



## The dark side of gratitude: Gratitude could lead to moral violation<sup>☆</sup>

Ruida Zhu<sup>a,b,c,d,1</sup>, Zhenhua Xu<sup>b,c,d,1</sup>, Honghong Tang<sup>a</sup>, Huagen Wang<sup>b,c,d</sup>, Sihui Zhang<sup>b,c,d</sup>, Zhiqi Zhang<sup>b,c,d</sup>, Xiaoqin Mai<sup>e</sup>, Chao Liu<sup>b,c,d,\*</sup>

<sup>a</sup> Business School, Beijing Normal University, 100875 Beijing, China

<sup>b</sup> State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, 100875 Beijing, China

<sup>c</sup> Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, 100875 Beijing, China

<sup>d</sup> Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, 100875 Beijing, China

<sup>e</sup> Department of Psychology, Renmin University of China, 100872 Beijing, China



### ARTICLE INFO

#### Keywords:

Gratitude  
Moral violation  
Relationship promotion  
Harm avoidance

### ABSTRACT

Past research has provided abundant evidence for the positive impacts of gratitude on individuals and society. However, based on the social function of gratitude, which is proposed to be personal relationship promotion, some negative effects of gratitude may exist, especially when the needs of personal relationship promotion and moral norm obedience are in conflict. The current studies investigated whether gratitude leads individuals to violate moral norms for their benefactors and characterized its underlying psychological mechanisms. Scenario-based and laboratory-based results from a series of six studies ( $N = 663$ ) showed that gratitude increases individuals' moral violations (lying and lowering due punishment) when the violation protects their benefactors from harm. Serial multiple mediation analyses and computational models further demonstrated the crucial roles of relationship-building tendency towards the benefactors and harm avoidance on behalf of the benefactors in the gratitude-induced moral violation. These findings deepen the understanding of gratitude and gratitude-induced behaviour.

### 1. Introduction

Known as a moral emotion (McCullough & Kilpatrick, 2001) and 'the parent of other virtue' (Cicero, 1851), gratitude is a universal emotional response to others' kindness (McCullough, 2004). It is suggested that gratitude promotes behaviours that may benefit individuals and the society as a whole. For example, studies have found that gratitude enhances prosociality (Bartlett & DeSteno, 2006; Ma et al., 2017; Tsang, 2006a; Yu et al., 2017; Yu et al., 2018), cooperation (DeSteno et al., 2010), altruistic (third-party) punishment (Vayness et al., 2019), and reduces cheating (DeSteno et al., 2019) and economic impatience (Desteno et al., 2014; Dickens & DeSteno, 2016). Besides, a number of gratitude intervention methods are developed for improving psychological well-being (see two reviews, Davis et al., 2016; Wood et al., 2010) and are given high expectation.

It is proposed that the social function of gratitude in nature is promoting high-quality personal relationship (e.g., find-remind-bind theory) (Algoe, 2012; Algoe et al., 2008). The quality of a relationship

is indicated by three relational features, including the intent of the partner, the cost to the partner in offering the benefit, and the value of the benefit (Forster et al., 2017; Tesser et al., 1968). Gratitude, which is sensitive to these three relational features, helps individuals to find high-quality partners (Algoe, 2012). When a high-quality partner is found, gratitude coordinates individuals' response to the partner, which promotes their relationship and benefits individuals in a long-term (Algoe, 2012). Studies have shown that gratitude boosts self-sacrifice (reflected by bearing the cost of time, money, and energy in prosocial tasks) (Bartlett & DeSteno, 2006; Ma et al., 2017) and facilitates self-control (reflected by decreasing impulsivity in risk decision and dis-favouring immediate reward in intertemporal choice) (Dickens & DeSteno, 2016; Zhang et al., 2020), both of which are conducive to relationship promotion (Desteno et al., 2014; Ma et al., 2017). In other words, gratitude prepares individuals for relationship promotion and leads them to give more weight to personal relationship. This characteristic may consequently result in neglect and devaluation of other objects on the opposite side of relationship promotion. Take the

<sup>☆</sup> This paper has been recommended for acceptance by Giner-Sorolla.

\* Corresponding author at: State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, 100875 Beijing, China.

E-mail address: [liuchao@bnu.edu.cn](mailto:liuchao@bnu.edu.cn) (C. Liu).

<sup>1</sup> Ruida Zhu and Zhenhua Xu contributed equally to this work.

gratitude-promoted prosociality as an example: in such a case, individuals are faced with a trade-off between immediate self-interest (e.g., save money and refuse to help) and long-term beneficial relationship (e.g., spend money and make a friendship). The findings that gratitude promotes prosociality suggest grateful individuals weight the potential personal relationship, while devalue the immediate self-interest (e.g., Ma et al., 2017). As both building relationship with others and sacrificing self-interest for others are praised in most social situations, many existing studies emphasize the positive side of gratitude.

The empirical literature focuses on the positive (or neutral) effects of gratitude, but it is theoretically probable that gratitude causes problems in certain conditions. Consider, for instance, a scenario in which a person's colleague, who is frequently helpful, is absent from work, and the one is asked by the boss where the colleague is. In this case, individuals need to weigh the pros and cons of personal relationship promotion (i.e., cover up the benefactor's absenteeism) and obedience of the norm of honesty (i.e., adhere to honesty). If gratitude here still biases individuals to weight personal relationship as previous studies found, it means that gratitude promotes individuals to violate the moral norm of honesty for their benefactors. Nevertheless, sacred values (e.g., moral norm) and secular values (e.g., monetary rewards) are represented differently in the human brain (Berns et al., 2012; Qu et al., 2019; Vilarroya, 2013). One spending money for their benefactors does not signify that they are willing to violate moral norm for their benefactors as well. Thus, based on existing findings, it is difficult to deduce whether grateful individuals would benefit their benefactors at the cost of violating moral norms.

To fill in the research gap, we investigated whether grateful individuals are willing to violate two widely-accepted moral norms, honesty and justice (Graham et al., 2013), for personal relationship promotion. They are considered to be important moral norms (foundations) across different cultures (Graham et al., 2011). Moreover, we attempted to uncover how gratitude leads to moral violation. Existing studies found that grateful individuals are more likely to conduct various behaviours that promote personal relationship (Algoe, 2012). However, the psychological mechanisms of these gratitude-induced behaviours are relatively unclear. For example, given the function of gratitude is relationship promotion, grateful individuals may have a strong motivation to build relationships with the benefactors, and the motivation specifically facilitates the behaviours that benefit the benefactors, which result in personal relationship promotion. It is also possible that grateful individuals don't have a specific relationship-building motivation in their mind. They unconsciously conduct some behaviours under the influence of gratitude, and the behaviours happen to result in relationship promotion. To answer this question, we explored whether (self-reported) relationship-building tendency towards the benefactors is involved in gratitude-induced violation.

Additionally, we were interested in the causes of relationship-building tendency. Namely, why relationship-building tendency occurs, which may enrich the theory of gratitude's social function. Three relational features, including the intent of the partner, the cost to the partner in offering the benefit, and the value of the benefit, were found to be the antecedents of gratitude (Forster et al., 2017; Tesser et al., 1968). It implicates gratitude tracks the three relational features, each of which indicates a part of relational utility. The relational utility is defined as the utility of a partner for the achievement of individuals' personal aims, such as obtaining benefits (Nelissen, 2014; Ohtsubo & Yagi, 2015; Zhu et al., 2017). For instance, the more valuable the benefit a partner can offer, the higher relational utility would be assigned to the relationship with them. As the function of gratitude is proposed to be identifying high-quality partners (Algoe, 2012; Algoe et al., 2008), a possible cause of relationship-building tendency is the high relational utility of partners. Thus, we explored whether relationship-building tendency correlates with relational utility.

Before the occurrence of moral violation, various psychological processing may have been carried out in violators' mind (Conway &

Gawronski, 2013; Moll & de Oliveira-Souza, 2007). Some of the psychological factors are directly related to the occurrence of moral violation. To understand the mechanistic underpinnings of gratitude-induced violation, we also set out to explore whether gratitude affects the psychological factors associated with moral violation, and whether the effects result in moral violation. Six psychological factors were found to be associated with moral violation: harm aversion (Cushman et al., 2012; Decety & Cowell, 2018), moral judgment (Cushman, 2008), moral principle (Gao et al., 2018; Greene, 2014), concerns for one's own benefits (Haidt, 2003), concerns for the victim's losses (Vitaglione & Barnett, 2003), and anger (Fehr & Schmidt, 1999; Nelissen & Zeelenberg, 2009). More introduction about these psychological factor can be seen in the Supplementary information (SI-1).

Self-report is a common and convenient method for measuring the psychological factors mentioned above. Besides self-report, computational modelling can also achieve the aim. Computational models are designed to estimate latent variables that are not directly observable from behaviour itself (Farrell & Lewandowsky, 2010). With this approach, we are able to probe individuals' psychological processes independent of the self-report method. Consistent findings from different methods can provide more convincing evidence.

In addition to searching the factors related to gratitude and moral violation, we used mediation analyses to examine the processes underlying the effect of gratitude on moral violation. Mediation processes are structured in terms of mediators between an independent variable and a dependent variable (Aglar & De Boeck, 2017). Mediation analysis can be used to test whether and how mediators (e.g., relationship-building tendency and psychological factors associated with violation) are involved in the effect of an independent variable (e.g., gratitude) on a dependent variable (e.g., violation).

In the current studies, we investigated whether gratitude leads individuals to violate moral norms for their benefactors, and explored the associated psychological mechanisms with self-report measures, computational modelling, and mediation analyses. Specifically, in Studies 1A ( $N = 127$ ), 1B ( $N = 142$ ), 2A ( $N = 95$ ), and 2B ( $N = 102$ ), we used hypothetical paradigms to examine how gratitude affects individuals' behaviour when they are facing a trade-off between personal relationship and two different moral norms, honesty (i.e., whether they lie to protect their benefactors in Studies 1A and 1B) and justice (i.e., whether they decrease due punishment to protect their benefactors in Studies 2A and 2B). In Study 3 ( $N = 102$ ), we replicated Study 2B in a laboratory experiment and used computational models to confirm the results based on participants' self-report. In Study 4 ( $N = 95$ ), we tested the generalization of the gratitude-induced violation. Instead of examining whether individuals decrease due punishment to their benefactors for protection (Studies 2A, 2B, and 3), we investigated whether individuals increase due punishment to strangers who unfairly treat the benefactors of the individuals. We report how we determined our sample sizes, all data exclusions, all manipulations, and all measures (except two, see SI-2) in the studies.

## 2. Study 1A

### 2.1. Methods

#### 2.1.1. Participants and design

We expected a moderate effect size of gratitude on violation.<sup>2</sup> Using

<sup>2</sup> At the beginning, we expected Study 1A might has a moderate to large effect size ( $f = 0.35$ ) and recruited 73 valid participants. A reviewer suggested that the sample size appeared lower than present standards. It reminded to us to reappraise Study 1A. We found that Study 1A only had a power of 0.66 ( $< 0.80$ ) with 73 participants. Therefore, we decided to redo a power analysis and enlarge the sample size. Given the actual effect size of the existing Studies 1A, 2A, 2B, and 3 on average was  $f = 0.31$  and a recent meta-analytic review

**Table 1**  
The measured variables, questions, and scale in Study 1.

Measured variable	Question	Scale
Moral violation (lying)	How likely you are to lie to your boss in order to help A.	1 to 9
Harm avoidance	To what extent you want to reduce A's potential losses when answering your boss's question.	1 = not at all, 9 = very much
Moral judgment	To what extent A's behaviour is immoral.	
Moral principle	To what extent you concern for your own honesty when answering your boss's question	
Concerns for one's own benefits	To what extent you concern for your own benefits when answering your boss's question.	

the G\*Power 3 software (Faul et al., 2007), we set the probability of type I error (0.05), expected effect size ( $f = 0.30$ ), and power ( $1 - \beta = 0.80$ ) and determined the minimum sample size to be 90 participants. One hundred and thirty-seven college students participated in the experiment in exchange for monetary payment. Ten participants did not pass an attention check were excluded (Table S1), leaving 127 participants (85 females, 42 males,  $M_{\text{age}} = 22.06$  years,  $SD_{\text{age}} = 2.47$ ; 64 in the gratitude condition, 63 in the control condition) in the analyses. This sample had 80% power to detect effects of  $f > 0.25$  at  $p < .05$  and a 2-tailed test. The study had a between-subject design.

### 2.1.2. Procedure

Participants were required to imagine that they were in a following scenario, which was revised from a vignette developed by Watkins et al. (2006):

*Recently, you got a job in a new company and met a new colleague, A. You and A chatted together occasionally. One day, A asked you what you would be doing next Saturday. You said that you found a new apartment and would be spending the whole day moving. A volunteered to help you move. That Saturday, you saw A outside of your apartment early in the morning, ready to help. A helped you for most of the day until you completed your move. With the help of A, you saved lots of time and effort. Afterwards, you learned the reason why A offered to help.*

In the control condition, participants read 'A hoped that you could undertake some of their work in the future'. In the gratitude condition, participants read 'Having experienced moving house alone, A knew how hard and tiring it was and did not want you to suffer'. The latter description of the intent of A's helping behaviour proved to be efficient in inducing gratitude (Watkins et al., 2006). Afterwards, participants rated how much gratitude, shame, unhappiness, anger, indebtedness, and guilt they felt towards A, to what extent they thought A concerned for their benefits, to what extent they wanted to build a friendship with A, and to what extent they thought a friendship with A would benefit them in the future (1 = not at all, 9 = very much).

