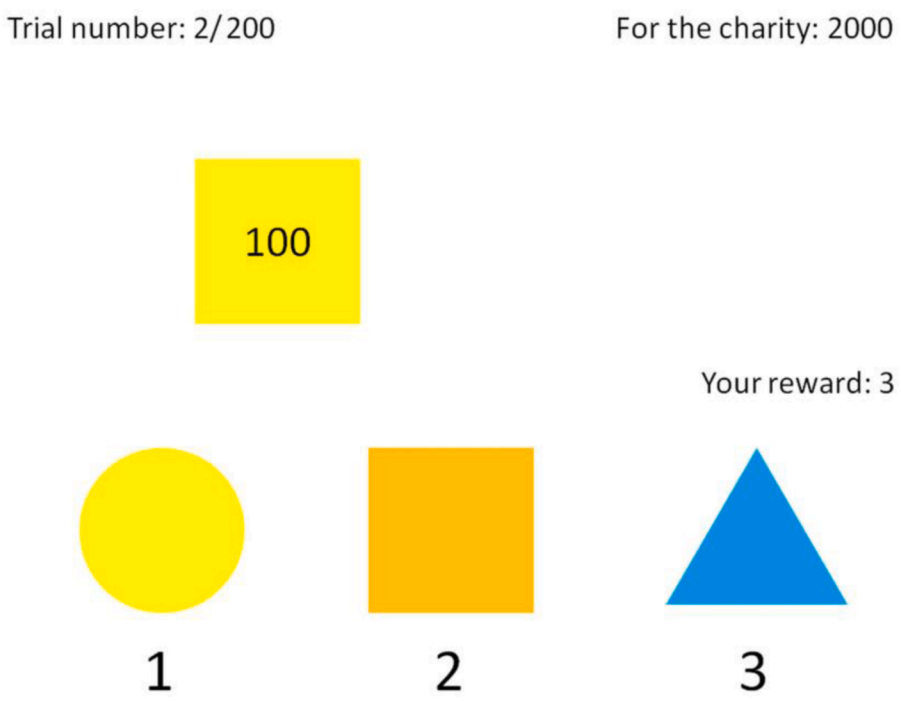

Fig. 1. An illustration of a computer screen seen by a participant.
The top row shows information about the number of the current trial, the total number of trials, and the number of points currently assigned to the charity organization. In the middle of the screen, an object (a yellow square in this case) is moving from the left side of the screen to the right. The current participant's reward in points is shown to the right of the screen. In the bottom row, a participant sees which shapes and colors are assigned to keys "1," "2," and "3" in this trial. If the participant presses "1," the object would be sorted by its color, that is correctly, and the participant would gain 3 points. If the participant presses "2," the object would be matched to a wrong color, and would cause a loss of 200 points for the charity, but it would be sorted according to its shape, gaining the participant the 100 points marked on the object in addition to the 3 points awarded for each sorted object. If the participant presses "3," the object would be matched to a wrong color and shape, the charity would therefore lose 200 points and the participants would gain only the 3 points for sorting the object. Adapted from "Bureaucracy game: A new computer task for the experimental study of corruption," by M. A. Vranka & Š. Bahník, 2018, *Frontiers in Psychology, 9:1511*, p. 3. Copyright 2018 by Vranka and Bahník.

at the expense of the charity (in the trial depicted in Fig. 1, if the participant pressed "2"). In trials without a bribe, sorting objects according to their shape led only to the loss to the charity and no gain for the participant.

Each participant went through 200 trials of the task; i.e., 200 objects to sort. The number of the current trial as well as the money earned for oneself and the charity were displayed on the screen during the whole task (see Fig. 1). Participants were given all the information in advance apart from the probability of the bribes and that the bribes are offered only in trials with mismatched sorting criteria. The shapes and colors of objects as well as bribes were randomly determined for each participant. The features of the objects and bribes were randomized at each trial, independent from other trials and participants' behavior in the task.

