Article

Social Psychological and Personality Science
1–9

DOI: 10.1177/19485506231168515
journals.sagepub.com/home/spp
SAGE

Are Guilt-Prone Power-Holders Less Corrupt? Evidence From Two Online Experiments

Yang Hu[1], Shiwei Qiu[1], Gaotong Wang[2], Kui Liu[2], Weijian Li[2], Hongbo Yu[3], and Xiaolin Zhou[1,2,4]

Abstract

Bribery is ubiquitous in human society. Yet it remains unknown how bribe-taking behaviors of power-holders and underlying psychological processes are affected by guilt-proneness, a crucial moral-related personality trait, and how this trait–behavior association depends on harm salience brought by bribery. To address these questions, we conducted two online experiments (N_{total} = 2, 082) combining economic games with personality measures. Experiment 1 showed that highly guilt-prone individuals were less willing to take bribes, especially when higher harm salience was involved. Leveraging a parametric design with computational modeling, Experiment 2 confirmed the moderation effect of harm salience, and revealed a mediation role of the concern for others' suffering, a key psychological construct in the trait–behavior association. Together, these findings demonstrate a critical function of guilt-proneness in curbing bribe-taking behaviors and suggest the concern for others' suffering as an underlying psychological mechanism.

Keywords

bribery, corruption, guilt-proneness, harm salience, computational model

Introduction

Corruption is generally considered a class of illegal/unethical conduct that leads to widespread and severe economic and social consequences (Graycar & Smith, 2011; Mungiu-Pippidi & Heywood, 2020). Bribery, as one of the most common forms of corruption, occurs in interpersonal contexts in which a power-holder abuses the entrusted power to seek personal gains through immoral collaboration with a briber, often at the expense of others' interests (Barr & Serra, 2009; Köbis et al., 2016). Although previous research has identified several macro- or micro-level antecedents that shape and influence bribery (Lambsdorff, 2006; Lange et al., 2022; Serra & Wantchekon, 2012), little effort has been made to explore the large interindividual differences in bribery affected by individuals' characteristics and the underlying psychological processes (Tanner et al., 2022; Thielmann et al., 2021).

A question of both theoretical and practical importance is what morally related personality traits of individuals in power would predict their acts when they face immoral economic temptation (i.e., taking or refusing bribes). Here, we focus on one particular trait, guilt-proneness, because of its general association with (less) unethical behaviors (Cohen et al., 2012). Guilt-proneness describes a predisposition to anticipate a negative feeling for personal misdeeds (e.g., causing harm to others due to one's own fault), rather than an affective state elicited by a specific event of ethical transgression (guilty feeling; Cohen et al., 2011). Prior evidence has shown a robust negative relationship between guilt-proneness and a variety of unethical behaviors (Cohen et al., 2012). Specifically, individuals who are more guilt-prone make fewer unethical decisions, such as cheating for personal gains (Cohen et al., 2011) or committing counterproductive work behaviors (Cohen et al., 2013). In the same

[1]Shanghai Key Laboratory of Mental Health and Psychological Crisis Intervention, School of Psychology and Cognitive Science, East China Normal University, Shanghai, China
[2]Institute of Psychological and Brain Sciences, Zhejiang Normal University, Jinhua, China
[3]Department of Psychological and Brain Sciences, University of California, Santa Barbara, USA
[4]School of Psychological and Cognitive Sciences, PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China

Corresponding Authors:
Hongbo Yu, Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93117, USA.
Email: hongbo.yu@psych.ucsb.edu

Xiaolin Zhou, School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China.
Email: xz104@psy.ecnu.edu.cn

vein, guilt-proneness might play a crucial role in curbing bribe-taking behaviors in power-holders.

A follow-up question is whether the strength of such an association depends on bribery contexts. One of the contextual variables that define the immoral feature of bribery is *harm salience*, namely, the degree of harm to the third party brought by bribery. In some moral psychology theories, harm is at the core of immorality and harm salience primarily determines how immoral an act is (Schein & Gray, 2018). As a specific type of immoral act, bribery often brings various forms of harm, such as incurring financial costs or physical harm to other individuals (Abbink et al., 2002; Barr & Serra, 2009). People feel guiltier when their behaviors bring a higher level of interpersonal harms (Berndsen et al., 2004; Koban et al., 2013; Zhang et al., 2021) and thus compensate the harmed persons more (Gao et al., 2021). Based on these findings, we could predict that highly guilt-prone power-holders are more concerned with the harm salience brought by bribery and thus are more likely to refuse the bribe than less guilt-prone power-holders, particularly when bribe-taking incurs more harm. However, how harm salience moderates the association between guilt-proneness and immoral acts has not been examined.

Experimentally manipulating harm salience in a bribery scenario would enable us to uncover participants' the concern for others' suffering, a key psychological construct underlying corrupt decision-making, and to further reveal its link to guilt-proneness. It is well-established that people avoid harming others (Graham et al., 2011; Gray et al., 2012; Schein & Gray, 2018; Yu et al., 2019). The concern for others' suffering is considered a core psychological construct of moral cognition and reflects the degree a person has a distaste for harming others (Crockett et al., 2014; Yu et al., 2019). Such a construct is often treated as a latent variable that contributes to the value computation during moral decision-making and can be quantified by implementing computational modeling on choice behavior observed in a task varying in harm salience (Yu et al., 2019). For example, Crockett and colleagues (2014) adopted a novel task in which participants chose between two options involving a trade-off between monetary reward and painful electric shocks inflicted on either themselves or an anonymous person. The authors manipulated harm salience (i.e., the number of shocks) in a parametric and trial-by-trial manner and estimated individuals' concern for others' suffering (labeled harm aversion in this study) by fitting a computational model on their choice behaviors responding to various levels of harm salience and monetary reward.

