

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

Štěpán Bahník and Marek A. Vranka*

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

* Corresponding author: vranka.marek@gmail.com

Declarations of interest: none

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.

Keywords: corruption; bribe-taking; punishment; laboratory experiment; HEXACO

JEL: C91, D73, D91, K42

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 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 violate any social or moral norms (Friesen, 2012). In the usually employed experimental tasks, either nobody is harmed or the harm done to other participants is seen as an intended part of playing the experimental “game”. 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. Moreover, punishment interacts with internal costs of wrongdoing, sometimes even increasing the likelihood of corrupt behavior by focusing on its positive expected value (Tenbrunsel & Messick, 1999). 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. Because the perceived internal costs of corruption depend on personality, there are also likely interindividual differences in effects of punishment, which is an area that did not get much attention in the previous research.

In our study, we aimed to fill these gaps in the knowledge of the effects of punishment in laboratory research of corruption. 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. Unlike the previous studies (e.g., Abbink et al., 2002; Banerjee & Mitra, 2018; Schildberg-Hörisch & Strassmair, 2010; Schulze & Frank, 2003), we employ multiple forms and sizes of punishment with independently varied probabilities, allowing us to test for their main effects as well as possible interactions. We employ 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 manipulate the probability of punishment after taking a bribe. While there was no punishment in the control group, participants in the nine experimental groups faced punishment of varying size (low fine, high fine, task termination) and probability (1%, 5%, 25%) everytime they broke the sorting rule. In case of the task termination, the sorting task ended and participants were not able to gain any additional reward.

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, because corrupt behavior in the experiment was associated with breaking rules and causing harm to a party not involved in the task itself, we were able to explore effects of punishment on perceived morality of various behaviors in the task which is associated with 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.

Our results show that both higher size and probability of fines deter bribe-taking and that there is no evidence of an interaction between the size and probability of punishment. Task termination had a similar effect as punishment by fines. Punishment generally worked by decreasing the probability of taking higher bribes rather than influencing behavior for low bribes. Personality characteristics moderated the effect of punishment. We find no strong evidence for punishment crowding out internalized norms against dishonest behavior and no evidence of deterrence effect of administered punishment on taking following bribes.

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.

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 penalty¹ 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 et al., 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 punishment² 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 and colleagues (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, Irlenbusch, & Renner, 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

¹ 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.

² Specifically, the probability of apprehension.

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 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 affect 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.

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, making them unsuitable for the study of the interplay between effects of external punishment and internal costs of wrongdoing. However, when one is interested also in factors related to the internalized psychological norms against corrupt behavior and their possible crowding out, it is essential that participants perceive the task in moral terms. Because the usually used tasks use abstract and neutral language and do not contain violation of any rules or norms, it is unclear whether participants even understand that the transaction in question represents corruption. To overcome these issues, we use a task developed by Vranka and Bahník (2018). The task models corruption in a bureaucratic setting, in which a public official can gain a reward for providing a favor for another person (Jain, 2001).

3 Hypotheses

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. In line with our pre-registered analysis plan, we

hypothesize that (H1) punishment size and (H2) punishment probability will influence the perception of morality of accepting and refusing an offered bribe in the task.

Furthermore, 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, we expect that (H3) punishment size and (H4) punishment probability will influence the likelihood of taking a bribe and that (H5) the effect of punishment probability on the likelihood of taking a bribe will be moderated by the punishment size.

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 (H6) an association between honest-humility and the likelihood of taking a bribe, however, we explore possible association between the bribe-taking and the remaining five HEXACO personality dimensions as well.

4 Methods³

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 an effect $d = 0.25$ with power .80. A more precise power analysis is complicated by the nature of the repeated measures design of the experiment and by complexity of mixed-effect regression, which was used for 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. Afterwards, they were explained the task and they were told that they have a 20% probability that the points they would earn during the task would be converted to a monetary reward using the conversion rate 10 points = 1 CZK.