To minimize reputational concerns of cheating, participants were given a lottery after the task, in which they were able to win additional money.[6] There were two tasks unrelated to the present experiment after the lottery and then the participants filled in the HEXACO questionnaire (Ashton & Lee, 2009) and demographic information. The HEXACO questionnaire measures six traits: honesty-humility, emotionality, extraversion, agreeableness, conscientiousness, openness to experience. Extraversion, conscientiousness, and openness to experience correspond more or less to their Big Five counterparts. Emotionality is associated with Big Five's neuroticism, but lacks anger-related aspects and instead contains traits related to sentimentality. Conversely, HEXACO's agreeableness lacks the sentimentality-related traits and contains those related to anger.. Honesty-humility is an additional trait, which consists of sincerity, fairness, greed-avoidance, and modesty facets which are loosely related to Big Five's agreeableness (Ashton et al., 2014; Thielmann, Moshagen, Hilbig, & Zettler, 2021).

To examine the perception of the task, we also asked the participants whether they consider breaking the sorting rule to gain additional points as "despicable", "dishonest", "unjust", and "immoral" and whether they consider ignoring the additional points to not lose money for charity as "just", "praiseworthy", "honest", and "moral" on a scale from one to four (1 – certainly not, 2 – rather not, 3 – rather yes, 4 – certainly yes). For analysis, we computed a composite score of task perception for each participant by averaging the eight ratings ($M = 2.58$, $SD = 0.59$; Cronbach's $\alpha = 0.82$, 95% CI = [0.80, 0.84]), and the evaluation of taking a bribe by averaging the negative ratings for taking the bribe ($M = 2.14$, $SD = 0.70$; Cronbach's $\alpha = 0.79$, 95% CI = [0.76, 0.82]).

After the study, participants were debriefed and paid for the study. In addition to the reward which they could win in the task and in the lottery, they were given 145 CZK (~6 USD) participation fee.

4.3. Design

Participants were divided into four groups according to the probability of punishment. A control group without punishment had the probability of 0%, three experimental groups had probabilities of 1%, 5%, and 25% that they would be punished after taking a bribe; i.e., when they sorted an object according to its shape in a trial with a bribe. Each experimental group with non-zero probability of punishment was further divided into three groups which differed in the severity of punishment. For one group, the punishment meant that the task ended and they could not earn any further reward (*end* condition). For the remaining participants, the punishment meant a loss of 40 or 400 points from the reward they had earned (low and high *fine* conditions). Participants were randomly assigned to one of the ten groups (see Table 1 for an overview) in which they stayed throughout the experiment and the experimental groups had the punishment described in the instructions before the task.

5. Results

5.1. Task performance

For trials in which participants classified the object according to one of the criteria, but not the other, the object was classified correctly according to color in 77.3% of trials when there was a bribe and in 97.3%

[6] In the lottery, participants started with 5 CZK and then they had a sequence of choices where they could take the money or decide to take a lottery with a 50% probability to double the money and 50% probability of losing all the money. They were told about the lottery at the beginning of the experiment and they were told that the experimenter handing them their final reward would not know the source of their winnings.

Table 1
Overview of the ten conditions used in the experiment.

Condition	Punishment	Punishment probability	Punishment size
Control	×	–	–
1% 40-fine	✓	1%	40-point fine
5% 40-fine	✓	5%	40-point fine
25% 40-fine	✓	25%	40-point fine
1% 400-fine	✓	1%	400-point fine
5% 400-fine	✓	5%	400-point fine
25% 400-fine	✓	25%	400-point fine
1% end	✓	1%	task termination
5% end	✓	5%	task termination
25% end	✓	25%	task termination

of trials when there was no bribe. Participants rarely made mistakes when there was no bribe and they were motivated to disregard the classification rule in the presence of a bribe.

On average, participants earned 1217 points for themselves (SD = 880, Mdn = 960, IQR = 952). Twenty percent of participants classified all the objects correctly according to color and thus earned 600 points for themselves and did not lose any points for the charity. The distribution of the final outcome for the charity was highly negatively skewed (skewness = −4.11) with a mean of 39 points (SD = 3076) and a median of 1000 points (IQR = 2000). Only a minority of participants had the final outcome for the charity negative (25.8%) or zero (4.5%) and only a small number of participants took more than 90% of all bribes (5.1%).