The participants then imagined the following situation and answered a series of questions (see Table 1):

*Colleague A, who helped you move before, do not come to work on a Friday afternoon. You know that A is going hiking without leave. Your boss is very angry because of A's absence and asks you why A is absent.*

## 2.2. Results and discussion

### 2.2.1. Manipulation check

Participants in the gratitude condition thought that A concerned for their benefits more ( $M = 6.77$ ,  $SD = 1.31$ ) and had higher gratitude ratings ( $M = 8.38$ ,  $SD = 1.09$ ) than the control condition (concern for

(footnote continued)

indicated that the effect size of gratitude was moderate (Ma et al., 2017), we set the expected effect size  $f$  to be 0.30, determined the minimum sample size to be 90 participants, and increased the sample size to be 127. The additional samples did not change the main statistical results of Study 1A.

their benefits:  $M = 4.19$ ,  $SD = 1.67$ ,  $F(1,125) = 93.59$ ,  $p < .001$ , partial  $\eta^2 = 0.428$ ; gratitude ratings:  $M = 4.52$ ,  $SD = 1.93$ ,  $F(1,125) = 191.90$ ,  $p < .001$ , partial  $\eta^2 = 0.606$ ) (Fig. 1A). In the gratitude condition, the gratitude ratings were significantly higher than the ratings of other emotions (all  $F_s > 351.02$ , all  $p_s < .001$ , all partial  $\eta^2_s > 0.848$ ) (Table S2). The manipulation check revealed that our manipulation of the participants' perceptions of A's intent and resulting grateful feelings was successful. The ratings of other emotions can be seen in the Supplementary information (SI-4).

### 2.2.2. Relationship-building tendency and relational utility

Compared with participants in the control condition (relationship-building tendency:  $M = 4.14$ ,  $SD = 1.69$ ; relational utility:  $M = 4.94$ ,  $SD = 1.70$ ), participants in the gratitude condition were more inclined to build a friendship with A ( $M = 7.94$ ,  $SD = 0.99$ ,  $F(1,125) = 238.63$ ,  $p < .001$ , partial  $\eta^2 = 0.656$ , Fig. 1A) and considered a friendship with A to be more beneficial ( $M = 7.34$ ,  $SD = 1.14$ ,  $F(1,125) = 87.71$ ,  $p < .001$ , partial  $\eta^2 = 0.412$ ). A significant positive correlation was found between relationship-building tendency and relational utility across all participants ( $r = 0.82$ ,  $p < .001$ ,  $N = 127$ ) (Fig. S1). The results suggest that relationship-building tendency emerges with gratitude and that the increased relationship-building tendency could be on account of the perceived high relational utility.

### 2.2.3. Moral violation (lying)

Compared to participants in the control condition ( $M = 5.27$ ,  $SD = 2.20$ ), participants in the gratitude condition were more likely to lie to their boss ( $M = 6.33$ ,  $SD = 1.89$ ;  $F(1,125) = 8.47$ ,  $p = .004$ , partial  $\eta^2 = 0.063$ ) (Fig. 1B). Grateful individuals do tend to benefit their benefactors at the cost of violating the norm of honesty.

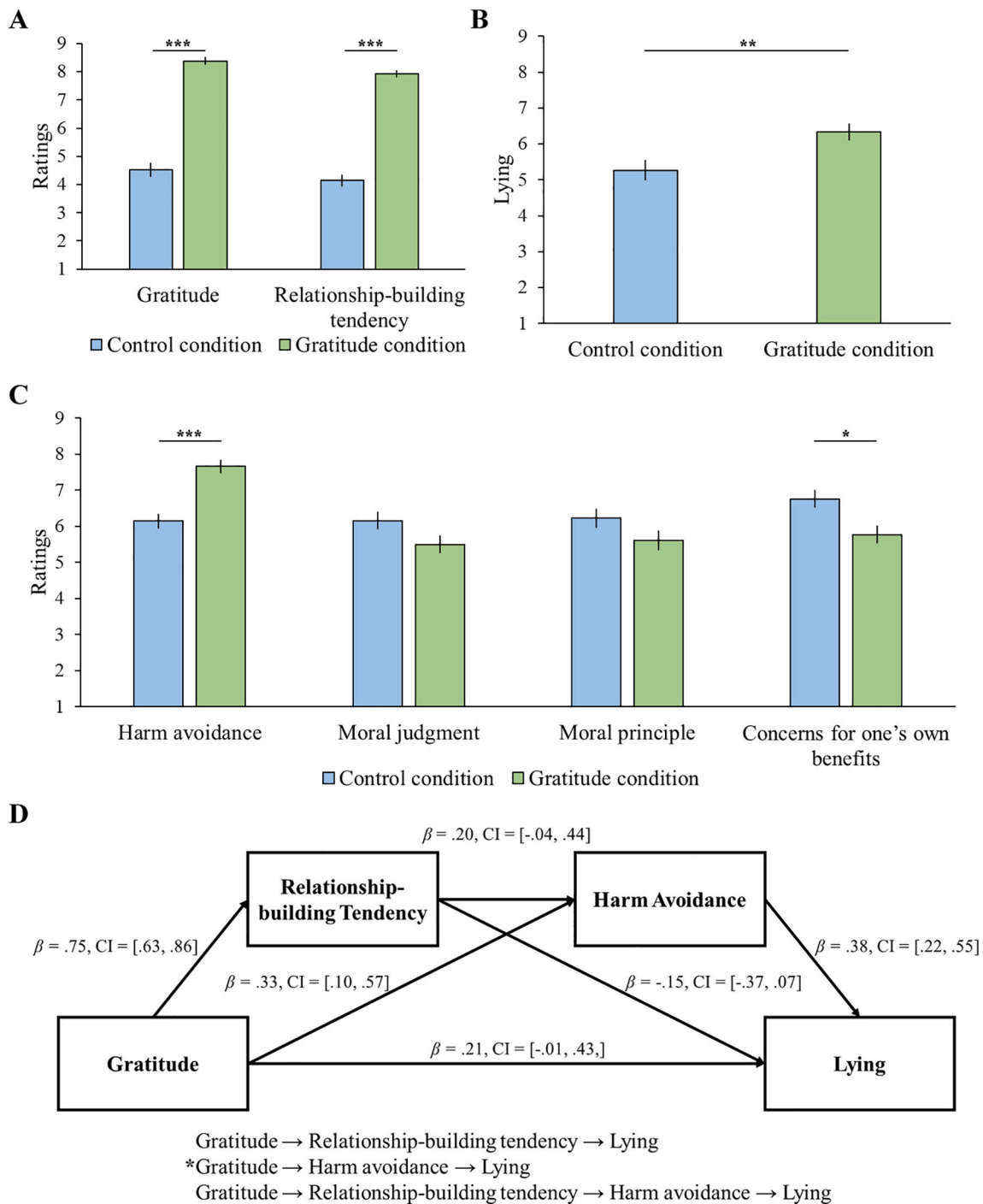
### 2.2.4. Simple mediation

Based on the social function of gratitude (e.g., find-remind-bind theory) (Algoe, 2012), an intuitive hypothesis is that relationship-building tendency mediates the effect of gratitude on violation. To test this hypothesis, we used SPSS and PROCESS macro software (<http://www.processmacro.org/index.html>) for data analysis (Preacher & Hayes, 2008). The data from the gratitude and control conditions were combined and the continuous variables (ratings of gratitude, friendship-building tendency, and lying) were standardized (z-scored). A simple mediation model was employed. A 5000 bootstrap resample procedure was implemented to obtain 95% confidence intervals (CI) for making statistical inference about indirect effects. Standardized coefficients ( $\beta$ ) were reported. The analysis strategy remained the same across all studies.

The results showed that the direct effect of gratitude on lying was significant ( $\beta = 0.35$ , CI = [0.08, 0.61], as the CI did not cover zero). However, the indirect effect of gratitude on lying through relationship-building tendency was not significant ( $\beta = 0.00$ , CI = [-0.17, 0.22]).

### 2.2.5. Psychological factors associated with violation

We explored the differences between the gratitude and control groups in each of the psychological factors associated with violations. We corrected for the statistical tests by adjusting  $p$  values using Bonferroni correction across all studies. Participants in the gratitude condition had significantly stronger motivation to reduce A's potential



**Fig. 1.** Results of Study 1A. A) Mean gratitude ratings and relationship-building tendency ( $\pm SE$ ) in the gratitude and control conditions.  $***p < .001$ . B) Mean lying ( $\pm SE$ ) in the gratitude and control conditions.  $**p < .01$ . C) Mean ratings of psychological factors associated with violation ( $\pm SE$ ) in the gratitude and control conditions.  $*p_{corrected} < .05$ ,  $***p_{corrected} < .001$ . D) There was a trend that the indirect effect of gratitude on lying through relationship-building tendency and harm avoidance in sequence was towards significance. The indirect effect of gratitude on lying through harm avoidance alone was significant. Significant indirect effect is indicated by symbol \*. CI represents 95% confidence interval.

loss ( $F(1,125) = 29.31$ ,  $p_{corrected} < .001$ , partial  $\eta^2 = 0.190$ ) and less concerns for their own benefits ( $F(1,125) = 8.40$ ,  $p_{corrected} = .018$ , partial  $\eta^2 = 0.063$ ) than participants in the control condition (Fig. 1C). There was no significant difference in the ratings of moral judgment ( $F(1,125) = 3.63$ ,  $p_{corrected} = .236$ , partial  $\eta^2 = 0.028$ ) and moral principle ( $F(1,125) = 2.66$ ,  $p_{corrected} = .421$ , partial  $\eta^2 = 0.021$ ).

We also explored whether each of the psychological factors correlates with relationship-building tendency. We used Bonferroni correction for controlling alpha inflation across all studies. Two significant

correlations were found between these psychological factors and relationship-building tendency, including harm avoidance ( $r = 0.51$ ,  $p_{corrected} < .001$ ,  $N = 127$ ) and concerns for one's own benefits ( $r = -0.26$ ,  $p_{corrected} = .013$ ,  $N = 127$ ). There was no significant correlation between relationship-building tendency and moral judgment ( $r = -0.18$ ,  $p_{corrected} = .165$ ,  $N = 127$ ) or moral principle ( $r = -0.17$ ,  $p_{corrected} = .253$ ,  $N = 127$ ).

### 2.2.6. Serial multiple mediation

Given the significant correlations between relationship-building tendency and psychological factors associated with violation, it is possible that an indirect effect of gratitude on lying is achieved through relationship-building tendency and a psychological factor associated with violation in sequence (i.e., gratitude → relationship-building tendency → a key psychological factor associated with violation → violation). For testing this possibility, similar preparation procedures in the simple mediation analysis were conducted and a serial multiple mediation model was employed. Relationship-building tendency was entered into the model as the first mediator. Each psychological factor associated with violation was entered into the model as the second mediator at a time. In the meanwhile, the other psychological factors associated with violation were involved in the model as covariates. It was for observing the unique effect of each psychological factor, considering that overlaps might exist among the psychological factors. The analysis strategy remained the same across all studies.

When harm avoidance was entered into the model as the second mediator, the direct effect of gratitude on lying was not significant ( $\beta = 0.21$ , 95% CI = [-0.01, 0.43]). The indirect effect of gratitude on lying through relationship-building tendency alone was not significant ( $\beta = -0.11$ , 95% CI = [-0.25, 0.05]) and the indirect effect of gratitude on lying through harm avoidance alone was significant ( $\beta = 0.13$ , 95% CI = [0.02, 0.25]). Although strictly speaking the indirect effect of gratitude on lying through relationship-building tendency and harm avoidance in sequence was not significant ( $\beta = 0.06$ , 95% CI = [-0.01, 0.14]), a trend towards significance existed. The path from gratitude to relationship-building tendency ( $\beta = 0.75$ , CI = [0.63, 0.86],  $p < .001$ ) and the path from harm avoidance to lying ( $\beta = 0.38$ , CI = [0.22, 0.55],  $p < .001$ ) revealed significant effects, while the path from relationship-building tendency to harm avoidance revealed a marginally significant effect ( $\beta = 0.20$ , CI = [-0.04, 0.44],  $p = .099$ ) (Fig. 1D).

When any other psychological factor (namely moral judgment, moral principle, or concerns for one's own benefits) was entered into the model as the second mediator, the indirect effect of gratitude on lying was not mediated through relationship-building tendency alone, through any psychological factor alone, or through relationship-building tendency and any psychological factor in sequence (Table S6).

In Study 1A, our results show that gratitude encourages individuals to benefit the benefactors by telling a lie. The effect of gratitude on lying may be achieved through a serial multiple mediation (gratitude → relationship-building tendency → harm avoidance → lying), in which relationship-building tendency and harm avoidance play important roles. In the Study 1B, we tested whether the findings of Study 1A could be replicated by running a preregistered study.