³ The materials used in the study, data, analysis scripts, as well as pre-registration of the study are available at: https://osf.io/szrtd/?view_only=88f0d85ec105456ead57fe57ce12391a

Participants were told to sort objects running on a computer screen according to their color. The objects had three possible shapes and colors. 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. Colors associated with the three keys were randomly determined for each trial and they were displayed prominently at the bottom of the screen. At the beginning of the task, 2000 points (corresponding to 200 CZK, ~9 USD) were allotted to a charity. If a key response led to an assignment to a wrong color, the charity lost 200 points. The loss simulated negative societal effects of not performing given work according to the given rule. 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, in trials where the two sorting criteria were mismatched, there was a 22.5% probability that a given object was associated with a “bribe”. These objects were shown with a number corresponding to the value of the bribe, which a participant got if they sorted the object according to its shape. The bribe size varied from 40 to 190 points. The loss to the charity was therefore always larger than a possible gain for the participant. Each participant went through 200 trials of the task; i.e., 200 objects to sort. The shapes and colors of objects as well as bribes were randomly determined for each participant. 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 Figure 1).

Trial number: 2/200

For the charity: 2000

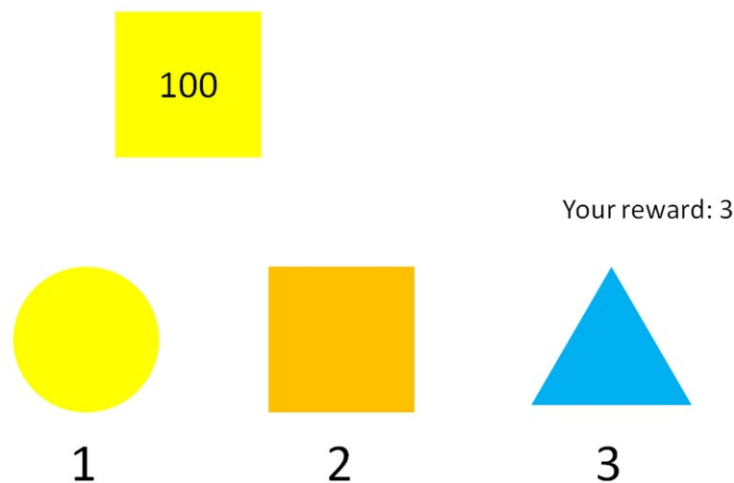


Figure 1. An illustration of a computer screen seen by a participant. The top row shows information about the number of the current trial, the total number of trials, and the number of points currently assigned to the charity organization. In the middle of the screen, an object (a yellow square in this 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. Adapted from "Bureaucracy game: A new computer task for the experimental study of corruption," by M. A. Vranka and Š. Bahník, 2018, *Frontiers in Psychology*, 9:1511, p. 3. Copyright 2018 by Vranka and Bahník.

To minimize reputational concerns of cheating, participants were given a lottery after the task, in which they were able to win additional money. 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. 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. 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) and the experimental groups had the punishment described in the instructions before the task.

Condition	Punishment	Punishment probability	Punishment size
Control	X	-	-
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

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

5 Results⁴

Section 5.1 describes the effect of presence of punishment on perception of the task and bribe taking. Section 5.2 focuses on overall performance in the task measured by final rewards and the proportion of correct classifications. The remaining sections report analyses on a trial level, looking at individual decisions to take a bribe or not. In particular, section 5.3 describes the effect of presence of any punishment, sections 5.4 and 5.5 focus on effects of fines and task termination respectively. Finally, section 5.6 builds on the model from section 5.3 by adding individual differences measures in analysis.

5.1 Task perception

Linear regression showed that in presence of a fine the task was considered in more moral terms, $t(362) = 2.21$, $p = .027$, $b = 0.18$, 95% CI [0.02, 0.34], and taking the bribe was perceived more negatively, $t(362) = 2.35$, $p = .020$, $b = 0.23$, 95% CI [0.04, 0.43]. On the other hand, the end condition differed from the control group neither in the evaluation of the task nor the taking the bribe, $ps > .17$.

⁴ Although we had pre-registered the study, we also conducted additional exploratory analyses. In particular, the analysis related to 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.

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 $ps > .33$. The sole exception was a negative association between punishment size and the evaluation of taking a bribe, $t(302) = 1.87, p = .062, b = 0.15, 95\% \text{ CI } [-0.01, 0.31]$, which was, however, still not statistically significant. Thus, we found a limited support for hypothesis 1: 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 2: 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), $r_s = -.24, 95\% \text{ CI } [-.31, -.15], p < .001$.

5.2 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% 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%).

5.3 Effect of punishment

Trial-level analysis was conducted using mixed-effect linear regression.⁵ The incorrectness of object classification, i.e. 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) contrasts were used for the

⁵ 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, 2020). 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.

probability of punishment.⁶ Random intercepts for participants were included in the model alongside random slopes for bribe size. 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, $t(489.0) = 14.37, p < .001, b = 0.232$, 95% CI [0.200, 0.263]. 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, $t(477.5) = -1.99, p = .048, b = -0.102$, 95% CI [-0.203, -0.001] (see Figure 2).