5.2. Effect of punishment

Trial-level analysis was conducted using mixed-effect linear regression.[7] The incorrectness of object classification, that is taking a bribe, served as a binary dependent variable. The trials incorrectly sorted according to both shape and color as well as trials where the two criteria were aligned were excluded. Bribe size, centered and rescaled to range from −0.5 to 0.5, was included as a covariate. Polynomial (i.e., linear and quadratic) coding was used for the probability of punishment.[8] Random intercepts for participants were included in the model alongside random slopes for bribe size. The random effects account for the dependency of data for a given participant. Order of a trial and squared order of a trial were included as covariates in analyses that did not include the end condition. Both were centered and rescaled to range from −0.5 to 0.5 and random slopes for participants were also included for trial order. The degrees of freedom and p-values are computed with Satterthwaite approximation using R package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017).

Participants were more likely to take higher bribes, b = 0.232, 95% CI [0.200, 0.263], p < .001 (see Fig. 2). While participants in most experimental conditions were less likely to take a bribe than in the control condition, out of the nine experimental conditions, only the 5% 400-fine condition significantly differed from the control condition, b = −0.102, 95% CI [−0.203, −0.001], p = .048 (see Fig. 3).

When an interaction of bribe size with condition was added in the model, the interaction with bribe size was significant for 25% end condition, b = −0.274, 95% CI [−0.427, −0.121], p < .001, and for 25% 400-fine condition, b = −0.147, 95% CI [−0.285, −0.009], p = .037. However, the interaction was negative for all punishment conditions, suggesting that punishment generally led to lower sensitivity to bribe size (see Fig. 2). Accordingly, when all punishment conditions were compared with the control condition, the interaction of bribe size with punishment was significantly negative, b = −0.114, 95% CI [−0.212, −0.017], p = .022 (see Fig. 2). That is, the effect of punishment was present only for high bribes, b = −0.099, 95% CI [−0.185, −0.013], p = .025,[9] and there was no effect for low bribes, b = −0.007, 95% CI [−0.082, 0.069], p = .865.

5.3. Effect of fines and task termination

Next, we conducted an analysis of the effect of punishment using only data from the fine conditions (see Table 2 for results). Participants were less likely to take a bribe with increasing size of punishment as well as with increasing probability of punishment, supporting hypotheses 1 and 2 in the fine condition. The quadratic effect of punishment probability was also significant, suggesting that the effect of increasing probability of punishment was smaller between the two higher probabilities (5 and 25%) than between the two lower probabilities (1 and 5%). At odds with hypothesis 3, there was no interaction between punishment probability and size, suggesting that the effect of punishment size was not moderated by punishment probability. Participants were also less likely to take bribes in later trials.

By selecting the trials following trials where a bribe was taken and adding the administration of punishment on these previous trials in the model, we also examined whether punishment after taking a bribe influenced the probability of taking the next bribe. We found that there was no difference between the probability of taking a bribe after a previous taken bribe was followed by a punishment or not, b = 0.002, 95% CI [−0.059, 0.063], p = .952 (see Table S1).[10]

The end condition was compared with the two fine conditions by adding the end condition in the model with fine conditions using treatment coding (see Table S2 for results). There was no significant difference between the probability of taking a bribe of participants in the end condition and either of the fine conditions, ps > 0.19. However, the effect of probability of punishment differed between the 40-fine and end conditions, b = −0.108, 95% CI [−0.214, −0.002], p = .046. The interaction showed that while the end condition led to lower probability of taking a bribe than 40-fine condition for 1% probability of punishment, b = 0.101, 95% CI [0.019, 0.182], p = .017, there was no difference between the two conditions for 5% punishment and 25% punishment probability, ps > 0.84. That is, when the probability of punishment was low, participants were less likely to take a bribe if they could be punished by termination of the task than if they could be punished by a small fine. None of the other interactions between probability and size of punishment was significant, ps > 0.24.