Previous research has implied a broad link between guilt-proneness and concern for others' suffering. For instance, the level of guilt-proneness is negatively associated with antisocial personality among inmates with a history of harmful criminal acts (Tangney et al., 2011). Highly guilt-prone individuals are less likely to establish a relationship with a competent partner to avoid disappointing the latter (Wiltermuth & Cohen, 2014). When facing a moral trade-off between avoiding harming others and obedience to authority, individuals who are more guilt-prone weigh more on avoiding harm and are more likely to prevent a partner from doing an unpleasant task, even at a cost of disobeying the authority (Ent & Baumeister, 2015). However, these studies did not formally characterize the concern for others' suffering and tended to overlook how such a latent psychological construct is associated with guilt-proneness (Köbis et al., 2016; Serra & Wantchekon, 2012). Answering this question would contribute to the understanding of the mechanisms underpinning the relationship between guilt-proneness and bribe-taking behaviors.

In addition to a trade-off between personal gain and the cost to the victim, deciding whether to take or refuse a bribe may involve other psychological processes. In a typical bribery setting, when deciding whether to take or refuse a bribe, a power-holder is also confronted with the social dilemma of how to properly allocate the obtained resources between himself/herself and the briber, thereby eliciting concern for (un)fairness (Fehr & Schmidt, 1999; Köbis et al., 2016). A plethora of evidence has shown that fairness concern plays a critical role in guiding our social behaviors. For example, people are generally averse to unequal resource distribution during an economic exchange (Fehr & Schmidt, 1999; McAuliffe et al., 2017). While guilt is conceptually linked to fairness concern (i.e., advantageous inequality aversion; Fehr & Schmidt, 1999), it remains an open question as to how guilt-proneness is related to such a concern during corrupt decision-making.

To address these questions, we conducted two online experiments that combined personality measures, incentivized behavioral tasks mimicking bribery-related situations, and computational modeling. In Experiment 1 (N = 1,613), we first examined to what extent guilt proneness would predict bribe-taking behaviors and then explored how the association was moderated by the salience of the harm caused by bribery. In particular, we adopted a between-subject design in which participants in the role of power-holder were randomly assigned to one of the two bribery scenarios varying in harm salience indexed by the explicitness of victimization. In a one-shot bribery game (BG), participants decided whether to accept a bribe and, if so, indicated how much they would request from the briber. Experiment 2 (N = 469) aimed to more rigorously test the role of harm salience in moderating the association between guilt-proneness and bribe-taking behaviors and to further investigate how guilt-proneness is related to different psychological constructs underlying bribe-taking behaviors. To this end, we manipulated harm salience in a multi-round BG by parametrically varying the number of victims affected by bribery and the payoff inequality between the briber and the participants in the proposed bribe. These experimental manipulations allowed us to

quantify these latent psychological constructs during corrupt decision-making by using a computational model with distinct parameters.

Based on previous findings, we developed the following hypotheses focusing on choice behavior. First, participants high in guilt-proneness, compared with those less prone to guilt, would be less likely to take bribes (H1). Second, the negative association of guilt-proneness on bribe-taking behaviors would be stronger when harm in the scenario is more salient (i.e., the explicitness of victimization in Experiment 1; the number of victims in Experiment 2; H2). Last, highly guilt-prone individuals would show stronger concern for others' suffering (characterized by the parameter in the computational model; H3). We performed an exploratory mediation analysis to examine whether this psychological construct serves as a key mechanism through which guilt proneness influences bribe-taking behaviors.

Experiment 1

Method

Participants. A total of 2,765 first-year undergraduate students (1,781 females and 984 males; mean age: 18.7 ± 0.8 years, ranging from 16 to 39 years[1]) at a university in southeast China completed the study as part of the university-led assessment of mental health. Among them, 1,152 participants were excluded because they provided invalid age information ($N = 19$) or failed the comprehension check of the task ($N = 1{,}133$), leaving a sample of $N = 1{,}613$ for further analyses (the valid sample; 1,018 females and 595 males; mean age: 18.7 ± 0.8 years). The sample size was determined by the number of students available on the day of data collection. Note that essentially the same pattern of results was obtained when all participants were included in the analyses (the whole sample, excluding those providing invalid age information, $N = 2{,}746$; see Supplementary Materials for details). Informed consent was obtained from all participants before the experiment. The study (including Experiments 1 and 2) was approved by the ethics committee of the University where the data collection took place.

Procedure and Measures. Experiment 1 consisted of two sessions. The task session included a main task (i.e., the bribe game, BG) that measures corrupt behaviors and a control task (i.e., the modified ultimatum game, UG) that measures participants' baseline fairness concern. The UG was always administered prior to the BG to ensure that participants' baseline fairness concern was not contaminated by their experiences in the BG. The questionnaire session included a series of questionnaires on relevant personality and demographic information. Among the valid sample, 849 participants performed the task session first and then completed the questionnaire session. The remaining 764 participants completed the two sessions in a reversed order.

Both sessions were computerized and presented to participants via an online survey platform (https://www.wjx.cn/). Depending on the room capacity, a group of 30 to 60 participants was tested together in a testing room. They were seated in front of desktops and completed the tasks via computers individually. No talk or interaction was allowed during the test.