When an interaction of bribe size with condition was added in the model, the interaction with bribe size was significant for 25% 400-fine condition, $t(461.4) = -2.09, p = .037, b = -0.147$, 95% CI [-0.285, -0.009], and for 25% end condition, $t(558.0) = -3.51, p < .001, b = -0.274$, 95% CI [-0.427, -0.121]. However, the interaction was negative for all punishment conditions, suggesting that punishment generally led to lower sensitivity to bribe size (see Figure 2). Accordingly, when all punishment conditions were compared with the control condition, the interaction of bribe size with punishment was significantly negative, $t(471.3) = -2.30, p = .022, b = -0.114$, 95% CI [-0.212, -0.017] (see Figure 3). That is, the effect of punishment was present only for high bribes, $t(488.2) = -2.25, p = .025, b = -0.099$, 95% CI [-0.185, -0.013],⁷ and there was no effect for low bribes, $t(484.8) = -0.17, p = .865, b = -0.007$, 95% CI [-0.082, 0.069].

⁶ The linear and quadratic contrast coding was -0.707, 0, 0.707 and 0.408, -0.816, 0.408 for the 1%, 5%, and 25% punishment probabilities respectively. The contrasts thus test linear and quadratic effects of punishment probability.

⁷ Note that the threshold for “high” bribes (130 and higher) was selected based on Figure 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 ($p = .024$).

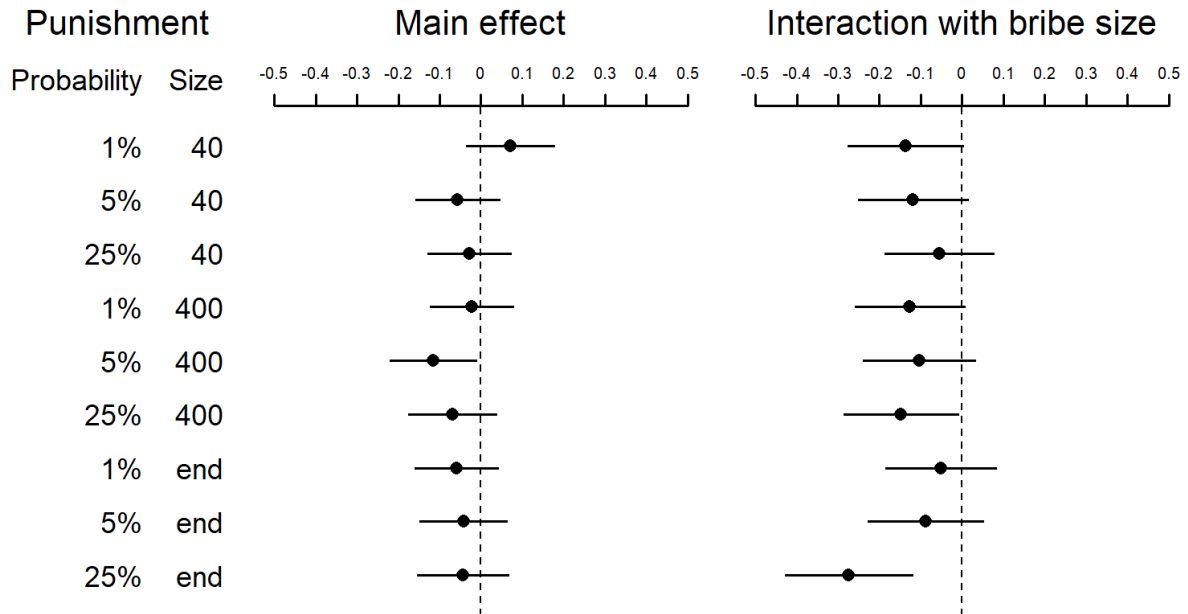


Figure 2. The effect of punishment on the probability of taking a bribe. The figure shows estimates of the difference in the probability of taking a bribe between punishment conditions and the control condition and the interaction of the condition effect with bribe size. The error bars show 95% confidence intervals.