Next, we conducted a linear regression only with participants in the end conditions (see Table 2 for results). At odds with hypothesis 2, the probability of punishment did not influence the probability of taking a bribe in the end conditions. However, the interaction of bribe size with the linear effect of the probability of punishment suggested that with

[7] While we pre-registered use of the mixed-effect logistic regression, the models did not converge properly, so we used the mixed-effect linear regression at the end. The linear regression approach is sometimes recommended because it usually leads to similar results and its results are easier to interpret (Gomila, 2021). We also pre-registered analyses using interaction of presence of a bribe with the factor of interest to test its effect on correct classification. However, the rate of incorrect classifications in trials without bribes was generally very low and did not differ appreciably between different conditions. Testing the effect using interactions was therefore deemed unnecessary and possibly leading to lower statistical power, so only results of models using trials with bribes are reported. We also pre-registered using the order of the trial as a covariate, but the order of the trial was confounded with taking a bribe in the end condition given that those who took bribes were less likely to finish all trials.

[8] The linear and quadratic coding was -0.707, 0, 0.707 and 0.408, -0.816, 0.408 for the 1%, 5%, and 25% punishment probabilities respectively. The coded variables thus test linear and quadratic effects of punishment probability and they are not correlated with each other.

[9] Note that the threshold for "high" bribes (130 and higher) was selected based on Fig. 3 and the significance of the effect is not robust to the threshold selection even if the pattern of results is similar. The effect of punishment for high bribes is not significant when the threshold is 110 (p = .088) and 120 (p = .058), but it is significant for threshold 140 (p = .022) and 150 (p = .024).

[10] Links to online supplementary results are at the end of the manuscript.

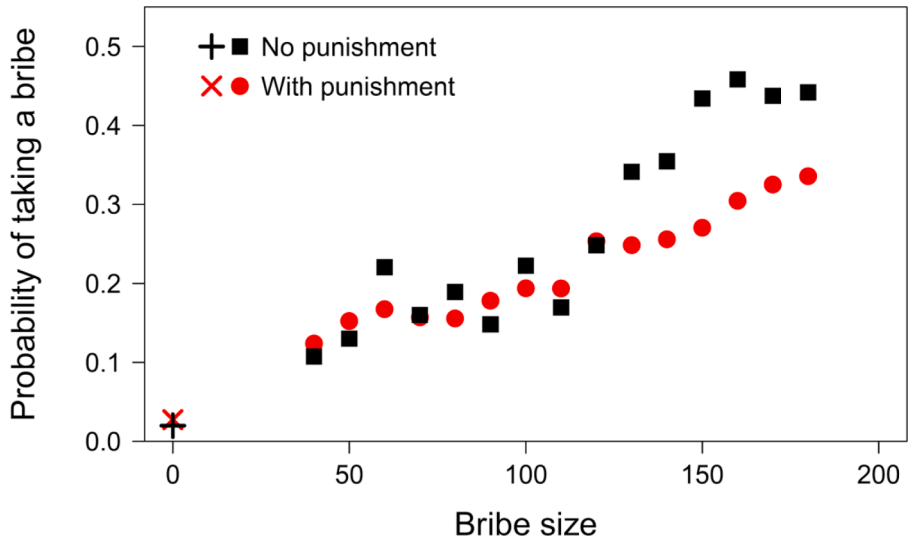

Fig. 2. The effect of bribe size on the probability of taking a bribe.
The figure shows the average probability of taking a bribe for bribes of different sizes. The cross and a saltire show the probability of sorting the object according to the shape in trials without a bribe for comparison.

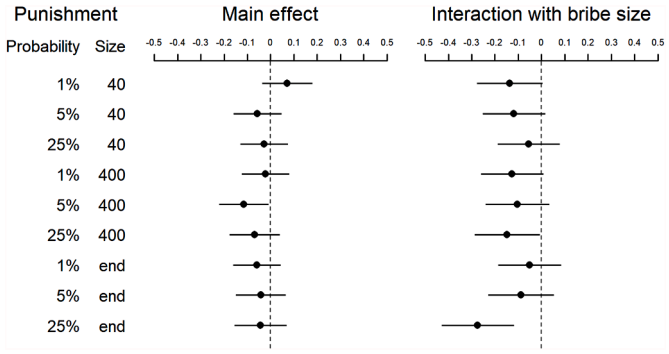

Fig. 3. The effect of punishment on the probability of taking a bribe.
The figure shows estimates of the difference in the probability of taking a bribe between punishment conditions and the control condition and the interaction of the condition effect with bribe size. The error bars show 95% confidence intervals.

increasing probability of punishment, the effect of bribe size was smaller.