One-Shot Bribe Game. Participants were randomly assigned to one of two groups to complete a one-shot bribe game (BG) either with an exam scenario (labeled *Exam* hereafter; $N = 784$) or a donation scenario (labeled *Dona* hereafter; $N = 829$). In both scenarios, participants played the role of the power-holder (framed as arbitrator). They were informed to complete the task together with an independent group of anonymous co-players (framed as undergraduates in another university) who, unbeknownst to participants, were fictitious.

In the *Exam* scenario, participants were informed that these co-players had taken an online "arithmetic test" in which they needed to complete 100 math problems (i.e., addition or subtraction) involving 2 two-digit numbers within 20 min. Only those co-players with an accuracy of 85% or above could earn a reward of 100 tokens (1 token = 0.3 CNY; same below). Those who failed the test would get 0 tokens, but they had a chance to modify their scores and earn the reward by bribing a power-holder. They could share a portion of the reward to persuade the arbitrator to approve of their choices. Note that due to the potential framing effect elicited by the wording in the instruction (Abbink & Hennig-Schmidt, 2006), the word "bribe" was never used throughout the study. Instead, the label "offer" was adopted.

In the *Dona* scenario, participants were told that these co-players would have a chance to keep to themselves a charitable donation of 100 tokens, which should be otherwise donated to a charity (i.e., the United Nations International Children's Emergency Fund; https://www.unicef.org), by bribing a power-holder in a similar way as in the *Exam* scenario. The key difference between the two scenarios lies in harm salience brought by bribery, indexed by the explicitness of the victimization: Keeping the fund for oneself would harm innocent third parties (i.e., children in this case) in the *Dona* scenario, whereas no one would be explicitly harmed in the *Exam* scenario.

Each participant was paired with a unique co-player who bribed and was required to decide whether to accept or reject the bribe from the paired co-player.[2] If accepted, participants needed to further indicate the minimum amount (ranging from 0 to 100) they would request from the co-player. This was important because the bribe would be implemented only when the requested amount was no more than the actual shared amount by the co-player, which was unknown to the participant. If rejected, neither the co-player nor the participants would earn anything.

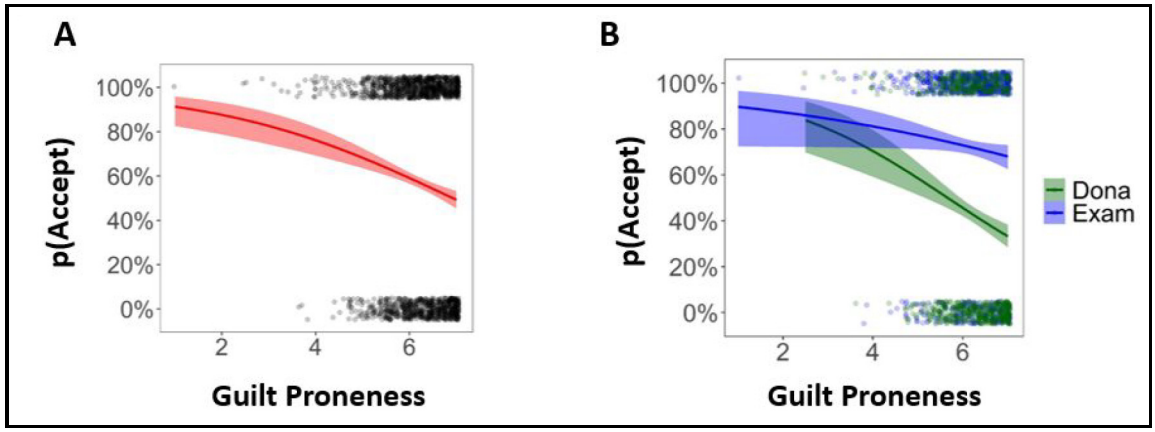

Figure 1. *Results of Choices in Experiment 1. Relationship Between Guilt Proneness and the Probability of Making a Corrupt Choice (i.e., Accepting the Bribe) for Both Scenarios (A) and for Each Scenario Across Participants (B)*
Note. Each dot represents the choice of one participant: dots clustered around 0 (0%) refer to rejection and those clustered around 1 (100%) refer to acceptance. Curves represent the best linear or logistic fits; shaded areas represent ± 1 SEM (For Figure 1B, the purple curve/shaded area and dots indicate the Exam scenario, the green curve/shaded area and dots indicate the Dona scenario). SEM = standard error of the mean.

To incentivize their choices, participants were told that they, together with the paired co-player, would be paid according to the decision they made either in this game or in the UG after the experiment, which was randomly determined. To ensure that they correctly understand the task, they answered comprehension questions before the BG task.

Guilt-Proneness Measures. We administered the Guilt and Shame Proneness scale to measure the guilt-proneness of participants (Cohen et al., 2011). The mean score over these two subscales assessing the negative behavioral evaluations and the repair tendencies to private transgressions was calculated as a measure of the guilt-proneness of each participant. We also measured other personality traits and demographical information, which, together with the baseline fairness concern (as measured by UG), were treated as covariates in data analyses.

Results and Discussion

Statistical analyses and data visualization were performed in *R* (http://www.r-project.org/; R Core Team, 2014). See supplementary material for details about analyses and additional results (Supplemental Tables S1, S2, and S6–S8; also see Supplemental Figures S3–S5).