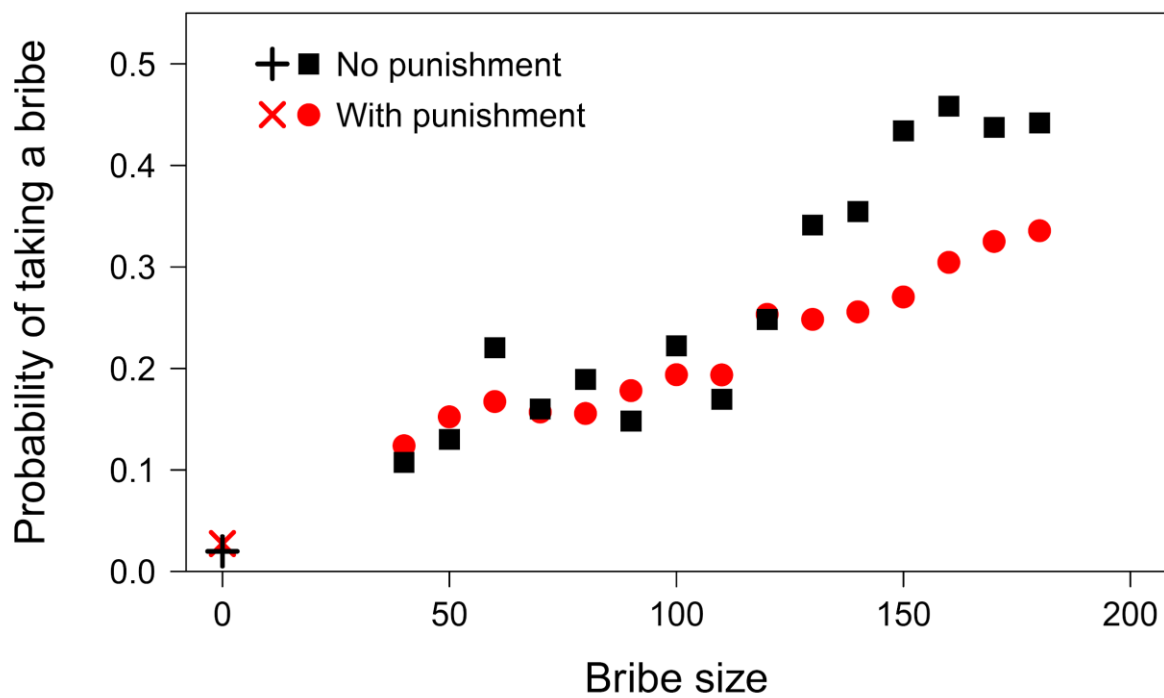


Figure 3. The effect of bribe size on the probability of taking a bribe. The figure shows the average probability of taking a bribe for bribes of different sizes. The cross and a saltire show the probability of sorting the object according to the shape in trials without a bribe for comparison.

5.4 Effect of fines

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 3 and 4 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 5, 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.

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, $t(440.7) = 0.06$, $p = .952$, $b = 0.002$, 95% CI [-0.059, 0.063].

	Fine conditions	End conditions
Punishment probability (linear)	-0.056* (-0.109, -0.003)	0.013 (-0.062, 0.088)
Punishment probability (quadratic)	0.057* (0.004, 0.109)	-0.007 (-0.082, 0.068)
Punishment size	-0.046* (-0.089, -0.003)	
Bribe size	0.228*** (0.188, 0.268)	0.201*** (0.142, 0.261)
Trial order (linear)	-0.063** (-0.103, -0.022)	
Trial order (quadratic)	-0.020 (-0.041, 0.002)	
Punishment probability (l) x Bribe size	0.017 (-0.052, 0.087)	-0.151** (-0.256, -0.046)
Punishment probability (q) x Bribe size	0.007 (-0.061, 0.076)	-0.057 (-0.158, 0.044)
Punishment size x Bribe size	-0.016 (-0.072, 0.041)	
Punishment probability (l) x Punishment size	0.036 (-0.037, 0.110)	
Punishment probability (q) x Punishment size	0.003 (-0.070, 0.076)	
Constant	0.237*** (0.206, 0.268)	0.230*** (0.186, 0.273)
Observations	8,971	3,567
<i>Note:</i>	* p<0.05; ** p<0.01; *** p<0.001	

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. 95% confidence intervals are in parentheses.