5.4. Individual differences

Adding standardized honesty-humility scores in the model with all conditions from the section 5.2 showed that participants higher in honesty-humility were less likely to take bribes, $b = -0.076$, 95% CI [−0.098, −0.053], $p < .001$. Hypothesis 4 was therefore supported. Similarly, participants higher in emotionality were less likely to take bribes, $b = -0.050$, 95% CI [−0.073, −0.027], $p < .001$. Adding an interaction of emotionality with the presence of punishment in the model (see Table S3), suggested that the association of bribe-taking with emotionality might be mostly driven by a decreased probability of taking a bribe in the presence of punishment, $b = -0.073$, 95% CI [−0.146, 0.000], $p = .051$. While emotionality was associated with a lower probability of taking a bribe for participants in conditions with punishment, $b = -0.058$, 95% CI [−0.082, −0.034], $p < .001$, there was no association between emotionality and bribe-taking without punishment, $b = 0.010$, 95% CI [−0.063, 0.084], $p = .785$. The interaction of honesty-humility with the presence of punishment suggested an opposite effect, $b = 0.063$, 95% CI [0.000, 0.126], $p = .050$. The difference in bribe-taking between people low and high in honesty-humility was larger in conditions without punishment, $b = -0.133$, 95% CI [−0.186, −0.079], $p < .001$, than in conditions with punishment, $b = -0.066$, 95% CI [−0.090, −0.041], $p < .001$.

Table 2
The effect of probability and size of punishment. The table shows results of mixed-effect regression models separate for fine and end conditions with incorrectness of object classification as a binary dependent variable. Standard errors are in parentheses.

	Fine conditions	End conditions
Punishment probability (linear)	−0.056**	0.013
	(0.027)	(0.038)
Punishment probability (quadratic)	0.057**	−0.007
	(0.027)	(0.038)
Punishment size	−0.046**	
	(0.022)	
Bribe size	0.228***	0.201***
	(0.020)	(0.030)
Trial order (linear)	−0.063***	
	(0.021)	
Trial order (quadratic)	−0.020*	
	(0.011)	
Punishment probability (l) x Bribe size	0.017	−0.151***
	(0.035)	(0.054)
Punishment probability (q) x Bribe size	0.007	−0.057
	(0.035)	(0.052)
Punishment size x Bribe size	−0.016	
	(0.029)	
Punishment probability (l) x Punishment size	0.036	
	(0.038)	
Punishment probability (q) x Punishment size	0.003	
	(0.037)	
Constant	0.237***	0.230***
	(0.016)	(0.022)
Observations	8971	3567

* $p<.1$.
** $p<.05$.
*** $p<.01$.

5.5. Task perception

Linear regression showed that in presence of a fine the task was considered in more moral terms, $b = 0.18$, 95% CI [0.02, 0.34], $p = .027$, and taking the bribe was perceived more negatively, $b = 0.23$, 95% CI

[0.04, 0.43], $p = .020$ (see Fig. S1). On the other hand, the end condition differed from the control group neither in the evaluation of the task nor the taking the bribe, $ps > 0.17$.

Comparison of perception of the task and taking a bribe between fine conditions did not show any effect of punishment probability, punishment size, or their interaction, all $ps > 0.33$. The sole exception was a negative association between punishment size and the evaluation of taking a bribe, $b = 0.15$, 95% CI [−0.01, 0.31], $p = .062$, which was, however, still not statistically significant. Thus, we found a limited support for hypothesis 5: only the presence of fines affected perceived morality of the behavior in the task; the evidence for the effect of size of punishment was mixed. We did not find support for hypothesis 6: probability of punishment was unrelated to the perceived morality.

Participants who considered taking bribes in more moral terms took a lower proportion of bribes as shown by Spearman correlation coefficient (r_S), $r_S = -.24$, 95% CI [−.31, −.15], $p < .001$.