## 3. Study 1B

### 3.1. Methods

#### 3.1.1. Participants and design

We used this preregistered study (<https://osf.io/ck3a2>) to test whether the findings of Study 1A could be replicated. We expected the effect size of gratitude on violation in Study 1B would be the same as Study 1A ( $f = 0.26$ ). We set the probability of type I error (0.05), expected effect size ( $f = 0.26$ ), and power ( $1 - \beta = 0.80$ ) and determined the minimum sample size to be 118 participants. One hundred and forty-four participants took part in the experiment in exchange for monetary payment. Two participants did not pass an attention check were excluded (Table S1), leaving 142 participants<sup>3</sup> (96 females, 46

males,  $M_{\text{age}} = 25.46$  years,  $SD_{\text{age}} = 7.78$ ; 71 in the gratitude condition, 71 in the control condition) in the analyses. This sample had 80% power to detect effects of  $f > 0.24$  at  $p < .05$  and a 2-tailed test. The study had a between-subject design.

#### 3.1.2. Procedure

The procedure of Study 1B was the same as Study 1A.

### 3.2. Results and discussion

#### 3.2.1. Manipulation check

Participants in the gratitude condition thought that A concerned for their benefits more ( $M = 6.76$ ,  $SD = 1.37$ ) and had higher gratitude ratings ( $M = 8.11$ ,  $SD = 1.39$ ) than the control condition (concern for their benefits:  $M = 4.04$ ,  $SD = 1.96$ ,  $F(1,140) = 91.86$ ,  $p < .001$ , partial  $\eta^2 = 0.396$ ; gratitude ratings:  $M = 4.83$ ,  $SD = 2.07$ ,  $F(1,140) = 123.03$ ,  $p < .001$ , partial  $\eta^2 = 0.468$ ) (Fig. 2A). In the gratitude condition, the gratitude ratings were significantly higher than the ratings of other emotions (all  $F_s > 229.34$ , all  $p_s < .001$ , all partial  $\eta^2_s > 0.766$ ) (Table S2). The manipulation check revealed that our manipulation of the participants' perceptions of A's intent and resulting grateful feelings was successful.

#### 3.2.2. Relationship-building tendency and relational utility

Compared with participants in the control condition (relationship-building tendency:  $M = 4.70$ ,  $SD = 2.16$ ; relational utility:  $M = 5.61$ ,  $SD = 1.73$ ), participants in the gratitude condition were more inclined to build a friendship with A ( $M = 8.01$ ,  $SD = 1.18$ ,  $F(1,140) = 128.48$ ,  $p < .001$ , partial  $\eta^2 = 0.479$ , Fig. 2A) and considered a friendship with A to be more beneficial ( $M = 7.51$ ,  $SD = 1.27$ ,  $F(1,140) = 55.68$ ,  $p < .001$ , partial  $\eta^2 = 0.285$ ). A significant positive correlation was found between relationship-building tendency and relational utility across all participants ( $r = 0.81$ ,  $p < .001$ ,  $N = 142$ ) (Fig. S1).

#### 3.2.3. Moral violation (lying)

Compared to participants in the control condition ( $M = 5.24$ ,  $SD = 2.26$ ), participants in the gratitude condition were more likely to lie to their boss ( $M = 6.52$ ,  $SD = 1.58$ ;  $F(1,140) = 15.39$ ,  $p < .001$ , partial  $\eta^2 = 0.099$ ) (Fig. 2B).

#### 3.2.4. Simple mediation

The results showed that the direct effect of gratitude on lying was not significant ( $\beta = 0.05$ , CI = [-0.20, 0.31]). The indirect effect of gratitude on lying through relationship-building tendency was significant ( $\beta = 0.31$ , CI = [0.10, 0.56]).

#### 3.2.5. Psychological factors associated with violation

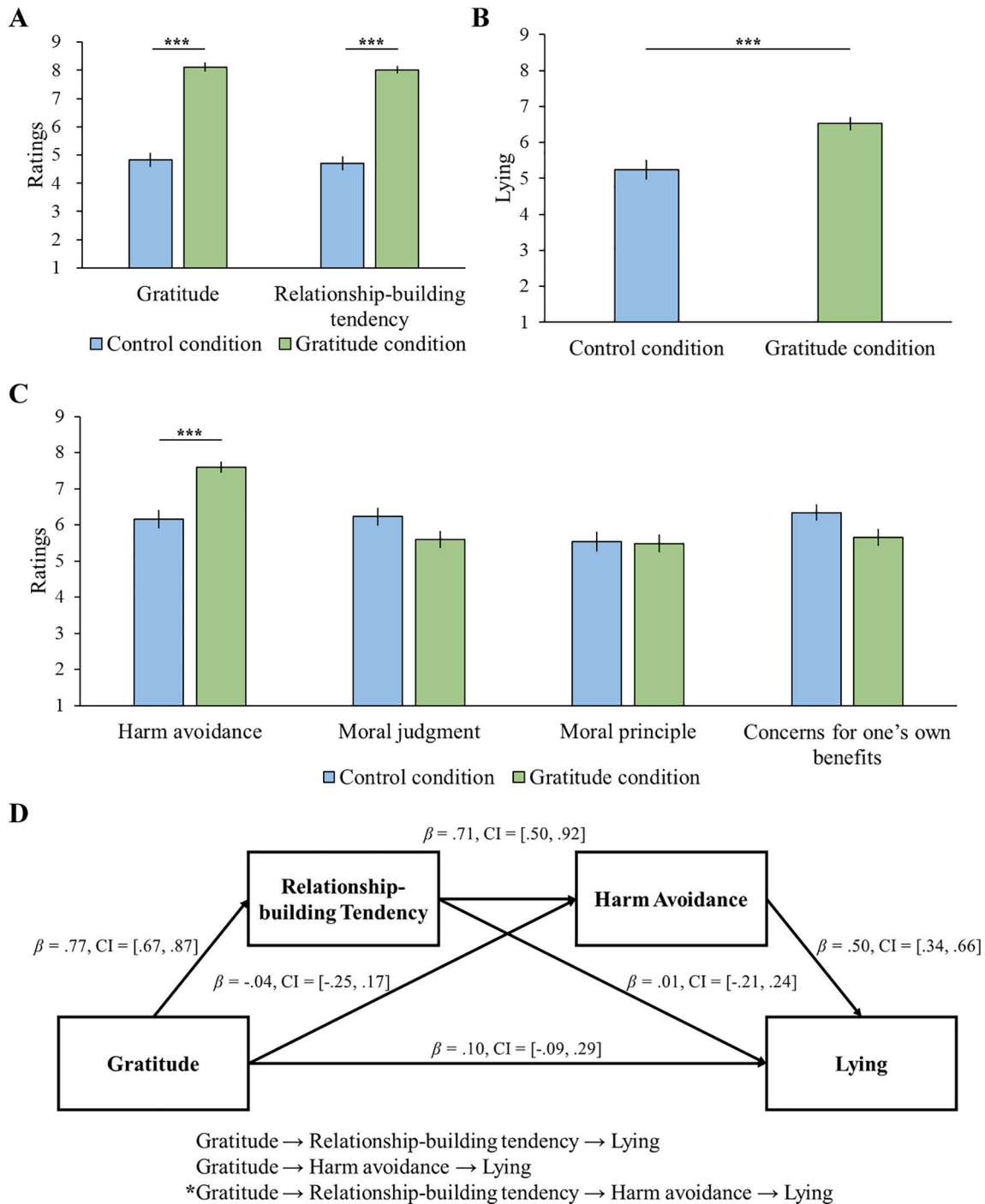
Participants in the gratitude condition had significantly stronger motivation to reduce A's potential loss ( $F(1,140) = 24.51$ ,  $p_{\text{corrected}} < .001$ , partial  $\eta^2 = 0.149$ ) (Fig. 2C). There was no significant difference in the ratings of moral judgment ( $F(1,140) = 3.51$ ,  $p_{\text{corrected}} = .253$ , partial  $\eta^2 = 0.024$ ), moral principle ( $F(1,140) = 0.02$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 < 0.001$ ), and concerns for one's own benefits ( $F(1,140) = 4.57$ ,  $p_{\text{corrected}} = .137$ , partial  $\eta^2 = 0.032$ ).

Three significant correlations were found between these psychological factors and relationship-building tendency, including harm avoidance ( $r = 0.67$ ,  $p_{\text{corrected}} < .001$ ,  $N = 142$ ), moral judgment ( $r = -0.22$ ,  $p_{\text{corrected}} = .031$ ,  $N = 142$ ) and concerns for one's own benefits ( $r = -0.22$ ,  $p_{\text{corrected}} = .041$ ,  $N = 142$ ). There was no significant correlation between relationship-building tendency and moral principle ( $r = -0.01$ ,  $p_{\text{corrected}} > .999$ ,  $N = 142$ ).

(footnote continued)

wouldn't change, no matter these two participants were involved in the analyses or not.

<sup>3</sup>As we wrote in the preregistration (<https://osf.io/ck3a2>), we planned to recruit 140 valid participants. But we finally had 142. Two extra participants were accidentally recruited due to a technical issue. The statistical results



**Fig. 2.** Results of Study 1B. A) Mean gratitude ratings and relationship-building tendency ( $\pm SE$ ) in the gratitude and control conditions.  $***p < .001$ . B) Mean lying ( $\pm SE$ ) in the gratitude and control conditions.  $***p < .001$ . C) Mean ratings of psychological factors associated with violation ( $\pm SE$ ) in the gratitude and control conditions.  $***p_{corrected} < .001$ . D) The indirect effect of gratitude on lying through relationship-building tendency and harm avoidance in sequence was significant. Significant indirect effect is indicated by symbol \*. CI represents 95% confidence interval.

### 3.2.6. Serial multiple mediation

When harm avoidance was entered into the model as the second mediator, the direct effect of gratitude on lying was not significant ( $\beta = 0.10$ , 95% CI =  $[-0.09, 0.29]$ ). The indirect effect of gratitude on lying through relationship-building tendency alone ( $\beta = 0.01$ , 95% CI =  $[-0.19, 0.17]$ ) or through harm avoidance alone ( $\beta = -0.02$ , 95% CI =  $[-0.14, 0.11]$ ) was not significant. The indirect effect of gratitude on lying through relationship-building tendency and harm avoidance in sequence was significant ( $\beta = 0.27$ , 95% CI =  $[0.15, 0.46]$ ) (Fig. 2D).

When any other psychological factor (namely moral judgment, moral principle, or concerns for one's own benefits) was entered into the model as the second mediator, the indirect effect of gratitude on lying was not mediated through relationship-building tendency alone, through any psychological factor alone, or through relationship-building tendency and any psychological factor in sequence (Table S7).

Consistent with Study 1A, Study 1B reveals that gratitude promotes individuals to benefit the benefactors by telling a lie. The relationship-building tendency and harm avoidance play roles in the gratitude-induced lying. In the following Study 2A, we tested whether gratitude-

induced violation can be observed in another moral domain (i.e., justice).

## 4. Study 2A

### 4.1. Methods

#### 4.1.1. Participants and design

We set the probability of type I error (0.05), expected effect size ( $f = 0.30$ ), and power ( $1 - \beta = 0.80$ ) and determined the minimum sample size to be 90 participants. One hundred and forty-two college students participated in the experiment in exchange for monetary payment. Of these, 47 participants who incorrectly responded on either of the two comprehension tests of the instructions or failed an attention check were excluded (Table S1), leaving 95 participants (57 females, 38 males,  $M_{\text{age}} = 22.26$  years,  $SD_{\text{age}} = 3.32$ ; 53 in the gratitude condition, 42 in the control condition) in the analyses. This sample had 80% power to detect effects of  $f > 0.29$  at  $p < .05$  and a 2-tailed test.

#### 4.1.2. Procedure

Participants imagined that they were in the following scenario, the development of which was inspired by a study on gratitude (Yu et al., 2017):

*You participated in an experiment and met a stranger, A. You and A were required to complete a choice game together. The rules of the game were as follow: the game contained 20 trials. In each trial, there were two cards, a grey and a yellow card. At the beginning, you saw the backs of the two cards and had to choose one of them. After you made your decision, the two cards were turned over. If the yellow card was chosen, your wrist might receive a strong, painful electric shock (within safe levels). If the grey card was chosen, nothing happened. Stranger A could see your choice and the corresponding outcome.*

In the gratitude condition, participants read:

*If you chose a yellow card in a trial, A was faced with the decision of whether to help you or not. If A chose to help, you and A each received a moderate electric shock, which caused moderate pain. Otherwise, you received a strong electric shock and nothing happened to A. After the game, only one trial was randomly chosen and executed. The ultimate results of the game turned out to be that you chose the grey card and the yellow card each ten times. Whenever you chose the yellow card, A decided to help you (i.e., A decided to help you ten times). In the chosen and executed trial, you chose a yellow card and A decided to help you.*

In the control condition, participants read:

*When you chose the yellow card in a trial, a computer program decided whether to require A to help you based on an algorithm. If the program required A to help, you and A each received a moderate electric shock, causing moderate pain. Otherwise, you received a strong electric shock and nothing happened to A. After the game, only one trial was randomly chosen and executed. The ultimate results of the game turned out to be that you chose the grey card and the yellow card each ten times. In the chosen and executed trial, you chose a yellow card and the program required A to help you.*

After the induction of emotion above, participants answered the same questions as those in the Study 1B, except that one additional question (i.e., to what extent the participant believed that A was willing to help them) was added. They also completed a comprehension test regarding the choice game.