Choices. To test Hypothesis 1, we performed a robust logistic regression analysis on choice in the BG with *Guilt-Proneness* (continuous variable) and *Scenario* (binary variable: *Exam* as reference level) as the predictor. As predicted, we showed that participants who were high in guilt proneness were less likely to accept the bribe proposed by the co-player regardless of the scenario (Odds Ratio [OR] = 0.67, $b = -0.41$, *Standard Error* [*SE*] = 0.08, $p < .001$, 95% confidence interval [CI]: [–0.57, –0.25]; Figure 1A and Supplemental Figure S1A). Participants were also less likely to accept the bribe in the *Dona* scenario than in the *Exam* scenario (OR = 0.29, b = –1.24, $SE = 0.11$, $p < .001$, 95% CI: [–1.45, –1.03]), suggesting that accepting the bribe in a scenario in which the victimization was more explicit brought an extra moral cost of harming the third party.

Consistent with Hypothesis 2, we found a trend-to-significant *Guilt-Proneness* × *Scenario* (*Dona* vs. *Exam*) interaction effect (Odds Ratio = 0.83, b = –0.19, $SE = 0.11$, $p = .092$, 95% CI: [–0.41, 0.03]; see *online supplemental material* and Supplemental Figure S6 for a note on achieved statistical power) when we incorporated the interaction term. To better understand this interaction trend, we performed simple-slope analyses, showing that highly guilt-prone participants were much less likely to accept a bribe in the *Dona* scenario (OR = 0.60, b = –0.51, $SE = 0.11$, $p < .001$, 95% CI: [–0.73, –0.29]) compared with the *Exam* scenario (OR = 0.79, b = –0.24, $SE = 0.12$, $p = .048$, 95% CI: [–0.47, –0.002]; Figure 1B, also see Supplemental Figures S1B and S2, Tables S3 and S4; see Supplemental Table S5 for effects of *Shame-Proneness*). Post-task subjective rating also showed that the briber's behavior in the *Dona* (vs. *Exam*) scenario was deemed more immoral (–2.7 ± 4.8 vs. –0.9 ± 4.6); $t(1,611) = -8.00$, $p < .001$ (Supplemental Table S2), suggesting a stronger harm salience involved in this scenario. Together, these findings suggest that the inhibitory effect of guilt-proneness on bribe-taking behaviors was stronger when the harm brought by bribery was more salient. However, it should be noted that the degree of harm salience was not directly evaluated and these two scenarios may differ in

other aspects as well. These limitations motivated us to more rigorously examine this moderating effect using a different design in Experiment 2.

Experiment 2

Method

Participants. A total of 847 participants completed Experiment 2 (526 females and 321 males; mean age: 19.2 ± 2.1 years, ranging from 16 to 35 years[3]). The sample size was determined by the number of participants available at the time of data collection. The majority of the participants (˜ 77%) were recruited via the university-led assessment of mental health as in Experiment 1. The remaining participants (˜ 23%) were recruited via posters and individually completed the task on their laptops or mobile devices at their own places. In total, 378 participants were excluded due to their failure in the comprehension checks, leaving a sample of 469 participants for further analyses (the valid sample; 287 females and 182 males; mean age: 19.3 ± 2.1 years, ranging from 17 to 28 years). Note that model-free analyses based on all participants without those providing invalid age information yielded the same pattern of results (the whole sample, $N = 845$).

Procedure and Measures. The structure and procedure of Experiment 2 were essentially the same as Experiment 1, except that all the participants performed the task session first and then completed the questionnaire session.

The multiround BG was the main task of the experiment. It was adapted from the one-shot BG with the *Exam* scenario in Experiment 1. On each trial, participants, acting as the power-holder, were told that they were paired with a unique co-player and received a bribe offer from this co-player and then decided whether to accept or reject this bribe (Supplemental Figure S7). Here, we adopted a parametric design that would allow us to better quantify different psychological processes underlying corrupt decision-making with computational modeling. Specifically, we manipulated harm salience by varying the number of victims (labeled N_{Victim} hereafter; 4 levels: 0, 1, 2, 3) to measure individuals' concern for others' financial suffering. To characterize the fairness concern involved during corrupt decision-making and further improve the external validity of our task, we also manipulated the payoff inequality between the co-player and the participant (labeled *Payoff Inequality*; 4 levels: 10, 20, 30, 40; here, the payoff inequality was always disadvantageous to the participant) and the participant's payoff (labeled *Personal Gain*; 4 levels: 20, 30, 40, 50). All parametric variables (i.e., N_{Victim}, *Payoff Inequality*, *Personal Gains*) were orthogonal and thus yielded 64 unique bribery settings, which were presented only once throughout the task (i.e., 64 trials in total). Similar to Experiment 1, we also conducted a multiround UG prior to the BG to measure participants' baseline fairness concerns.

Data Analyses. We implemented mixed-effect regression analyses on choices in the BG via the *glmer* function in the "lme4" package of *R* (Bates et al., 2013). Furthermore, we applied computational models on choices to formally characterize the psychological constructs underlying corrupt decision-making with distinct parameters and how they were related to guilt-proneness. Model-based analyses were performed using the hierarchical Bayesian approach via *rstan* (Stan Development Team, 2016; https://mc-stan.org/users/interfaces/rstan). See *online supplementary material* for details regarding procedures, computational modeling, and analyses on subjective ratings (Supplemental Figure S8).