6. Discussion

Larger size and higher probability of a fine both decreased the probability of taking a bribe, showing that participants were deterred by the possibility of punishment. Increasing the probability of punishment from 1% to 5% had a larger effect than increasing the probability of punishment from 5% to 25%. It is possible that participants perceive subjectively the increase in probability as more salient for lower probabilities (Tversky & Kahneman, 1992). We did not observe a strong evidence for the hypothesis that punishment crowds out internalized norms against dishonest behavior. Although the presence of a small fine with a low probability seemingly led to an increase of the probability of bribe-taking (see Fig. 2), the probability was not significantly higher in comparison to the control condition. It is possible that, the cost of bribe-taking incurred by the charity in all conditions might have precluded the crowding-out effect as taking a bribe had clear negative consequences despite the low expected costs of the fine.

Participants were more likely to take higher bribes, showing that they responded to the monetary incentives. Given that the possibility of punishment means that taking a bribe is associated with costs, from the economic standpoint, we would expect that punishment would lead participants to not take lower bribes, which may not overcome the negative consequences of punishment. Surprisingly, punishment had an appreciable effect only on the proportion of the higher bribes taken and not on the proportion of the smaller bribes taken. One possible explanation is that the possibility of punishment influences predominantly participants who would have taken only higher bribes without punishment, because only those would pass their threshold for taking a bribe. However, only a small number of participants took all the bribes that they were offered, suggesting that most of the participants were selective in which bribes to take, and they should have been therefore less likely to take smaller bribes if they wanted to maximize their reward. In the real world, higher benefits are associated with higher risks (Pleskac & Hertwig, 2014), which could make the possibility of punishment more salient when a higher bribe is offered. In a previous study using the same paradigm, participants spent more time on trials with higher bribes (Vranka & Bahník, 2018), which suggests that they have more time to consider the negative impact of punishment. Additionally, we cannot rule out the possibility that participants who took even lower bribes were those who paid less attention to the instructions and were thus also less affected by the manipulations of punishment.

The actual administration of punishment did not deter participants in taking a bribe when the opportunity occurred again. The effect of punishment therefore seemed to be mostly in reducing the probability of taking a bribe just by its presence rather than by the actual administration. It is also possible that the punishment dissuaded some participants from taking the next bribe while other participants wanted to compensate for the financial loss from the punishment and thus took the next bribe with a higher probability. These two effects could have canceled each other in the aggregate. Unlike in the study by Chaudhuri, Paichayontvijit, and Sbai (2016) which showed reduction of bribe acceptance after punishment in a bribery game, punishment administration provided no information in our study given that participants knew the precise probability and size of punishment in advance. It is possible that administration of punishment has a deterrent effect when it leads people to perceive the punishment as more likely.

Replicating the results of Vranka and Bahník (2018), honesty-humility negatively correlated with the proportion of bribes taken, showing that it consistently predicts cheating in the task—especially when there is no punishment. It is possible that punishment introduces an additional motivation for not taking a bribe than honesty of the person and the predictive value of honesty-humility therefore decreases. While honesty-humility has been shown to predict dishonest behavior in a number of laboratory tasks (Hilbig & Zettler, 2015), our results show that its predictive power could depend on contextual factors such as the presence of punishment. Somewhat surprisingly, higher emotionality was also associated with a lower proportion of bribes taken. A previous study using a different paradigm found that participants with higher neuroticism, a trait closely related to emotionality, behaved more dishonestly (Conrads, Irlenbusch, Rilke, & Walkowitz, 2013). In our study, the association of higher emotionality with abstaining from taking bribes seemed to be driven by participants who could be punished for taking a bribe, which could explain the difference in results. People with higher emotionality may be more worried by punishment or overestimate its likelihood and they might therefore respond to it more strongly. When there is no punishment, emotionality does not seem to play a role (Vranka & Bahník, 2018) or might be positively associated with dishonest behavior (Conrads et al., 2013).