Then participants imagined that they played a third-party punishment game (TPP) and completed the corresponding questions (Table 2) and a comprehension test about the TPP. The rules of the TPP were described as follow:

*After having received the electric shock, you and A participated in*

*another game, called a decision game, together with four strangers: B, C, D and E. In this game, there were three players: an allocator, a receiver, and a decider. In each trial, the allocator freely divided 100 Chinese yuan between themselves and the receiver. The receiver had to accept the division. Afterwards, the decider, who obtained 50 Chinese yuan in each trial, had a chance to reduce the allocator's money by spending their own money: each yuan spent deducted 3 yuan from the allocator's payoff. The game contained four trials. In all four trials, you and A respectively played as the decider and the allocator. B, C, D, and E each played the game only once as the receiver. In the first stage, A completed the four distributions between B, C, D, E, and themselves. In the second stage, you saw A's distributions one by one and decided how much money to spend in reducing A's payoff.*

### 4.2. Results and discussion

#### 4.2.1. Manipulation check

Participants in the gratitude condition thought that A concerned for their benefits more ( $M = 6.09$ ,  $SD = 1.86$ ), and that A was more willing to help them ( $M = 6.02$ ,  $SD = 1.79$ ), and had higher gratitude ratings ( $M = 7.55$ ,  $SD = 1.86$ ) than participants in the control condition (concern for their benefits:  $M = 4.00$ ,  $SD = 2.02$ ,  $F(1,93) = 27.43$ ,  $p < .001$ , partial  $\eta^2 = 0.228$ ; willingness to help:  $M = 3.48$ ,  $SD = 2.03$ ,  $F(1,93) = 42.00$ ,  $p < .001$ , partial  $\eta^2 = 0.311$ ; gratitude ratings:  $M = 6.38$ ,  $SD = 2.19$ ,  $F(1,93) = 7.90$ ,  $p = .006$ , partial  $\eta^2 = 0.078$ ) (Fig. 3A). In the gratitude condition, the gratitude ratings were significantly higher than the ratings of other emotions (all  $F_s > 59.06$ , all  $p_s < .001$ , all partial  $\eta^2_s > 0.532$ ) (Table S3). The manipulation checks revealed that our manipulation of the participants' perceptions of A's intent and resulting grateful feelings was successful.

#### 4.2.2. Relationship-building tendency and relational utility

Compared with participants in the control condition (relationship-building tendency:  $M = 6.33$ ,  $SD = 1.93$ ; relational utility:  $M = 6.12$ ,  $SD = 1.77$ ), participants in the gratitude condition were more inclined to build a friendship with A ( $M = 7.68$ ,  $SD = 1.37$ ,  $F(1,93) = 15.73$ ,  $p < .001$ , partial  $\eta^2 = 0.145$ ) (Fig. 3A) and considered a friendship with A to be more beneficial ( $M = 7.53$ ,  $SD = 1.20$ ,  $F(1,93) = 21.26$ ,  $p < .001$ , partial  $\eta^2 = 0.186$ ). A significant positive correlation was found between relationship-building tendency and relational utility across all participants ( $r = 0.73$ ,  $p < .001$ ,  $N = 95$ ) (Fig. S1).

#### 4.2.3. Moral violation (decreased punishment)

Participants in the gratitude condition ( $M = 17.46$ ,  $SD = 20.35$ ) deducted the allocator less money than participants in the control condition did ( $M = 29.84$ ,  $SD = 27.52$ ;  $F(1,93) = 6.35$ ,  $p = .013$ , partial  $\eta^2 = 0.064$ ) (Fig. 3B).

#### 4.2.4. Simple mediation

The results showed that the direct effect of gratitude on punishment was not significant ( $\beta = -0.07$ ,  $CI = [-0.29, 0.15]$ ). The indirect effect of gratitude on punishment through relationship-building tendency was not significant ( $\beta = 0.02$ ,  $CI = [-0.05, 0.10]$ ).

#### 4.2.5. Psychological factors associated with violation

Participants in the gratitude condition showed stronger motivation to reduce A's potential loss than participants in the control condition ( $F(1,93) = 7.58$ ,  $p_{\text{corrected}} = .036$ , partial  $\eta^2 = 0.075$ ) (Fig. 3C). There was no significant difference in the ratings of moral judgment ( $F(1,93) = 1.53$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 = 0.016$ ), moral principle ( $F(1,93) = 0.40$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 = 0.004$ ), concerns for one's own benefits ( $F(1,93) = 0.26$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 = 0.003$ ), or concerns for receivers' losses ( $F(1,93) = 0.51$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 = 0.005$ ) between the two conditions.

A significant correlation was found between harm avoidance and

**Table 2**  
The measured variables, questions, and scale in Study 2A.

Measured variable	Question	Scale
Moral violation (decreased punishment)	How much money you decide to spend in deducting A's money (the four distributions of A were 100:0 (100 Chinese yuan for A and 0 for the receiver), 90:10, 80:20, and 70:30; the question was asked after each distribution was presented).	0–50 yuan
Harm avoidance	To what extent you want to reduce A's losses when making decisions.	1–9
Moral judgment	To what extent A's behaviour is immoral (this question was asked after each of A's distribution was presented).	1 = not at all, 9 = very much
Moral principle	To what extent you concern for the justice of your behaviour when making decisions.	
Concerns for one's own benefits	To what extent you concern for your own benefits when making decisions.	
Concerns for the receivers' losses	To what extent you concern for the receivers' losses when making decisions.	

Note: we averaged the punishment the allocators received (the amount of money participants spent multiplied by three) and participants' moral judgment ratings in four trials to indicate moral violation and moral judgment, respectively.

relationship-building tendency ( $r = 0.40$ ,  $p_{\text{corrected}} < .001$ ,  $N = 95$ ). There was no significant correlation between relationship-building tendency and moral judgment ( $r = -0.00$ ,  $p_{\text{corrected}} > .999$ ,  $N = 95$ ), moral principle ( $r = -0.02$ ,  $p_{\text{corrected}} > .999$ ,  $N = 95$ ), concerns for one's own benefits ( $r = -0.06$ ,  $p_{\text{corrected}} > .999$ ,  $N = 95$ ), or concerns for receivers' losses ( $r = 0.01$ ,  $p_{\text{corrected}} > .999$ ,  $N = 95$ ).

#### 4.2.6. Serial multiple mediation

When harm avoidance was entered into the model as the second mediator, the direct effect of gratitude on violation was not significant ( $\beta = -0.11$ , 95% CI =  $[-0.28, 0.05]$ ). The indirect effect of gratitude on punishment through relationship-building tendency alone ( $\beta = 0.06$ , 95% CI =  $[-0.00, 0.15]$ ) or through harm avoidance alone ( $\beta = -0.06$ , 95% CI =  $[-0.14, 0.00]$ ) was not significant. Importantly, the indirect effect of gratitude on punishment through relationship-building tendency and harm avoidance in sequence was significant ( $\beta = -0.03$ , 95% CI =  $[-0.08, -0.00]$ ) (Fig. 3D).

When any other psychological factor (namely moral judgment, moral principle, concerns for one's own benefits, or concerns for receivers' losses) was entered into the model as the second mediator, the indirect effect of gratitude on punishment was not mediated through relationship-building tendency alone, through any psychological factor alone, or through relationship-building tendency and any psychological factor in sequence (Table S8).

In Study 2A, we observed gratitude-induced violation in another moral domain (i.e., justice). Gratitude promotes individuals to benefit the benefactors by lowering due punishment. The effect of gratitude on punishment is achieved through a serial multiple mediation (gratitude  $\rightarrow$  relationship-building tendency  $\rightarrow$  harm avoidance  $\rightarrow$  decreased punishment), in which relationship-building tendency and harm avoidance serve as important nodes. A limitation of Study 2A was that the rules of the choice game were complicated. Therefore, we simplified them in Study 2B and examined whether the findings could be replicated. In addition, participants' anger towards A's unfair allocations was measured, considering that anger is a psychological drive of punishment (Fehr & Gächter, 2002; Nelissen & Zeelenberg, 2009).

## 5. Study 2B

### 5.1. Methods

#### 5.1.1. Participants and design

Same as Study 2A, the determined the minimum sample size was 90 participants. One hundred and thirty-two college students participated in the experiment in exchange for course credits. Of these, 30 participants who gave any incorrect answer during the two comprehension tests of the instructions or failed an attention check were excluded (Table S1), leaving 102 participants (63 females, 39 males,  $M_{\text{age}} = 20.37$  years,  $SD_{\text{age}} = 2.67$ ; 52 in the gratitude condition, 50 in the control condition) in following analyses. This sample had 80%

power to detect effects of  $f > 0.28$  at  $p < .05$  and a 2-tailed test.

#### 5.1.2. Procedure

Study 2B was same as Study 2A, except for two changes.<sup>4</sup> Firstly, the rules of the choice game were revised to be less complicated. Participants imagined that the electric shock would be immediately implemented according to the outcome at the end of each trial, rather than only one trial being chosen and executed at the end of the experiment. Besides, in the control condition, the rules required A to help the participants whenever the participants chose a yellow card. Secondly, we asked participants to report their anger levels when they saw A's distributions during the TPP.

## 5.2. Results and discussion

### 5.2.1. Manipulation check

Participants in the gratitude condition thought that A concerned for their benefits more ( $M = 6.77$ ,  $SD = 1.77$ ), and that A was more willing to help them ( $M = 6.46$ ,  $SD = 1.91$ ), and had higher gratitude ratings ( $M = 8.06$ ,  $SD = 1.60$ ) than participants in the control condition (concern for their benefits:  $M = 4.32$ ,  $SD = 1.89$ ,  $F(1,100) = 45.77$ ,  $p < .001$ , partial  $\eta^2 = 0.314$ ; willingness to help:  $M = 3.78$ ,  $SD = 1.89$ ,  $F(1,100) = 50.70$ ,  $p < .001$ , partial  $\eta^2 = 0.336$ ; gratitude ratings:  $M = 3.14$ ,  $SD = 2.19$ ,  $F(1,100) = 168.96$ ,  $p < .001$ , partial  $\eta^2 = 0.628$ ) (Fig. 4A). In the gratitude condition, the gratitude ratings were significantly higher than the ratings of other emotions (all  $F_s > 49.91$ , all  $p_s < .001$ , all partial  $\eta^2_s > 0.495$ ) (Table S3). The manipulation checks revealed that our manipulation of the participants' perceptions of A's intent and resulting grateful feelings was successful.

### 5.2.2. Relationship-building tendency and relational utility

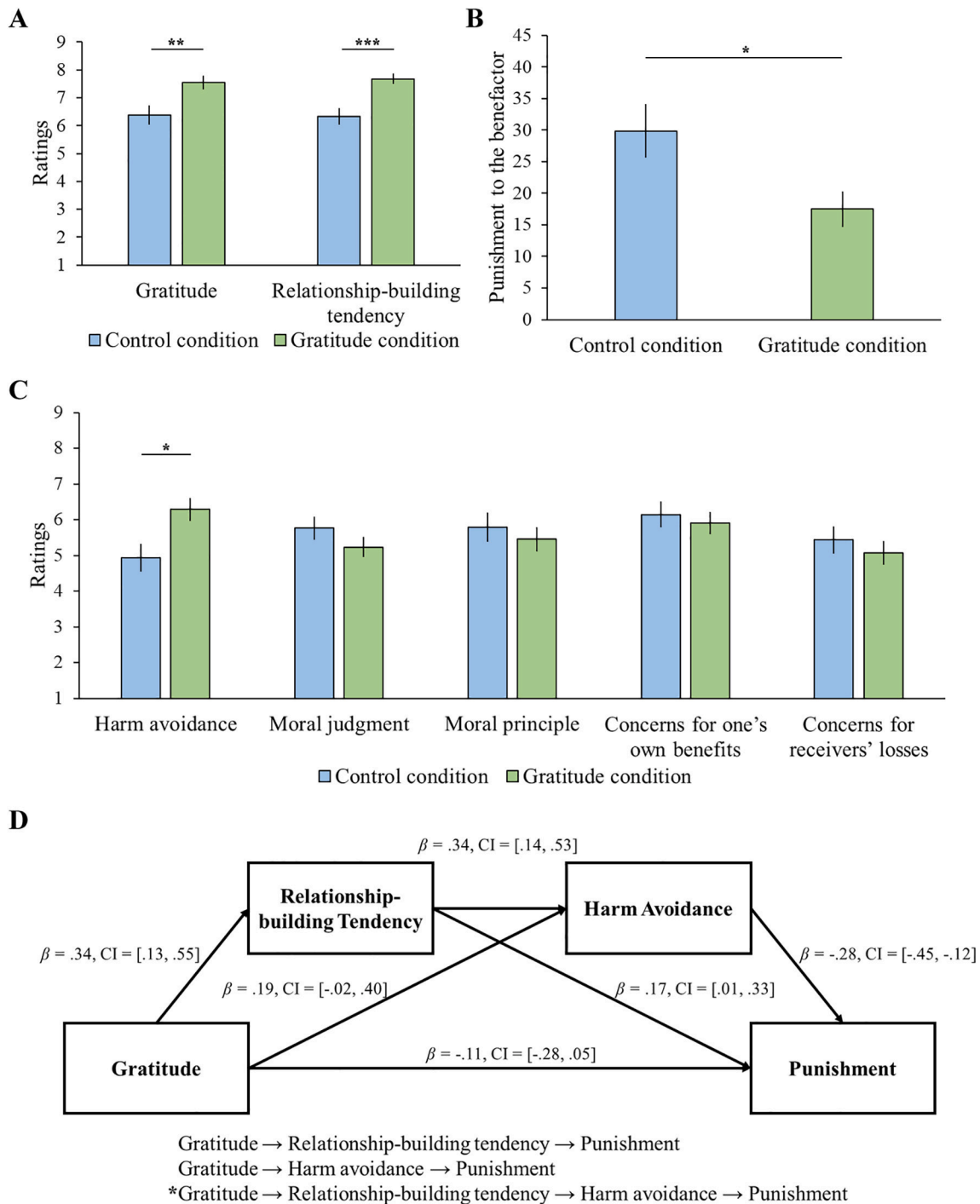
Compared with participants in the control condition (relationship-building tendency:  $M = 6.12$ ,  $SD = 1.70$ ; relational utility:  $M = 6.47$ ,  $SD = 1.70$ ), participants in the gratitude condition were more inclined to build a friendship with A ( $M = 7.85$ ,  $SD = 1.49$ ,  $F(1,100) = 29.90$ ,  $p < .001$ , partial  $\eta^2 = 0.230$ ) (Fig. 4A) and considered a friendship with A to be more beneficial ( $M = 7.67$ ,  $SD = 1.41$ ,  $F(1,99) = 15.10$ ,  $p < .001$ , partial  $\eta^2 = 0.132$ ). A significant positive correlation was found between relationship-building tendency and relational utility across participants ( $r = 0.68$ ,  $p < .001$ ,  $N = 101$ ) (Fig. S1).