5.5 Effect of task termination

The end condition was compared with the two fine conditions by adding the end condition in the model with fine conditions using treatment contrasts. There was no significant difference between the probability of taking a bribe of participants in the end condition and either of the fine conditions, $ps > .19$. However, the effect of probability of punishment differed between

the 40-fine and end conditions, $t(447.6) = -2.00, p = .046, b = -0.108, 95\% \text{ CI } [-0.214, -0.002]$. 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, $t(95.5) = 2.43, p = .017, b = 0.101, 95\% \text{ CI } [0.019, 0.182]$, there was no difference between the two conditions for 5% punishment and 25% punishment probability, $ps > .84$. 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, $ps > .24$.

Next, we conducted a linear regression only with participants in the end conditions (see Table 2 for results). At odds with hypothesis 4, 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 contrast for the probability of punishment suggested that with increasing probability of punishment, the effect of bribe size was smaller.

5.6 Individual differences

Adding standardized honesty-humility scores in the model with all conditions from the section 5.3 showed that participants higher in honesty-humility were less likely to take bribes, $t(486.4) = -6.68, p < .001, b = -0.076, 95\% \text{ CI } [-0.098, -0.053]$. Hypothesis 6 was therefore supported. Similarly, participants higher in emotionality were less likely to take bribes, $t(492.7) = -4.24, p < .001, b = -0.050, 95\% \text{ CI } [-0.073, -0.027]$. Adding an interaction of emotionality with the presence of punishment in the model, 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, $t(478.1) = -1.96, p = .051, b = -0.073, 95\% \text{ CI } [-0.146, 0.000]$. While emotionality was associated with a lower probability of taking a bribe for participants in conditions with punishment, $t(435.6) = -4.69, p < .001, b = -0.058, 95\% \text{ CI } [-0.082, -0.034]$, there was no association between emotionality and bribe-taking without punishment, $t(55.7) = 0.27, p = .785, b = 0.010, 95\% \text{ CI } [-0.063, 0.084]$. The interaction of honesty-humility with the presence of punishment suggested an opposite effect, $t(475.7) = 1.96, p = .050, b = 0.063, 95\% \text{ CI } [0.000, 0.126]$. Participants higher in honesty-humility in conditions without punishment were even less likely to take a bribe, $t(55.6) = -4.85, p < .001, b = -0.133, 95\% \text{ CI } [-0.186, -0.079]$, than participants higher in honesty-humility in conditions with punishment, $t(430.5) = -5.31, p < .001, b = -0.066, 95\% \text{ CI } [-0.090, -0.041]$.

6 Discussion and conclusions

We found that participants were more likely to classify an object incorrectly—to take a “bribe”—when the misclassification was associated with a monetary gain. The probability of misclassification was further increased in case of higher bribes. The presence of punishment decreased the rate of taking high bribes, but not of taking low bribes—which were, however, taken only relatively rarely even in the control condition without any punishment. The punishment in the form of a fine led to a lower rate of taking bribes when the fine was larger and more probable. Getting a fine after taking a bribe did not lead to a decreased probability

of taking a bribe when the opportunity was again presented. Apart from the effects of punishment, we also found that participants higher on honesty-humility and higher on emotionality were less likely to take bribes.

The task was perceived only in slightly more moral terms when a fine for cheating was present, but not when the punishment led to termination of the task. The effect of fines on task perception did not seem to depend on the size and probability of the fine. It is notable that the effect on task perception persisted even after a delay given that the questions about task perception were given at the end of the experiment. People who perceived the task in more moral terms were also less likely to take bribes in the task. Similarly as in the previous study using the task (Vranka & Bahník, 2018), participants tended to view not taking a bribe positively, but they did not generally agree that taking a bribe would be negative. This could, however, be at least partially caused by ex-post rationalization of their own conduct during the task as the task perception and bribe-taking were correlated.

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 be dependent 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).

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 taken higher bribes and not on the proportion of taken smaller bribes. 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.

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). From the economic analysis perspective, it could be expected that the effect of size and probability of punishment would be multiplicative rather than additive. However, we found no interaction between the two factors. It is possible that people evaluate the two factors separately rather than computing expected costs of taking a bribe from the probability and size of punishment. This is in line, for example, with results by Banerjee and Mitra (2018) who show that the deterrence effect of punishment depends more on its probability rather than its overall expected costs.

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 Figure 2), the probability was not significantly higher in comparison to the control condition. It is possible that while the fine was quite small and unlikely, it still was not small and unlikely enough and worked as a deterrent at least for some of the participants. However, this seems unlikely, as the fine in question would occur only once for one hundred bribes taken on average and its size was equal to the size of the smallest bribe offered. Alternatively, 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.