7. Conclusion

In summary, we showed that bribe-taking in a laboratory task is deterred by fines and the deterrent effect increases with the increase of size and probability of punishment. Unlike previous studies, we did not observe any clear indication of the crowding-out effect of small and unlikely punishment, even in a condition in which the punishment had virtually no consequential monetary effects. On the other hand, our results suggest that effects of punishment might depend on personality characteristics and do not affect behavior of all participants uniformly. Specifically, the risk of punishment might deter those with higher emotionality more strongly, while those higher on honesty-humility behave honestly even when no punishment is present. These findings support the idea that by focusing on average effects of incentives, studies may miss many important interindividual differences and that policies informed by research findings might lead to unexpected results when applied to a real-world setting to which people with specific characteristics self-select (Houdek, Bahník, Hudík, & Vranka, 2021).

Declaration of Competing Interest

None.

Acknowledgments

The research was supported by GAČR Project No. 19–10781S. The funder had no role in study design, data collection and analysis, or preparation of the manuscript. We would like to thank Nikola Frollová, Jan Kasalický, Dominik Stříbrný, Markéta Sýkorová, and Vojtěch Zíka for research assistance.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at https://osf.io/zfcw9/.

References

Abbink, K., Irlenbusch, B., & Renner, E. (2002). An experimental bribery game. *Journal of Law, Economics, and Organization, 18*, 428–454.

Abbink, K., & Serra, D. (2012). Anticorruption policies: Lessons from the lab. In D. Serra, & L. Wantchekon (Eds.), *New advances in experimental research on corruption. research in experimental economics* (pp. 77–115). Bingley, UK: Emerald Group Publishing *(Vol. 15)*.

Alatas, V., Cameron, L., Chaudhuri, A., Erkal, N., & Gangadharan, L. (2009). Subject pool effects in a corruption experiment: A comparison of Indonesian public servants and Indonesian students. *Experimental Economics, 12*(1), 113–132.

Armantier, O., & Boly, A. (2012). On the external validity of laboratory experiments on corruption. In D. Serra, & L. Wantchekon (Eds.), *New advances in experimental research on corruption. research in experimental economics* (pp. 117–144). Bingley, UK: Emerald Group Publishing *(Vol. 15)*.

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*(4), 340–345.

Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO honesty-humility, agreeableness, and emotionality factors: A review of research and theory. *Personality and Social Psychology Review, 18*(2), 139–152.

Banerjee, R., & Mitra, A. (2018). On monetary and non-monetary interventions to combat corruption. *Journal of Economic Behavior & Organization, 149*, 332–355.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy, 76*, 169–217.

Boly, A., & Gillanders, R. (2018). Anti-corruption policy making, discretionary power and institutional quality: An experimental analysis. *Journal of Economic Behavior & Organization, 152*, 314–327.

Boly, A., Gillanders, R., & Miettinen, T. (2019). Deterrence, contagion, and legitimacy in anticorruption policy making: An experimental analysis. *The Journal of Legal Studies, 48*(2), 275–305.

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

Chaudhuri, A., Paichayontvijit, T., & Sbai, E. (2016). The role of framing, inequity and history in a corruption game: Some experimental evidence. *Games, 7*(2), 13.

Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology, 34*, 1–7.

Cubitt, R. P., Starmer, C., & Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics, 1*(2), 115–131.

da Hora, K. L., & Sampaio, A. A. (2019). Units of analysis for corruption experiments: Operant, culturobehavioral lineage, culturant, and macrobehavior. *Perspectives on Behavior Science, 42*(4), 751–771.

Drugov, M., Hamman, J., & Serra, D. (2014). Intermediaries in corruption: An experiment. *Experimental Economics, 17*(1), 78–99.

Dušek, L., & Traxler, C. (2017). Experience with punishment and specific deterrence: Evidence from speeding tickets. *Mimeo*.

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys, 15*(5), 589–611.

Friesen, L. (2012). Certainty of punishment versus severity of punishment: An experimental investigation. *Southern Economic Journal, 79*(2), 399–421.

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies, 29*, 1–17.

Goel, R. K., & Nelson, M. A. (2011). Measures of corruption and determinants of US corruption. *Economics of Governance, 12*(2), 155–176.

Goel, R. K., & Rich, D. P. (1989). On the economic incentives for taking bribes. *Public choice, 61*, 269–275.

Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General, 150*(4), 700–709.