### 5.2.3. Moral violation (decreased punishment)

Participants in the gratitude condition ( $M = 21.50$ ,  $SD = 22.41$ ) deducted the allocator less money than participants in the control condition did ( $M = 37.67$ ,  $SD = 26.29$ ;  $F(1,100) = 11.19$ ,  $p = .001$ ,

<sup>4</sup> Additionally, in Studies 2B and 4, the scale participants used for answering the questions was 0 to 8 (0 = not at all, 8 = very much). To increase the comparability across studies, we transformed the range to be 1 to 9 by adding one to the data.





**Fig. 3.** Results of Study 2A. A) Mean gratitude ratings and relationship-building tendency ( $\pm SE$ ) in the gratitude and control conditions.  $**p < .01$ ,  $***p < .001$ . B) Mean amount of money of the allocator deducted by participants ( $\pm SE$ ) in the gratitude and control conditions.  $*p < .05$ . C) Mean ratings of psychological factors associated with violation ( $\pm SE$ ) in the gratitude and control conditions.  $*p_{corrected} < .05$ . D) The indirect effect of gratitude on punishment through relationship-building tendency and harm avoidance in sequence was significant. Significant indirect effects are indicated by symbol \*. CI represents 95% confidence interval.

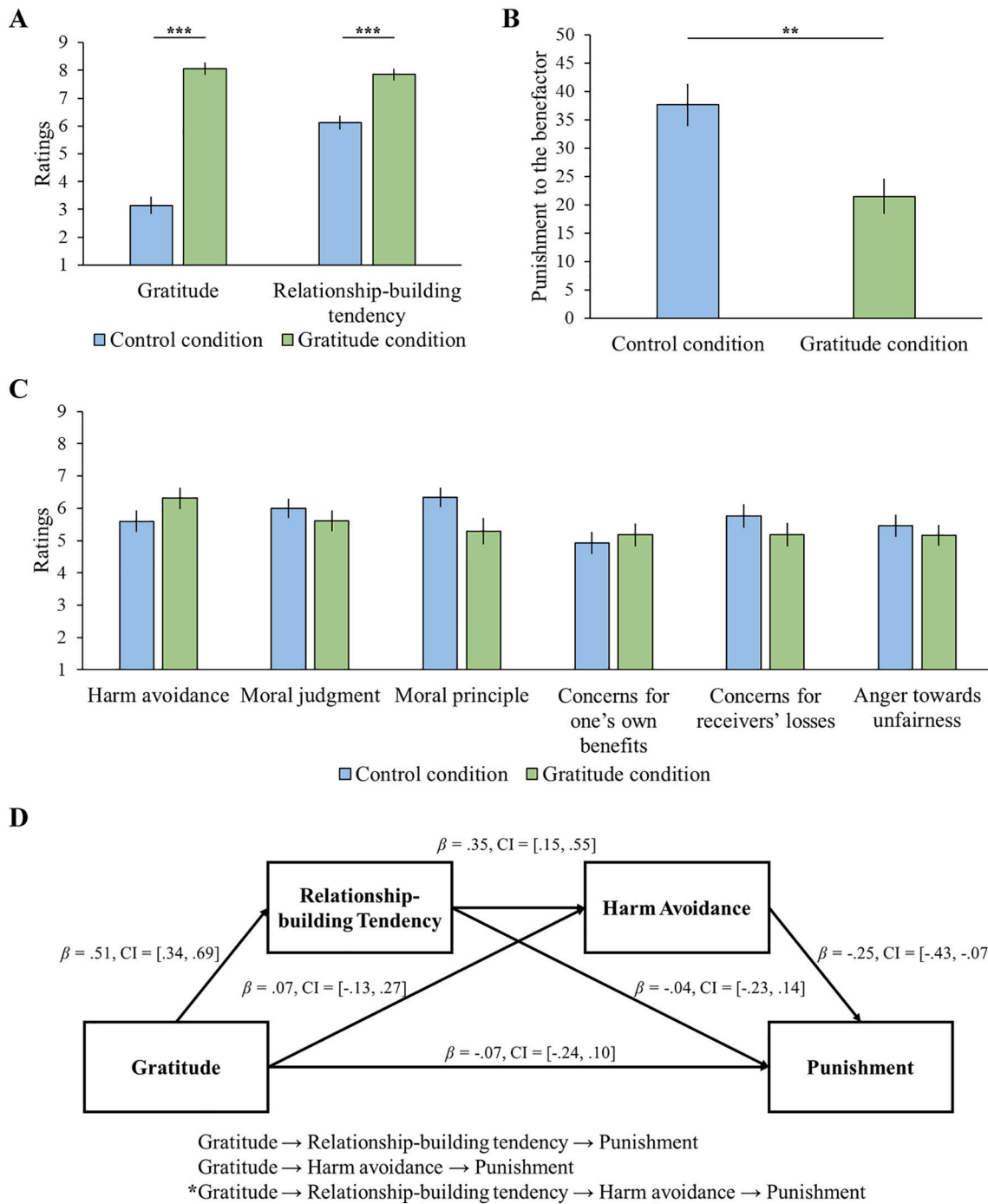
partial  $\eta^2 = 0.101$ ) (Fig. 4B).

**5.2.4. Simple mediation**

The results showed that the direct effect of gratitude on punishment was not significant ( $\beta = -0.06$ ,  $CI = [-0.29, 0.16]$ ). The indirect effect of gratitude on punishment through relationship-building tendency was not significant ( $\beta = -0.08$ ,  $CI = [-0.23, 0.06]$ ).

**5.2.5. Psychological factors associated with violation**

There was no significant difference in the ratings of harm avoidance ( $F(1,100) = 2.29$ ,  $p_{corrected} = .799$ , partial  $\eta^2 = 0.022$ ), moral judgment ( $F(1,100) = 0.76$ ,  $p_{corrected} > .999$ , partial  $\eta^2 = 0.008$ ), moral principle ( $F(1,100) = 4.32$ ,  $p_{corrected} = .241$ , partial  $\eta^2 = 0.041$ ), concerns for one's own benefits ( $F(1,100) = 0.27$ ,  $p_{corrected} > .999$ , partial  $\eta^2 = 0.003$ ), concerns for receivers' losses ( $F(1,100) = 1.23$ ,  $p_{corrected} > .999$ , partial  $\eta^2 = 0.012$ ), or anger towards unfairness ( $F(1,100) = 0.41$ ,  $p_{corrected} > .999$ , partial  $\eta^2 = 0.004$ ) between the two



**Fig. 4.** Results of Study 2B. A) Mean gratitude ratings and relationship-building tendency ( $\pm SE$ ) in the gratitude and control conditions.  $***p < .001$ . B) Mean amount of money of the allocator deducted by participants ( $\pm SE$ ) in the gratitude and control conditions.  $**p < .01$ . C) Mean ratings of psychological factors associated with violation ( $\pm SE$ ) in the gratitude and control conditions. D) The indirect effect of gratitude on punishment through relationship-building tendency and harm avoidance in sequence was significant. Significant indirect effects are indicated by symbol \*. CI represents 95% confidence interval.

conditions (Fig. 4C).

A significant correlation was found between harm avoidance and relationship-building tendency ( $r = 0.41, p_{corrected} < .001, N = 102$ ). There was no significant correlation between relationship-building tendency and moral judgment ( $r = -0.04, p_{corrected} > .999, N = 102$ ), moral principle ( $r = 0.00, p_{corrected} > .999, N = 102$ ), concerns for one's own benefits ( $r = -0.00, p_{corrected} > .999, N = 102$ ), concerns for receivers' losses ( $r = 0.07, p_{corrected} > .999, N = 102$ ), or anger towards unfairness ( $r = -0.06, p_{corrected} > .999, N = 102$ ).

#### 5.2.6. Serial multiple mediation

When harm avoidance was entered into the model as the second mediator, the direct effect of gratitude on punishment was not significant ( $\beta = -0.07, 95\% CI = [-0.24, 0.10]$ ). The indirect effect of gratitude on punishment through relationship-building tendency alone ( $\beta = -0.02, 95\% CI = [-0.11, 0.06]$ ) or through harm avoidance alone ( $\beta = -0.02, 95\% CI = [-0.08, 0.03]$ ) was not significant. Importantly, the indirect effect of gratitude on punishment through relationship-building tendency and harm avoidance in sequence was significant ( $\beta = -0.04, 95\% CI = [-0.10, -0.01]$ ) (Fig. 4D).

When any other psychological factor (namely moral judgment, moral principle, concerns for one's own benefits, concerns for receivers' losses, or anger towards unfairness) was entered into the model as the second mediator, the indirect effect of gratitude on punishment was not mediated through relationship-building tendency alone, through any psychological factor alone, or through relationship-building tendency and any psychological factor in sequence (Table S9).

In line with Study 2A, the results of Study 2B reveals that gratitude prompts individuals to benefit the benefactors at the cost of offending justice. The effect of gratitude on punishment is mediated by relationship-building tendency and harm avoidance in sequence. The Study 2B can still be improved in several aspects: 1) even though hypothetical paradigms are commonly used by studies on gratitude (e.g., Fox et al., 2015; Tsang, 2006b; Watkins et al., 2006; Wood et al., 2011), the validity and credibility of the results could be enhanced if the findings were replicated in a laboratory experiment; 2) computational models independent of participants' self-report can provide additional results revealing the psychological process underlying gratitude-induced violation; 3) gratitude increases individuals' desire for money can be an explanation for why we observed that, compared to participants in the control condition, participants in the gratitude condition were less likely to spend money on punishment. But this explanation is unlikely, as previous studies consistently showed that gratitude promotes individuals to spend money in helping others (e.g., Ma et al., 2017). However, to be doubly sure, confirming that gratitude does not increase participants' desire for money in our study is necessary to exclude the explanation above. Thus, in Study 3 we conducted a laboratory experiment, developed computational models, and measured participants' desire for money by using the social value orientation (SVO) scale (Kuhlman & Marshello, 1975).

## 6. Study 3

### 6.1. Methods

#### 6.1.1. Participants and design

We set the probability of type I error (0.05), expected effect size ( $f = 0.30$ ), and power ( $1 - \beta = 0.80$ ) and determined the minimum sample size to be 90 participants.<sup>5</sup> One hundred and ten healthy college students participated in the experiment in exchange for monetary payment. All participants reported no history of psychiatric, neurological, or cognitive diseases. Participants first read the instructions by themselves; then, the experimenter thoroughly explained the instructions to them if they had any questions. Of these, 8 participants who missed more than five questions, doubted the authenticity of the experiment, or dropped out for personal reasons were excluded (Table S1), leaving 102 participants (52 females, 50 males,  $M_{\text{age}} = 22.39$  years,  $SD_{\text{age}} = 2.32$ ; 52 in the gratitude condition, 50 in the control condition) in the analyses. This sample had 80% power to detect effects of  $f > 0.28$  at  $p < .05$  and a 2-tailed test.