Given that the prospect of termination of the task did not appreciably deter participants in taking bribes and it did not lead participant to perceive the task in more moral terms, unlike the presence of a fine, it does not seem that the punishment in the form of termination of the task could improve appreciably laboratory models of corruption or dishonest behavior in general. The possibility of termination introduces various problems in data analysis given that some of the cheating participants will not go through the same number of trials as other participants. We did not test the effect of punishment in the form of termination of the task alongside punishment in the form of a fine, which would better correspond to the punishment in many environments in the real world. It is possible that this combination would lead to higher external validity of the manipulation than the presence of a fine itself. While the

punishment in the form of termination did not seem to change participants' behavior significantly, the punishment in the form of the fine influenced both perception of the task and behavior and it could therefore be used in future laboratory studies of dishonesty to increase their external validity.

Finally, 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 cancelled each other in the aggregate.

Although corruption in the form of bribery always requires two parties—a person offering a bribe and a person taking a bribe—we focused on the latter, similar as Schulze and Frank (2003), because the dishonest behavior of a bribee is what institutions could realistically attempt to target and eliminate. Ignoring behavior of a briber limits generalizability of the findings for circumstances where behavior of the briber is important, such as when the interaction of the briber and bribee is repeated. The used task therefore better models situations where the interaction is one-time or anonymous or situations where corruption takes place without interaction with any other party, as in embezzlement.

Despite the relatively large sample size, p-values associated with some of the significant results were still rather large and did not reach a stricter definition of significance (Benjamin et al., 2018). Some of the analyses were also only exploratory and the analysis deviated from the pre-registration due to various issues that we have not anticipated before seeing the data. The results of exploratory analyses and the results that do not pass a stricter significance threshold should be treated as suggestive, and their further replication is needed to show whether they are reliable.

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 more strongly those with higher emotionality, 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, in press).

Acknowledgements

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.

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 (Vol. 15)* (pp. 77-115). Bingley, UK: Emerald Group Publishing.
- 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 (Vol. 15)* (pp. 117-144). Bingley, UK: Emerald Group Publishing.
- 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.
- 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.
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... & Cesarini, D. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6-10.
- 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.
- Chetwynd, E., Chetwynd, F., & Spector, B. (2003). Corruption and poverty: A review of recent literature. *Management Systems International*.
- Conrads, J., Irlenbusch, B., Rilke, R. M., & Walkowitz, G. (2013). Lying and team incentives. *Journal of Economic Psychology*, 34, 1-7.
- 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.

- De Vries, R. E., Tybur, J. M., Pollet, T. V., & van Vugt, M. (2016). Evolution, situational affordances, and the HEXACO model of personality. *Evolution and Human Behavior*, 37(5), 407-421.
- Drury, A. C., Kriekhaus, J., & Lusztig, M. (2006). Corruption, democracy, and economic growth. *International Political Science Review*, 27, 121-136.
- Dušek, L., & Traxler, C. (2017). Experience with Punishment and Specific Deterrence: Evidence from Speeding Tickets, *Mimeo*.
- Elkamel, H. (2019). Corruption and inflation: Evidence from US states. *Journal of Financial Economic Policy*, 11(2), 251-262.
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11, 525-547.
- 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. (2020). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*. Advance online publication. <https://doi.org/10.1037/xge0000920>
- Gupta, S., Davoodi, H., & Alonso-Terme, R. (2002). Does corruption affect income inequality and poverty? *Economics of Governance*, 3, 23-45.
- 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. (in press). Selection Effects on Dishonest Behavior. *Judgment and Decision Making*.

- Jain, A. K. (2001). Corruption: A review. *Journal of Economic Surveys*, 15, 71-121.
- Justesen, M. K., & Bjørnskov, C. (2014). Exploiting the poor: Bureaucratic corruption and poverty in Africa. *World Development*, 58, 106-115.
- 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. & Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy & Marketing*, 25, 11-126.
- 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.
- Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 110, 681-712.
- 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.
- Ryvkin, D., Serra, D., & Tremewan, J. (2017). I paid a bribe: An experiment on information sharing and extortionary corruption. *European Economic Review*, 94, 1-22.
- 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 (Vol. 15)* (pp. 145-175). Bingley, UK: Emerald Group Publishing.

- 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 (Vol. 15)* (pp. 1-11). Bingley, UK: Emerald Group Publishing.
- Tenbrunsel, A. E., & Messick, D. M. (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly*, 44(4), 684-707.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297-323.
- 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. doi: 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.