Results and Discussion

Choice. Consistent with Experiment 1, participants with higher levels of guilt proneness were less likely to accept a bribe (OR = 0.57, b = –0.56, $SE = 0.22$, $p = .010$, 95% CI: [–0.99, –0.14]), as predicted by Hypothesis 1, and participants were less likely to accept a bribe when the number of victims increased (OR = 0.09, b = –2.38, $SE = 0.13$, $p < .001$, 95% CI: [–2.65, –2.12]), after controlling for *Payoff Inequality* and *Personal Gain*. Consistent with Hypothesis 2, a *Guilt-Proneness* × N_{Victim} interaction was observed on corrupt acts (OR = 0.69, b = –0.38, $SE = 0.15$, $p = .012$, 95% CI: [–0.67, –0.08]; Figure 2A and Supplemental Figure S9; see *online supplementary material* for additional analyses), such that participants who were high (vs. low) in guilt-proneness were less likely to commit bribe-taking choices when the number of victims increased from 0. Notably, all these effects still held after controlling for the covariates (Supplemental Tables S9 and S10; see Supplemental Table S11 for effects of *Shame-Proneness*).

Computational Modeling. Bayesian model comparison showed that the *Inequality-Loss* model (Equation 1) had the lowest leave-one-out information criterion LOOIC scores and outperformed other candidate models (Supplemental Figure S10A).

$$U = P_P - \alpha * (P_B - P_P) - \gamma * (L * N_{Victim}) \qquad (1)$$

Here, U denotes the participant's utility for the given choice (i.e., accepting or rejecting the bribe); P_P and P_B represent the payoff for the participant and the co-player denoted in that choice (same below), respectively. L refers to tokens (i.e., a constant: 40) each victim would lose due to bribe-taking; N_{Victim} represents the number of victims involved in that bribe. Regarding free parameters, α measures the concern for payoff inequality between the participant and the co-player, while γ measures the concern for victims' suffering (Supplemental Table S12). A posterior

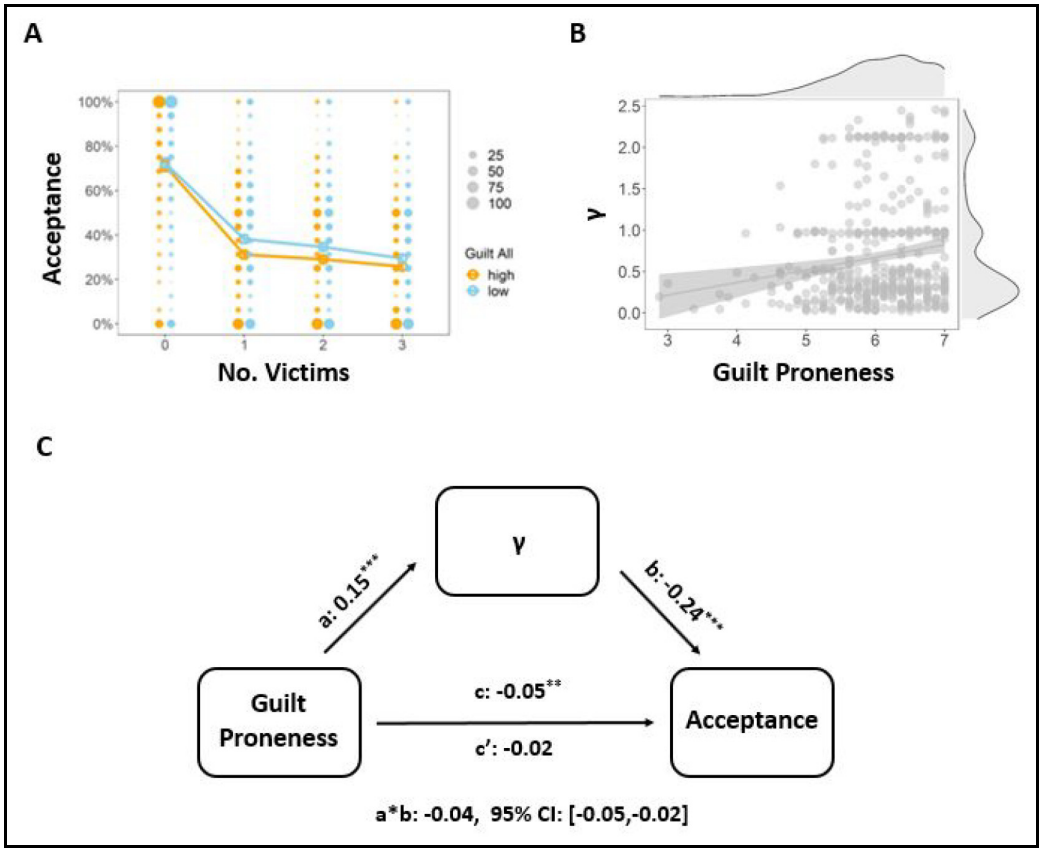

Figure 2. *Main results in Experiment 2. (A) Corrupt Choice (Acceptance Rate) Plotted as a Function of No. Victim and Guilt-Proneness. (B) Relationship Between Guilt-Proneness and γ Across Participants. (C) Mediation Results. γ Fully Mediates the Relationship Between Guilt-Proneness and Bribe-Taking Behaviors (i.e., Overall Acceptance Rates)*

Note. (A) We split participants into two sub-groups based on the median of guilt-proneness score (the orange line and dots indicate individuals with high guilt proneness; the cyan line and dots indicate individuals with low guilt proneness). Circles represent the group-level mean. Error bars represent the SEM. Each dot represents the data of a single participant; the size of the dot represents the number of participants with the same acceptance rate. (B) Each dot represents the choice of a single participant; the line represents the best linear fit; shaded areas represent ± 1 SEM. Density curves in the marginal plots indicate the distribution of each variable respectively (horizon: *Guilt-Proneness*; vertical: γ). (C) The bootstrap approach was used to illustrate that a*b indirect effect was significant after 5,000 bootstraps. SEM = standard error of the mean.