#### 6.1.2. Procedure

Upon arrival, each participant met a same-sex confederate who later ostensibly interacted with the participant on computer via an internal network. The participant was then led to another room, where they underwent an individual pain calibration procedure with a BIOPAC

<sup>5</sup> At the beginning, we expected a larger effect size ( $f$  was set to be 0.40 instead of 0.30) in the laboratory-based than scenario-based study and recruited 62 valid participants in Study 3. However, two reviewers disagreed with our expectation and cast doubt on the original power analysis. Therefore, consistent with Studies 2A and 2B, we set the expected effect size  $f$  to be 0.30, determined the minimum sample size to be 90 participants, and increased the sample size to be 102. The additional samples did not change the main statistical results of Study 3.

MP160 and STMISO stimulus isolation adapter. After a brief introduction of the calibration process and precautions, the participant's left forearm was cleaned. Two disposable electrodes were placed on the back of the left wrist at intervals of 5 cm. The pain stimulation was set as three repeated square waveform electrical stimulation pulses (0.5 ms duration of each pulse and an interval of 100 ms) using AcqKnowledge 5 software. Titration began with a low-voltage electric shock (1 V) that increased or decreased (the rates of increase and decrease were about 3:1) in small increments each time (Crockett et al., 2014). The participant was asked to rate the experiences of pain on an 11-point scale (level 0 = no sensation; level 10 = maximum tolerable pain) after each shock. Voltage values corresponding to self-reported pain experience levels of four and seven were respectively recorded as moderate and strong pain stimuli (e.g., Yu et al., 2017). The moderate and strong pain stimuli were given to participants once again before the choice game and all of them reported the two levels of pain stimuli were clearly distinguishable.

After calibration, the participant played the choice game for 20 trials (Fig. 5). The rules, procedures, and outcomes of the choice game were almost the same as they were described in Study 2B. The only difference was that the shocks were to be executed at the very end of the experiment, not during the choice game.

The participant was then directed to play the TPP for 50 trials (Fig. 6). The rules of the TPP were similar to those described in Study 2B. In each trial, participants were presented with one of five types of distributions, in which the confederate distributed 200 tokens between themselves and a receiver (i.e., 200 (for the confederate):0 (for the receiver); 180:20, 160:40, 140:60, and 120:80, each type repeated 10 times in the whole task) and the receiver had to accept the distribution. The participant, owning 100 tokens, then had a chance to deduct the confederate's tokens by spending their own tokens (options: 0 to 100 tokens, in increments of 10 tokens); each 10 tokens spent deducted 30 tokens from the confederate. The tokens were convertible to a monetary bonus and that all of the players would be paid according to the tokens they obtained, in addition to a fixed show-up fee. It was informed that the receivers were different in each trial and the confederate could see the portraits of receivers before the distribution, but the participant could not see the portraits. Although receivers were not present in the lab at that time, they had come to the lab, taken time to understand the rules of the game, provided their portraits, agreed to participate as receivers, and would be paid according to the tokens they received. After the TPP, the participant answered the same questions as described in Study 2B. Additionally, they rated their emotions for the second time and completed a social value orientation (SVO) scale<sup>6</sup> (Kuhlman & Marshello, 1975), which identifies individuals' social interaction styles as 'prosocial', 'individual' (selfish), or 'competitive'. Afterwards, they received 10 moderate electric shocks and rated how painful they felt (see results in the SI-10). In the end, each participant was debriefed.

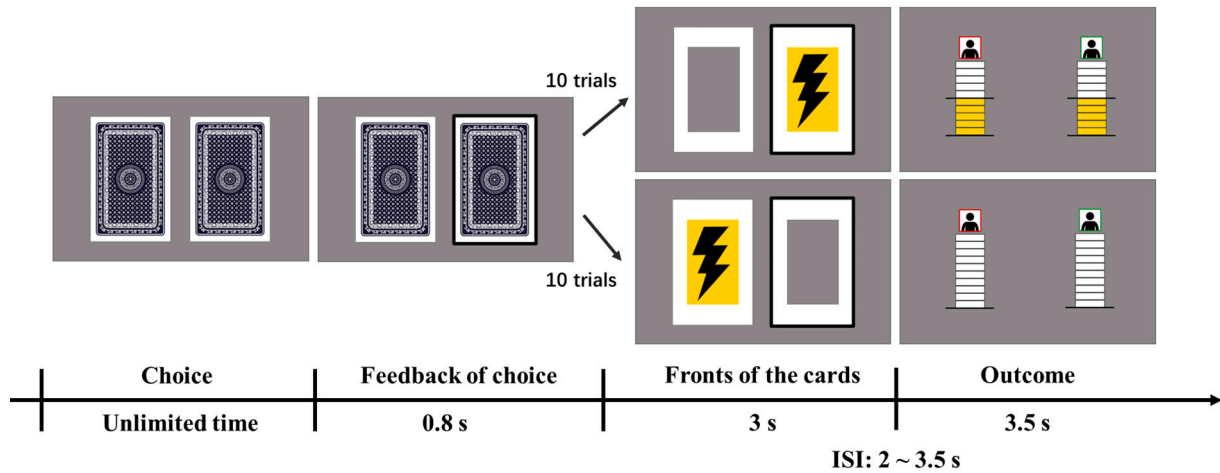
#### 6.1.3. Computational modelling

Inspired by the Fehr-Schmidt inequity aversion model (Fehr & Schmidt, 1999), we developed a computational utility model to capture the trade-off between harm aversion and inequity aversion. In this model, we defined the utility ( $U$ ) of a punishment decision as follows:

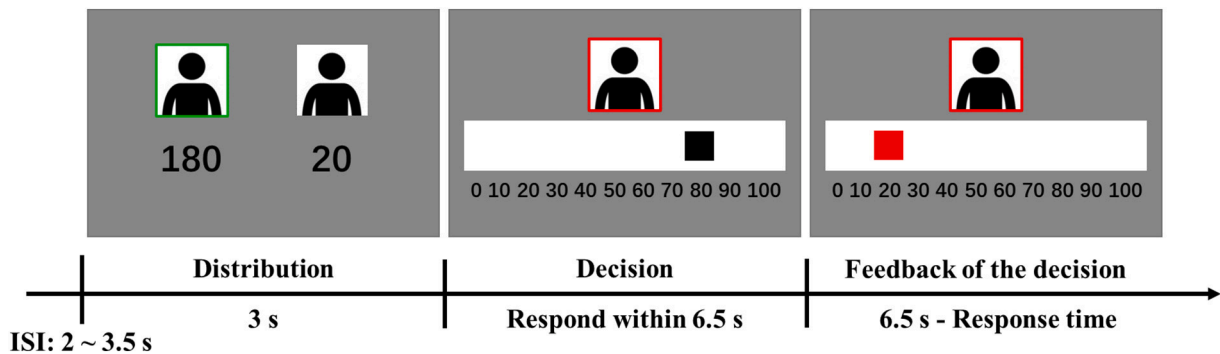
$$U(D) = \alpha \cdot (T_a - D) - \max(T_a - D - T_r, 0)$$

where  $T_a$  and  $T_r$  respectively refer to the tokens the allocator distributes to themselves and to the receiver, and  $D$  refers to the punishment the allocator receives from the decider (the number of tokens deducted by the participant).  $T_a - D$  is the final payoff of the allocator.  $\max(T_a - D - T_r, 0)$  is the inequity between the allocator and the receiver

<sup>6</sup> The participants were offered a chance to distribute money between a new stranger and them. They were told their distribution would affect the real monetary reward of the stranger and themselves.



**Fig. 5.** The timeline of the choice game. At the beginning of each trial, the participant saw the backs of two cards and chose one of them. The chosen card would be surrounded by a black frame. The cards were then turned over. After a short while, during which a green cross was shown on the screen (2–3.5 s), the participant saw the strength of the shock that they and the confederate would receive. The agent surrounded by a red frame represented the participant, while the one surrounded by a green frame represented the confederate. In 10 of the trials, the grey card was chosen and no one received a shock. In the other 10 trials, the yellow card was chosen and both the participant and the confederate received a moderate shock. The outcomes in the gratitude and control conditions were exactly the same. The only difference was that, when the yellow card was chosen, participants in the gratitude condition thought the reason why they received a moderate shock (instead of a strong shock) was that the confederate voluntarily decided to help, while participants in the control condition thought it was the rules that required the confederate to do so. ISI: interstimulus interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 6.** The timeline of the third party punishment game. Each trial started with a green cross on the screen (for 2 to 3.5 s). The participant then saw the distribution of the confederate. The agents surrounded by a green frame and a grey frame represented the confederate and a stranger, respectively. Afterwards, the participant moved the black block to indicated how many tokens they wanted to spend on punishment. The initial place of the black block was randomized. When the block was moved to the target place, the participant pressed a corresponding key to confirm their decision, and the block then turned red. The decision had to be made within 6.5 s. The agent surrounded by a red frame represented the participant. ISI: interstimulus interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

after the punishment. Thus,  $\alpha$  characterizes the relative weight on the allocator's payoff compared with the inequity between the allocator and the receiver ( $\alpha \in [0, 10]$ ) (Luo et al., 2018). The larger the parameter  $\alpha$  is, the more the decider intends to avoid harming the benefits of the allocator.

We adopted the softmax choice rule to represent the probability of a punishment decision and the maximum likelihood method to estimate parameters at the group level (see the details in the SI-7). For making statistical inference, the 95% confidence intervals of estimated parameters were obtained by a bootstrap procedure with 200 iterations (Sáez et al., 2015). To test whether this model provided a good representation of participants' behaviour, we simulated the model with the fit parameter values and compared the model to three other possible models (see the SI-8).

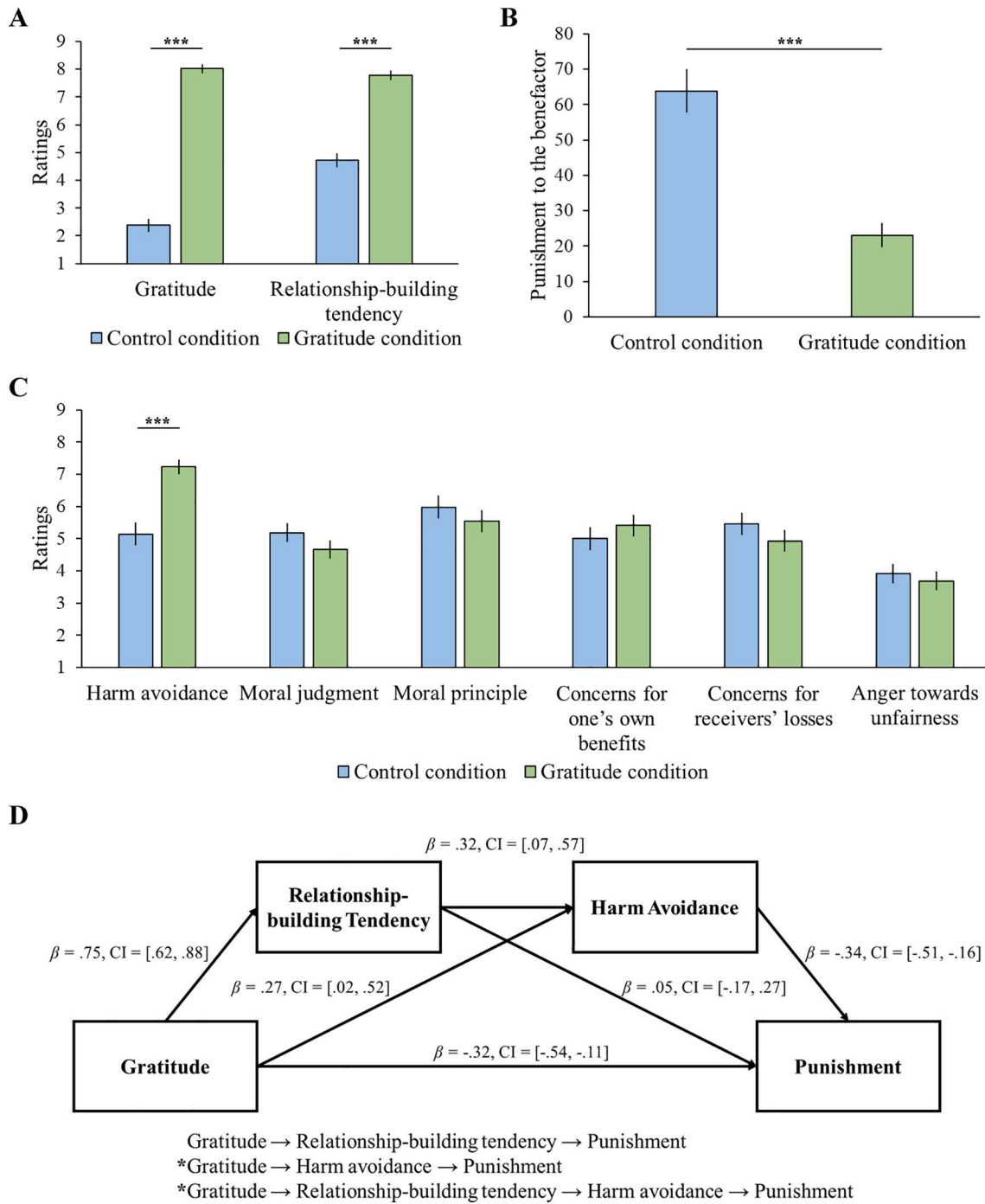
## 6.2. Results and discussion

### 6.2.1. Manipulation check

Participants in the gratitude condition thought that A concerned for their benefits more ( $M = 8.29, SD = 1.13$ ), and that A was more

willing to help them ( $M = 8.65, SD = 0.86$ ), and had higher gratitude ratings after the choice game ( $M = 8.02, SD = 1.18$ ) than participants in the control condition (concern for their benefits:  $M = 4.48, SD = 1.89, F(1,100) = 154.60, p < .001, \text{partial } \eta^2 = 0.607$ ; willingness to help:  $M = 4.74, SD = 1.98, F(1,100) = 170.22, p < .001, \text{partial } \eta^2 = 0.630$ ; gratitude ratings:  $M = 2.38, SD = 1.66, F(1,100) = 392.05, p < .001, \text{partial } \eta^2 = 0.797$ ) (Fig. 7A). In the gratitude condition, the gratitude ratings were significantly higher than the ratings of other emotions (all  $F_s > 111.31$ , all  $p_s < .001$ , all partial  $\eta^2_s > 0.686$ ) (Table S4). The manipulation checks revealed that our manipulation of the participants' perceptions of A's intent and resulting grateful feelings was successful.