Hanna, R., Bishop, S., Nadel, S., Scheffler, G., & Durlacher, K. (2011). The effectiveness of anti-corruption policy. *EPPI Centre Report, 3*(1).

Hilbig, B. E., & Zettler, I. (2015). When the cat's away, some mice will play: A basic trait account of dishonest behavior. *Journal of Research in Personality, 57*, 72–88.

Hilbig, B. E., Zettler, I., & Heydasch, T. (2012). Personality, punishment and public goods: Strategic shifts towards cooperation as a matter of dispositional honesty–humility. *European Journal of Personality, 26*(3), 245–254.

Houdek, P., Bahník, Š., Hudík, M., & Vranka, M. A. (2021). Selection effects on dishonest behavior. *Judgment and Decision Making, 16*, 238–266.

Jain, A. K. (2001). Corruption: A review. *Journal of Economic Surveys, 15*, 71–121.

Khadjavi, M. (2014). On the interaction of deterrence and emotions. *The Journal of Law, Economics, & Organization, 31*(2), 287–319.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26.

Laske, K., Saccardo, S., & Gneezy, U. (2018). Do fines deter unethical behavior? The effect of systematically varying the size and probability of punishment. *The effect of systematically varying the size and probability of punishment (April 5, 2018)*. Available at SSRN https://ssrn.com/abstract=3157387.

Lawpoolsri, S., Li, J., & Braver, E. R. (2007). Do speeding tickets reduce the likelihood of receiving subsequent speeding tickets? A longitudinal study of speeding violators in Maryland. *Traffic Injury Prevention, 8*(1), 26–34.

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.

Mazar, N., & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing, 25*, 11–126.

Nagin, D. S. (2013). Deterrence in the twenty-first century. *Crime and Justice, 42*, 199–263.

Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General, 143*(5), 2000–2019.

Qin, X., & Wang, S. (2013). Using an exogenous mechanism to examine efficient probabilistic punishment. *Journal of Economic Psychology, 39*, 1–10.

Salmon, T. C., & Serra, D. (2017). Corruption, social judgment and culture: An experiment. *Journal of Economic Behavior & Organization, 142*, 64–78.

Schildberg-Hörisch, H., & Strassmair, C. (2010). An experimental test of the deterrence hypothesis. *The Journal of Law, Economics, & Organization, 28*(3), 447–459.

Schulze, G. G., & Frank, B. (2003). Deterrence versus intrinsic motivation: Experimental evidence on the determinants of corruptibility. *Economics of Governance, 4*(2), 143–160.

Sequeira, S. (2012). Advances in measuring corruption in the field. In D. Serra, & L. Wantchekon (Eds.), *New advances in experimental research on corruption. research in experimental economics* (pp. 145–175). Bingley, UK: Emerald Group Publishing *(Vol. 15)*.

Serra, D., & Wantchekon, L. (2012). Experimental research on corruption: introduction and overview. In D. Serra, & L. Wantchekon (Eds.), *New advances in experimental research on corruption. research in experimental economics* (pp. 1–11). Bingley, UK: Emerald Group Publishing (Vol. 15).

Starmer, C., & Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *The American Economic Review, 81*(4), 971–978.

Thielmann, I., Moshagen, M., Hilbig, B., & Zettler, I. (2021). On the comparability of basic personality models: Meta-analytic correspondence, scope, and orthogonality of the big five and HEXACO dimensions. *European Journal of Personality*. https://doi.org/10.1177/08902070211026793

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323.

van Veldhuizen, R. (2013). The influence of wages on public officials' corruptibility: A laboratory investigation. *Journal of Economic Psychology, 39*, 341–356.

Van Winden, F., & Ash, E. (2009). On the behavioral economics of crime. *Review of Law & Economics, 8*(1), 181–213.

Vranka, M. A., & Bahník, Š. (2018). Bureaucracy game: A new computer task for the experimental study of corruption. *Frontiers in Psychology, 9*, 1511. https://doi.org/10.3389/fpsyg.2018.01511

Zettler, I., & Hilbig, B. E. (2010). Honesty–humility and a person–situation interaction at work. *European Journal of Personality, 24*(7), 569–582.