$^{**}p < .01$. $^{***}p < .001$.

predictive check showed that this model could effectively capture actual behaviors in the BG (Supplemental Figure S10B). Thus, the model suggests that while deciding whether to take a bribe, participants not only considered victims' interests but also cared about the payoff inequality between themselves and the co-player, in addition to their own gains. These findings were also aligned with subjective reports of decision-making strategy after the experiment (Supplemental Figure S11).

Given the main goal of Experiment 2, we investigated whether the concern for victims' financial losses due to bribe-taking, as measured by γ, could be selectively predicted by guilt-proneness. Consistent with H3, participants who were high in guilt-proneness displayed a higher γ (robust correlation: $r = 0.15$, $p = .001$; Figure 2B). Importantly, such a correlation was neither observed between shame proneness and γ (robust correlation: $r = 0.02$, $p = .650$) nor between guilt-proneness and α (robust correlation: $r = 0.03$, $p = .520$), suggesting a specific relationship between guilt-proneness and γ. Furthermore, exploratory mediation analysis showed that the positive relationship between guilt-proneness and overall acceptance rate in the BG could be fully mediated by γ (direct effect: –0.017, $p = .252$, 95% CI: [–0.048, 0.010]; indirect effect: –0.036, $p < .001$, 95% CI: [–0.051, –0.020]; Figure 2C and Supplemental Figure S12; also see Supplemental Table S13). This result indicates that the concern for others' suffering serves as a critical psychological mechanism that may underpin the inhibitory modulation of guilt-proneness on bribe-taking behaviors.

General Discussion

Why are some power-holders able to resist the erosion of corruption while others succumb to it? In two online

experiments combining incentivized tasks with computational modeling, we addressed this question by highlighting the crucial role of guilt-proneness—a crucial morally relevant emotion trait—in influencing bribe-taking behaviors of power-holders and its underlying psychological processes.

As predicted by our Hypothesis 1, we found that highly guilt-prone participants, when acting as power-holders, were more likely to reject the bribe in either a one-shot case (Experiment 1) or a multi-round situation (Experiment 2). These findings are consistent with past research demonstrating that guilt-proneness predicts unethical behaviors in various situations, such as dishonest behaviors in a lab setting, counterproductive behaviors in the workplace, and criminal behaviors among jail inmates (Cohen et al., 2011; Tangney et al., 2011).

Our results further revealed a moderating role of harm salience in tuning the effect of guilt-proneness on bribe-taking behaviors, supporting Hypothesis 2. By manipulating harm salience in different ways, we observed that the negative link between guilt-proneness and bribe-taking behaviors was strengthened when harm brought by bribery was more salient. Although ample research has demonstrated that highly guilt-prone individuals are less likely to behave immorally (Cohen et al., 2012), whether such a relationship holds across contexts is not well understood. Our findings suggest that this trait–behavior association is indeed context-sensitive. Given that harm salience is closely related to harm-doer guilt (Berndsen et al., 2004; Zhang et al., 2021), the observed moderation effect demonstrates that highly guilt-prone individuals are more sensitive to increased harm, and their urge to take bribes is inhibited more strongly when the harm involved in bribery becomes more salient. Note that, this finding could help explain the effect of guilt-proneness on a broad range of unethical behaviors varying in interpersonal harm. Such trait–behavior link is tightened when the transgression involves more interpersonal harm (e.g., theft, murder) but reduced when less interpersonal harm is involved (e.g., jaywalking on an empty street).

We additionally observed that the negative effect of the number of victims on bribe-taking behaviors was dramatically reduced for highly guilt-prone individuals once someone is harmed (vs. no one is harmed). This interesting pattern indicates that the relationship between harm salience and bribe-taking behaviors might not be linear, especially for highly guilt-prone individuals, who seem to care more about the presence of the victim but become less sensitive to the cumulative negative consequences. One possible explanation for this pattern is that highly guilt-prone individuals might be inclined to utilize the deontology principle to guide their behaviors (Greene, 2014). In their view, once the bribe-taking behavior harms the other's interests, it is intrinsically wrong regardless of its consequence.

More importantly, our findings enhance the understanding of how guilt-proneness is related to latent psychological constructs underlying corrupt decision-making, which is rarely examined in previous literature. In Experiment 2, we identified a positive association between the concern for others' suffering (indexed by γ) and guilt proneness, as predicted by Hypothesis 3. Moreover, the parameter γ fully mediated the link between guilt-proneness and overall acceptance rates across participants, suggesting that the concern for others' suffering serves as a crucial psychological mechanism through which guilt proneness impacts bribe-taking behaviors. These findings are consistent with previous studies indicating a connection between guilt-proneness and concern for others' suffering (Ent & Baumeister, 2015; Wiltermuth & Cohen, 2014). Nevertheless, unlike these studies that did not formally quantify the concern for others' suffering, the present study adopted a novel task in which harm salience was parametrically manipulated (Experiment 2), allowing us to quantify this latent psychological construct through computational modeling. Compared with the self-report measures, this model-based approach decomposes the psychological processes underpinning social or moral behaviors and reduces the impact of social desirability bias (Crockett, 2016; Konovalov et al., 2018; Yu et al., 2019).