Besides, after the TPP, the participants' gratitude ratings were still higher in the gratitude condition ( $M = 4.21, SD = 2.32$ ) than the control condition ( $M = 2.66, SD = 1.81, F(1,100) = 14.01, p < .001, \text{partial } \eta^2 = 0.123$ ) (Table S4). In the gratitude condition, the gratitude ratings were significantly higher than the ratings of other emotions (all  $F_s > 8.89$ , all  $p_s < .004$ , all partial  $\eta^2_s > 0.148$ ). It suggested that the induced grateful feeling remained when participants filled out the SVO scale.



**Fig. 7.** Results of Study 3. A) Mean gratitude ratings and relationship-building tendency ( $\pm SE$ ) in the gratitude and control conditions.  $***p < .001$ . B) Mean amount of tokens of the allocator deducted by participants ( $\pm SE$ ) in the gratitude and control conditions.  $***p < .001$ . C) Mean ratings of psychological factors associated with violation ( $\pm SE$ ) in the gratitude and control conditions.  $***p_{corrected} < .001$ . D) The indirect effects of gratitude on punishment through relationship-building tendency and harm avoidance in sequence and through harm avoidance alone were significant. Significant indirect effects are indicated by symbol \*. CI represents 95% confidence interval.

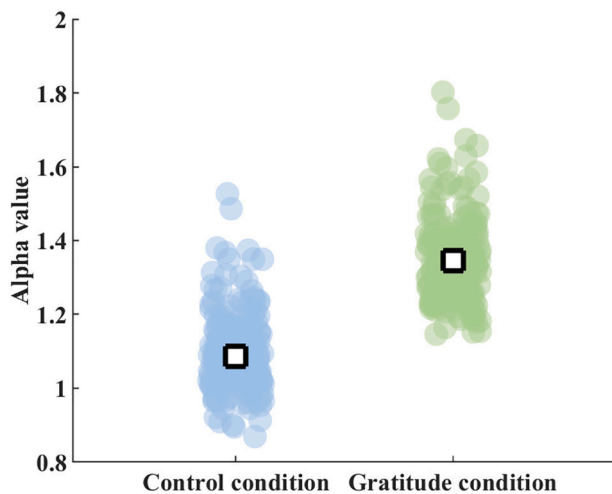
**6.2.2. Relationship-building tendency and relational utility**

Compared with participants in the control condition (relationship-building tendency:  $M = 4.72, SD = 1.73$ ; relational utility:  $M = 6.34, SD = 1.93$ ), participants in the gratitude condition were more inclined to build a friendship with A ( $M = 7.79, SD = 1.26, F(1,100) = 105.84, p < .001, \text{partial } \eta^2 = 0.514$ ) (Fig. 7A) and considered a friendship with A to be more beneficial ( $M = 8.02, SD = 1.04, F(1,100) = 30.18, p < .001, \text{partial } \eta^2 = 0.232$ ). A significant positive correlation was found between relationship-building tendency and relational utility

across all participants ( $r = 0.57, p < .001, N = 102$ ) (Fig. S1).

**6.2.3. Moral violation (decreased punishment)**

Participants in the gratitude condition ( $M = 23.05, SD = 25.05$ ) deducted the allocator less tokens than participants in the control condition did ( $M = 63.85, SD = 43.34; F(1,100) = 34.20, p < .001, \text{partial } \eta^2 = 0.255$ ) (Fig. 7B).



**Fig. 8.** Results of computational modelling. The blue dots represent the bootstrap pseudosample estimates for the control condition, while the green dots represent the bootstrap pseudosample estimates for the gratitude condition. The black squares represent the real sample estimates for the control and gratitude conditions respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

#### 6.2.4. Simple mediation

The results showed that the direct effect of gratitude on punishment was significant ( $\beta = -0.54$ ,  $CI = [-0.80, -0.27]$ ). The indirect effect of gratitude on punishment through relationship-building tendency was not significant ( $\beta = 0.07$ ,  $CI = [-0.15, 0.27]$ ).

#### 6.2.5. Psychological factors associated with violation

Participants in the gratitude condition had stronger motivation to reduce A's potential loss compared to participants in the control condition ( $F(1,100) = 25.48$ ,  $p_{corrected} < .001$ , partial  $\eta^2 = 0.203$ ) (Fig. 7C). There was no significant difference in the ratings of moral judgment ( $F(1,100) = 1.68$ ,  $p_{corrected} > .999$ , partial  $\eta^2 = 0.017$ ), moral principle ( $F(1,100) = 0.78$ ,  $p_{corrected} > .999$ , partial  $\eta^2 = 0.008$ ), concerns for one's own benefits ( $F(1,100) = 0.67$ ,  $p_{corrected} > .999$ , partial  $\eta^2 = 0.007$ ), concerns for receivers' losses ( $F(1,100) = 1.25$ ,  $p_{corrected} > .999$ , partial  $\eta^2 = 0.012$ ), or anger towards unfairness ( $F(1,100) = 0.30$ ,  $p_{corrected} > .999$ , partial  $\eta^2 = 0.003$ ) between the two conditions.

A significant correlation was found between harm avoidance and relationship-building tendency ( $r = 0.51$ ,  $p_{corrected} < .001$ ,  $N = 102$ ). There was no significant correlation between relationship-building tendency and moral judgment ( $r = 0.12$ ,  $p_{corrected} > .999$ ,  $N = 102$ ), moral principle ( $r = 0.16$ ,  $p_{corrected} = .666$ ,  $N = 102$ ), concerns for one's own benefits ( $r = 0.10$ ,  $p_{corrected} > .999$ ,  $N = 102$ ), concerns for receivers' losses ( $r = 0.08$ ,  $p_{corrected} > .999$ ,  $N = 102$ ), or anger towards unfairness ( $r = 0.15$ ,  $p_{corrected} = .747$ ,  $N = 102$ ).

#### 6.2.6. Serial multiple mediation

When harm avoidance was entered into the model as the second mediator, the direct effect of gratitude on punishment was significant ( $\beta = -0.32$ , 95%  $CI = [-0.54, -0.11]$ ). The indirect effect of gratitude on punishment through relationship-building tendency alone was not significant ( $\beta = 0.04$ , 95%  $CI = [-0.13, 0.20]$ ) and the indirect effect of gratitude on punishment through harm avoidance alone was significant ( $\beta = -0.09$ , 95%  $CI = [-0.22, -0.00]$ ). Importantly, the indirect effect of gratitude on punishment through relationship-building tendency and harm avoidance in sequence was significant ( $\beta = -0.08$ , 95%  $CI = [-0.19, -0.01]$ ) (Fig. 7D).

When any other psychological factor (namely moral judgment, moral principle, concerns for one's own benefits, concerns for receivers' losses, or anger towards unfairness) was entered into the model as the

second mediator, the indirect effect of gratitude on punishment was not mediated through relationship-building tendency alone, through any psychological factor alone, or through relationship-building tendency and any psychological factor in sequence (Table S10).

#### 6.2.7. Social value orientation

The results of the SVO scale showed that one participant in the gratitude condition and two participants in the control condition could not be classified under any particular interaction type. The rest of the participants were classified as either the prosocial or individual (selfish) type. Compared to individuals classified into the prosocial type, those who are classified into the individual type allocate more money to themselves and less money to others. The proportion of participants who act with prosocial or individual style in the gratitude condition (prosocial: 76.47%, individual: 23.53%) and control condition (prosocial: 56.25%, individual: 43.75%) revealed significant difference ( $\chi^2(1, N = 99) = 4.55$ ,  $p = .033$ , Cramer's  $V = 0.214$ ). The results were aligned with the previous findings that gratitude promotes individuals to bear a cost to benefit others (Ma et al., 2017). On one hand, the results of the SVO scale confirmed the validity of our manipulation of gratitude. On the other hand, they excluded the explanation that the gratitude condition participants' preference not to spend money on punishment was because gratitude increased their desire for money.

#### 6.2.8. Results of computational modelling

Gratitude compared with the control emotion significantly increased the parameter  $\alpha$  by 0.26 (bootstrap 95% confidence interval = [0.23, 0.27];  $\alpha_{control} = 1.09$ ,  $\alpha_{gratitude} = 1.35$ ) (Fig. 8). It suggested that compared with the control condition, participants were more inclined to avoid harming the benefits of the allocator in the gratitude condition.

Additionally, the harm aversion and inequity aversion model here best captured participants' behaviour compared against the other models based on the Bayesian Information Criterion (Table S12) and the simulated results coincided well with the real results (Fig. S2). These results demonstrated the validity of the model and denoted the importance of harm aversion (and inequity aversion) in the gratitude-induced violation.

Again, Study 3 show that gratitude urges participants to benefit their benefactors by lowering due punishment. The effect of gratitude on punishment is achieved through relationship-building tendency and harm avoidance in sequence. Moreover, independent of participants' self-report, the model-based results manifest a vital role of harm aversion in punishment decision and reveal that grateful individuals are inclined to avoid harming the benefits of their benefactors.

Notably, since justice means giving transgressors the proper punishment they deserve, not only decreasing due punishment, but also increasing due punishment offend justice (e.g., van Den Haag, 1980). Till now, our findings showed that grateful individuals decrease due punishment to their benefactors for protection. Nevertheless, it is unclear whether grateful individuals increase due punishment to others who unfairly treat the benefactors of the individuals (gratitude-induced harm). To explore the generalization of gratitude-induced moral violation, in Study 4 we investigated whether gratitude leads individuals to avenge their benefactors by harming others (e.g., increasing due punishment).

## 7. Study 4

### 7.1. Methods

#### 7.1.1. Participants and design

We set the probability of type I error (0.05), expected effect size ( $f = 0.30$ ), and power ( $1 - \beta = 0.80$ ) and determined the minimum sample size to be 90 participants. One hundred and thirty-seven college students participated in the experiment in exchange for course credits.

Of these, 42 participants who gave an incorrect answer during the two comprehension tests of the instructions, failed an attention check, or missed more than five questions were excluded (Table S1), leaving 95 participants (53 females, 42 males,  $M_{\text{age}} = 20.54$  years,  $SD_{\text{age}} = 3.05$ ; 50 in the gratitude condition, 45 in the control condition) in the analyses. This sample had 80% power to detect effects of  $f > 0.29$  at  $p < .05$  and a 2-tailed test.

We used the paradigm of Study 2B, but changed the role of participants' benefactors from the allocator who made unfair divisions into the receiver who received unfair divisions from other strange allocators. That is, the benefactor became a victim instead of a violator. In this case, we focused on examining whether gratitude affects participants' punishment to other strange allocators who divided unfair divisions to their benefactors.

### 7.1.2. Procedure

The procedure of Study 4 was same as that in Study 2B, except for the following changes. First, A was the receiver in the four trials. Strangers B, C, D, and E played the role of the allocator. Additionally, the question measuring harm avoidance was changed to read 'to what extent you want to reduce the losses of the allocators when making decisions', and the question measuring concerns for the receiver's losses changed to read 'to what extent do you concern for A's losses when making decisions'.

## 7.2. Results and discussion

### 7.2.1. Manipulation check

Participants in the gratitude condition thought that A concerned for their benefits more ( $M = 6.88$ ,  $SD = 1.85$ ), and that A was more willing to help them ( $M = 6.60$ ,  $SD = 1.74$ ), and had higher gratitude ratings ( $M = 7.64$ ,  $SD = 2.12$ ) than participants in the control condition (concern for their benefits:  $M = 4.69$ ,  $SD = 2.27$ ,  $F(1,93) = 26.78$ ,  $p < .001$ , partial  $\eta^2 = 0.224$ ; willingness to help:  $M = 4.27$ ,  $SD = 1.88$ ,  $F(1,93) = 39.60$ ,  $p < .001$ , partial  $\eta^2 = 0.299$ ; gratitude ratings:  $M = 4.11$ ,  $SD = 2.76$ ,  $F(1,93) = 49.52$ ,  $p < .001$ , partial  $\eta^2 = 0.347$ ) (Fig. 9A). In the gratitude condition, the gratitude ratings were significantly higher than the ratings of other emotions (all  $F$ s  $> 56.56$ , all  $p$ s  $< .001$ , all partial  $\eta^2$ s  $> 0.536$ ) (Table S5). The manipulation checks revealed that our manipulation of the participants' perceptions of A's intent and resulting grateful feelings was successful.