We performed additional analyses to rule out alternative mechanisms that might underpin the association between guilt-proneness and corrupt behaviors and to confirm the specific role of guilt-proneness in predicting bribe-taking behaviors. First, we did not find evidence for the correlation between guilt-proneness and the concern for fairness during corrupt decision-making (indexed by inequality aversion in the BG, α). Here we only considered the bribe disadvantageously unequal to the participants, mainly because this is much more common in real-life bribery situations. Interestingly, by adopting a dimensional approach, Yu and colleagues (2021), in a large-scale online study, showed that the guilt-related trait dimension is predictive of fairness concern in the advantageous unequal context, but not in the disadvantageous context, suggesting context-dependent guilt-proneness–fairness relationship. Future studies should create advantageous unequal bribes to clarify the relationship between guilt-proneness and advantageous inequality aversion during corrupt decision-making.

Second, we did not observe a reliable effect of shame proneness on bribe-taking behaviors. Compared with guilt, shame focuses on a negative evaluation of the global self ("I'm a bad person") rather than a specific behavior ("I did something wrong") and is more "public" by nature. In the current study, participants may focus on the ethical nature of their specific behaviors rather than on their own moral characteristics, and all their decisions were made anonymously, which potentially explains why shame proneness showed no effect here.

The present study sparks interesting questions for future research. First, guilt-proneness is never the only moral-related trait that could predict corrupt behaviors. Indeed, a

recent study has identified a crucial role of the honesty-humility trait and integrity in predicting corrupt acts (Tanner et al., 2022). Therefore, it would be intriguing to investigate whether and how guilt-proneness, together with other personality traits, can serve as a reliable anti-corruption predictor in personnel selection (Cohen et al., 2011, 2012). It is also promising to build anti-corruption tools based on artificial intelligence that utilize machine learning algorithms to identify personality traits reliably predicting real-life corrupt behaviors (Köbis et al., 2022). Second, alternative mechanisms besides the concern for others' suffering might exist through which guilt-proneness impacts corrupt acts. For instance, a recent study showed that interpersonal responsibility is a key mechanism underlying the relationship between guilt-proneness and trustworthiness (Levine et al., 2018). Future research is needed to test the predictive role of more traits and examine other possible mechanisms underlying corrupt decision-making.

Some limitations of the current study should be noted. First, the exclusion rate of the participants was relatively high. To ensure data quality, we applied the strictest criteria that only allowed participants who correctly answered all these questions to be included in subsequent analyses. While our main results still held when the whole sample was used for the analyses, future studies should optimize the procedure to minimize the exclusion rate. Second, we used deception in our task for practical reasons. If applicable, future studies on corruption may consider adopting more realistic interaction tasks.

In summary, the present study reveals a pivotal role of guilt-proneness in inhibiting bribe-taking behaviors that could be committed by the power-holder and how this trait–behavior association is moderated by harm salience. Moreover, this study demonstrates that the concern for others' suffering, a latent psychological construct underlying corrupt decision-making, might serve as a key psychological mechanism through which guilt proneness influences bribe-taking behaviors. Overall, these findings extend our knowledge of the social function of guilt-proneness and provide novel evidence for a deeper understanding of individual differences in corrupt acts.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by research grants from National Natural Science Foundation of China awarded to Y.H. (32200853) and X.Z. (71942001, 31630034), and the Fundamental Research Funds for the Central Universities (2022ECNU-XWK-XK003).

ORCID iDs

Hongbo Yu https://orcid.org/0000-0002-3384-7772
Xiaolin Zhou https://orcid.org/0000-0001-7363-4360

Supplemental Material

The supplemental material is available in the online version of the article.

Notes

1. This was calculated based on 2,746 participants because 19 participants did not provide correct date of birth.
2. We actually asked participants whether they were willing to change the score for the co-player failing the test who bribed (the *Exam* scenario) or to allow the co-player who bribed to keep the charity donation (the *Dona* scenario). Here, choosing "yes" means that they accepted the bribe.
3. The age information was provided based on 845 participants because two participants failed to provide correct date of birth.

References

Abbink, K., & Hennig-Schmidt, H. (2006). Neutral versus loaded instructions in a bribery experiment. *Experimental Economics*, *9*(2), 103–121.

Abbink, K., Irlenbusch, B., & Renner, E. (2002). An experimental bribery game. *Journal of Law, Economics, and Organization*, *18*(2), 428–454.

Barr, A., & Serra, D. (2009). The effects of externalities and framing on bribery in a petty corruption experiment. *Experimental Economics*, *12*(4), 488–503. https://doi.org/10.1007/s10683-009-9225-9

Bates, D., Maechler, M., & Bolker, B. (2013). *lme4: Linear mixed-effects models using S4 classes* (R Package Version 0.999999-0; 2012). http://CRAN.R-project.org/package = lme4

Berndsen, M., van der Pligt, J., Doosje, B., & Manstead, A. (2004). Guilt and regret: The determining role of interpersonal and intrapersonal harm. *Cognition and Emotion*, *18*(1), 55–70.

Cohen, T. R., Panter, A. T., & Turan, N. (2012). Guilt proneness and moral character. *Current Directions in Psychological Science*, *21*(5), 355–359.

Cohen, T. R., Panter, A. T., & Turan, N. (2013). Predicting counterproductive work behavior from guilt proneness. *Journal of Business Ethics*, *114*(1), 45–53.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP Scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, *100*(5), Article 947.

Crockett, M. J. (2016). How formal models can illuminate mechanisms of moral judgment and decision making. *Current Directions in Psychological Science*, *25*(2), 85–90.

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences*, *111*(48), 17320–17325.

Ent, M. R., & Baumeister, R. F. (2015). Individual differences in guilt proneness affect how people respond to moral tradeoffs

between harm avoidance and obedience to authority. *Personality and Individual Differences*, *74*, 231–234.
Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.
Gao, X., Yu, H., Peng, L., Gong, X., Xiang, Y., Jiang, C., & Zhou, X. (2021). The mutuality of social emotions: How the victim's reactive attitude influences the transgressor's emotional responses. *Neuroimage*, *244*, Article 118631.
Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, *101*(2), Article 366.
Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101–124.
Graycar, A., & Smith, R. G. (2011). *Handbook of global research and practice in corruption*. Edward Elgar.
Greene, J. D. (2014). The cognitive neuroscience of moral judgment and decision making. In M. S. Gazzaniga, & G. R. Mangun (Eds.), *The cognitive neurosciences* (pp. 1013–1023). MIT Press.
Koban, L., Corradi-Dell'Acqua, C., & Vuilleumier, P. (2013). Integration of error agency and representation of others' pain in the anterior insula. *Journal of Cognitive Neuroscience*, *25*(2), 258–272.
Köbis, N. C., Starke, C., & Rahwan, I. (2022). The promise and perils of using artificial intelligence to fight corruption. *Nature Machine Intelligence*, *4*, 418–424.
Köbis, N. C., van Prooijen, J.-W., Righetti, F., & Van Lange, P. A. (2016). Prospection in individual and interpersonal corruption dilemmas. *Review of General Psychology*, *20*(1), Article 71.
Konovalov, A., Hu, J., & Ruff, C. C. (2018). Neurocomputational approaches to social behavior. *Current Opinion in Psychology*, *24*, 41–47. https://doi.org/10.1016/j.copsyc.2018.04.009
Lambsdorff, J. G. (2006). Measuring corruption—The validity and precision of subjective indicators (CPI). In C. J. G. Sampford, A. Shacklock, C. Connors, & F. Galtung (Eds.), *Measuring corruption* (pp. 81–100). Ashgate.
Lange, J., Kobis, N. C., & Herrmann, C. (2022). *A systematic literature review and integrative framework on the link between emotions and corruption*. https://osf.io/8uzkf
Levine, E. E., Bitterly, T. B., Cohen, T. R., & Schweitzer, M. E. (2018). Who is trustworthy? Predicting trustworthy intentions and behavior. *Journal of Personality and Social Psychology*, *115*(3), Article 468.
McAuliffe, K., Blake, P. R., Steinbeis, N., & Warneken, F. (2017). The developmental foundations of human fairness. *Nature Human Behaviour*, *1*(2), 1–9.
Mungiu-Pippidi, A., & Heywood, P. M. (2020). *A research agenda for studies of corruption*. Edward Elgar.
R Core Team. (2014). *R: A language and environment for statistical computing*.
Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*(1), 32–70.
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* (pp. 1–12). Emerald Group.
Stan Development Team. (2016). *Stan: A C++ library for probability and sampling*.
Tangney, J. P., Stuewig, J., Mashek, D., & Hastings, M. (2011). Assessing jail inmates' proneness to shame and guilt: Feeling bad about the behavior or the self? *Criminal Justice and Behavior*, *38*(7), 710–734.
Tanner, C., Linder, S., & Sohn, M. (2022). Does moral commitment predict resistance to corruption? Experimental evidence from a bribery game. *PLOS ONE*, *17*(1), Article e0262201.
Thielmann, I., Böhm, R., & Hilbig, B. E. (2021). Buying unethical loyalty: A behavioral paradigm and empirical test. *Social Psychological and Personality Science*, *12*(3), 363–370.
Wiltermuth, S. S., & Cohen, T. R. (2014). "I'd only let you down": Guilt proneness and the avoidance of harmful interdependence. *Journal of Personality and Social Psychology*, *107*(5), 925–942.
Yu, H., Lu, C., Gao, X., Shen, B., Liu, K., Li, W., Xiao, Y., Yang, B., Zhao, X., Crockett, M. J., & Zhou, X. (2021). Explaining individual differences in advantageous inequity aversion by social-affective trait dimensions and family environment. *Social Psychological and Personality Science*, *13*(2), 626–637.
Yu, H., Siegel, J. Z., & Crockett, M. J. (2019). Modeling morality in 3-D: Decision-making, judgment, and inference. *Topics in Cognitive Science*, *11*(2), 409–432.
Zhang, X., Zeelenberg, M., & Breugelmans, S. M. (2021). Anticipated guilt and going against one's self-interest. *Emotion*, *21*(7), 1417–1426.

Author Biographies

Yang Hu is an associate professor in the School of Psychology and Cognitive Science, East China Normal University (Shanghai, China). He received his Ph.D. in psychology in 2017 from University of Bonn (Bonn, Germany) and did his postdoctoral research at CNRS (Lyon, France) and later at Peking University (Beijing, China). Dr. Hu's research interests lie primarily in understanding the cognitive and neural underpinnings of moral cognition using an interdisciplinary approach that combines behavioral economics games, computational modeling, and brain imaging techniques.

Hongbo Yu is an assistant professor in the Department of Psychological and Brain Sciences at the University of California Santa Barbara. He received his Ph.D. in psychology in 2016 from Peking University, China. Dr. Yu did his postdoctoral research first at the University of Oxford, where he was awarded a British Academy Newton International Fellowship, and then at Yale University. Dr. Yu joined the faculty at UCSB and started the Yu Emotion Science (YES) Lab in 2019.

Xiaolin Zhou is the dean and a full professor in the School of Psychology and Cognitive Science, East China Normal University. His research interests cover a range of fields, including social cognition and experimental social psychology, and decision making and neuroeconomics.

Handling Editor: Miyamoto, Yuri