### 7.2.2. Relationship-building tendency and relational utility

Compared with participants in the control condition (relationship-building tendency:  $M = 5.42$ ,  $SD = 2.22$ ; relational utility:  $M = 5.91$ ,  $SD = 1.99$ ), the participants in the gratitude condition were more inclined to build a friendship with A ( $M = 7.62$ ,  $SD = 1.46$ ,  $F(1,93) = 33.17$ ,  $p < .001$ , partial  $\eta^2 = 0.263$ ) (Fig. 9A) and considered a friendship with A to be more beneficial ( $M = 7.34$ ,  $SD = 1.19$ ,  $F(1,93) = 18.52$ ,  $p < .001$ , partial  $\eta^2 = 0.166$ ). A significant positive correlation was found between relationship-building tendency and relational utility across all participants ( $r = 0.66$ ,  $p < .001$ ,  $N = 95$ ) (Fig. S1).

### 7.2.3. Moral violation (increased punishment)

The money of the allocators deducted by participants did not differ significantly between the gratitude ( $M = 58.65$ ,  $SD = 29.64$ ) and control conditions ( $M = 56.48$ ,  $SD = 32.98$ ;  $F(1,93) = 0.11$ ,  $p = .737$ , partial  $\eta^2 = 0.001$ ) (Fig. 9B).

### 7.2.4. Simple mediation

The results showed that the direct effect of gratitude on punishment was not significant ( $\beta = 0.07$ ,  $CI = [-0.16, 0.30]$ ). The indirect effect of gratitude on punishment through relationship-building tendency was not significant ( $\beta = 0.05$ ,  $CI = [-0.05, 0.18]$ ).

### 7.2.5. Psychological factors associated with violation

There was no significant difference in the ratings of harm avoidance ( $F(1,93) = 1.50$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 = 0.016$ ), moral judgment ( $F(1,93) < 0.01$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 < 0.001$ ), moral principle ( $F(1,93) = 0.40$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 = 0.004$ ), concerns for one's own benefits ( $F(1,93) < 0.01$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 < 0.001$ ), concerns for receivers' losses ( $F(1,93) = 0.32$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 = 0.003$ ), or anger towards unfairness ( $F(1,93) = 0.568$ ,  $p_{\text{corrected}} > .999$ , partial  $\eta^2 = 0.006$ ) between the two conditions (Fig. 9C).

Significant or marginally significant correlations were found between relationship-building tendency and several psychological factors, including moral judgment ( $r = 0.33$ ,  $p_{\text{corrected}} = .007$ ,  $N = 95$ ), moral principle ( $r = 0.28$ ,  $p_{\text{corrected}} = .037$ ,  $N = 95$ ), concerns for receivers' losses ( $r = 0.41$ ,  $p_{\text{corrected}} < .001$ ,  $N = 95$ ), and anger towards unfairness ( $r = 0.26$ ,  $p_{\text{corrected}} = .070$ ,  $N = 95$ ). No significant correlation was found between relationship-building tendency and harm avoidance ( $r = 0.17$ ,  $p_{\text{corrected}} = .553$ ,  $N = 95$ ) or concerns for one's own benefits ( $r = -0.16$ ,  $p_{\text{corrected}} = .685$ ,  $N = 95$ ).

### 7.2.6. Serial multiple mediation

Whichever psychological factor was entered into the model as the second mediator, there was no significant direct or indirect effect of gratitude on punishment (Table S11).

The results of Study 4 show that gratitude doesn't increase individuals' punishment to strangers who treat their benefactors unfairly. It suggests that grateful individuals refuse to engage in moral violation when the violation causes harm.

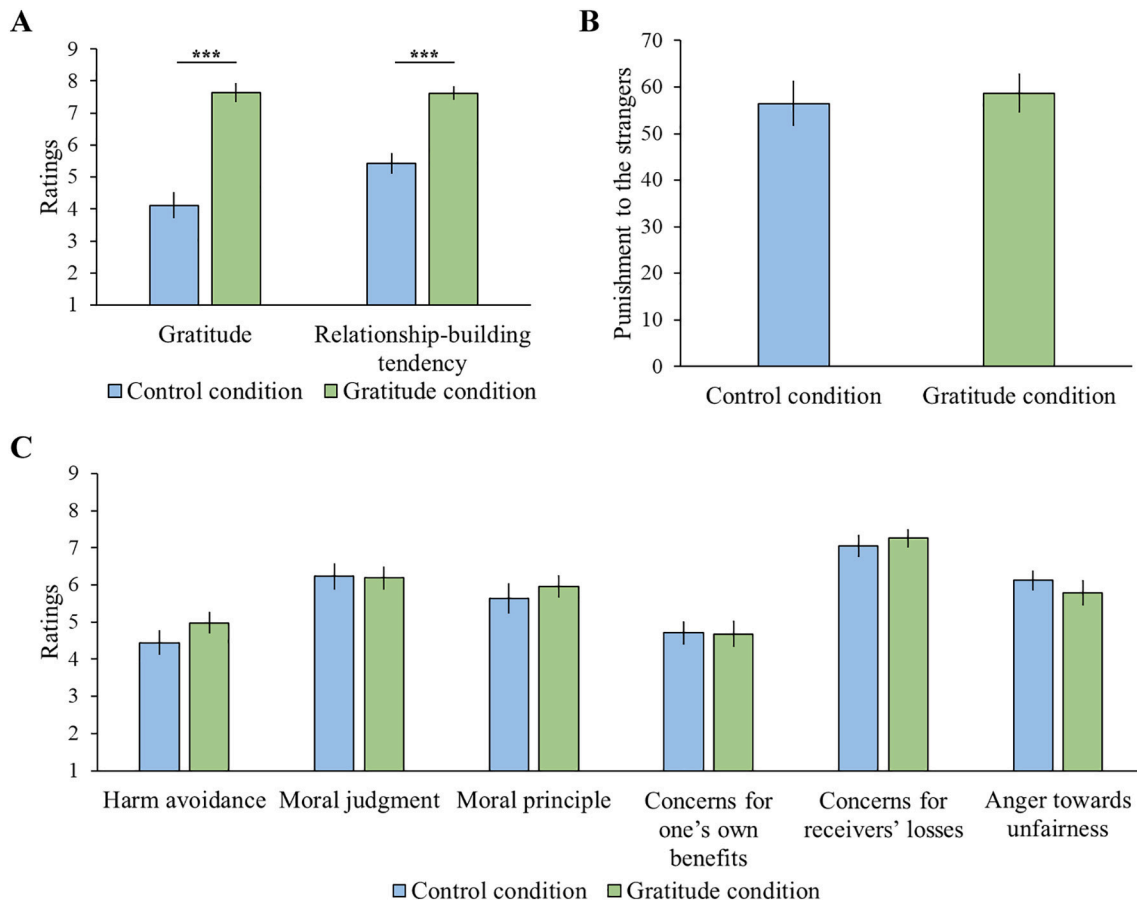
## 8. Meta-analyses

The effect sizes of gratitude on harm avoidance varied across studies. To provide more information about the consistency of the findings, we conducted random effect meta-analyses across Studies 1A, 1B, 2A, 2B and 3, using the JASP 0.11.1 software. The meta-analyses utilized Cohen's  $d$  and its standard error as inputs. The results showed that gratitude significantly increases participants' harm avoidance motivation (Cohen's  $d = 0.73$ , 95%  $CI = [0.47, 0.99]$ ,  $Z = 5.56$ ,  $p < .001$ ). The finding indicates a strong influence of gratitude on individuals' harm avoidance motivation towards their benefactors.

## 9. General discussion

Using self-reported, behavioural and computational levels of measures, we studied the effect of gratitude on moral violation with the purpose of understanding underlying psychological mechanisms. We have presented experimental evidence that gratitude increases individuals' moral violation when the violation protects their benefactors from harm. Explorations on the underlying mechanisms manifest the important roles of relationship-building tendency towards the benefactors and harm avoidance on behalf of the benefactors.

It has been found gratitude promotes a variety of behaviours which may benefit individuals and society (e.g., enhancing prosociality and reducing cheating and economic impatience) (DeSteno et al., 2019; Dickens & DeSteno, 2016; Ma et al., 2017; Tsang, 2006a). Different from previous studies, we revealed a potential dark side of gratitude by showing that grateful individuals are willing to benefit their benefactors at the cost of violating moral norms. However, it does not mean that our results have theoretical conflicts with previous findings. The social function of gratitude has been proposed to promote high-quality personal relationship (e.g., find-remind-bind theory) (Algoe, 2012; Algoe et al., 2008). In this theoretical framework, both the current and previous studies furnish evidence for the proposition that gratitude encourages individuals to weight personal relationship, while devalues the other objects. The difference is that in the previous studies, participants attempted to promote personal relationship with the



**Fig. 9.** Results of Study 4. A) Mean gratitude ratings and relationship-building tendency ( $\pm SE$ ) in the gratitude and control conditions.  $***p < .001$ . B) Mean amount of money of the allocator deducted by participants ( $\pm SE$ ) in the gratitude and control conditions. C) Mean ratings of psychological factors associated with violation ( $\pm SE$ ) in the gratitude and control conditions.

benefactors at the expense of secular values (e.g., time, money and effort), while in our studies, participants did it at the expense of sacred values (e.g., moral norm). As sacred values and secular values have distinct representations in the human mind (Berns et al., 2012; Vilarroya, 2013), our findings broaden the understanding of the scope of gratitude's effect.

Our findings are consistent with a previous study on guilt, another so-called moral emotion (de Hooghe et al., 2011). de Hooghe et al. (2011) found that guilt promotes individuals to benefit the victim at the cost of a third party's benefits. They advocated that guilt concentrates individuals' attention on repairing the damaged relationship with the victim and, consequently, reduces individuals' concerns for a third party (de Hooghe et al., 2011). Analogously, our findings suggest gratitude draws individuals' attention to promoting personal relationship, and hence causes the neglect of moral norm.

Beyond reporting a new phenomenon about gratitude, we offer insights of the underlying psychological mechanisms of gratitude-induced moral violation. In the first step, we identified several factors closely related to gratitude and moral behaviour. We found that grateful individuals report a stronger tendency to promote the relationship with their benefactors. The results echo previous findings that gratitude increases closeness between individuals and their benefactors (Algoe et al., 2008; Yu et al., 2017). A strong relationship-building tendency may lead grateful individuals to treat their benefactors as friends and be averse to harm the benefits of the benefactors (e.g., Decety & Cowell, 2018; Ma et al., 2011). Additionally, we observed a positive correlation between relationship-building tendency and relational utility. Gratitude has been found to be a function of three relational features (intent, cost, and benefit) (Forster et al., 2017; Tesser et al., 1968). The sensitivity to

the relational features implies gratitude helps individuals to identify high-quality partners (Algoe, 2012). Given relationship-building tendency emerges with gratitude, our result suggests grateful individuals form relationship-building tendency towards the benefactors, because they believe the relationship with the benefactors can benefit them in the future. In support of the social function of gratitude being relationship promotion, these findings enrich the understanding of the connections among gratitude, relationship-building tendency, and relational utility.

Based on participants' self-report, the results of meta-analyses for Studies 1A, 1B, 2A, 2B and 3 provide an indication that that gratitude has a strong effect on individuals' harm avoidance motivation towards their benefactors. In line with participants' self-report, the results of computational modelling in Study 3 also found that gratitude increases individuals' aversion to harming the benefits of the benefactors. Diverse evidence from previous studies has revealed a vital role of harm aversion in moral decision-making. It is found that individuals averse to harm others physically or economically, even when it would save lives or support charity (Cushman et al., 2012; Perera et al., 2016). Crockett et al. (2014) showed that individuals are reluctant to harm others for self-interest, and aversion to harming others outweighs aversion to harming self. Thus, our findings indicate that gratitude psychologically prepares individuals for violating moral norms for their benefactors by enhancing their aversion to harming the benefactors.

In the second step, we explored the relationship among gratitude, moral violation and identified related factors. Given gratitude's social function is proposed to be relationship promotion (Algoe, 2012), a possible assumption is that relationship-building tendency mediates the effect of gratitude on moral violation. However, the mediation effect of



relationship-building tendency (alone) was not significant in most studies (Studies 1A, 2A, 2B and 3). Instead, the serial multiple mediation analysis showed that the effect of gratitude on moral violation was mediated by relationship-building tendency and harm avoidance in sequence even when the effects of other psychological factors associated with moral violation being controlled in Studies 1A, 1B, 2A, 2B and 3. The significant serial multiple mediation effect reveals a process like this: 1) A grateful individual generates relationship-building tendency; 2) This tendency guides the grateful individual to consider the benefactor as a friend and to averse harming the benefactor just like averse harming their friend; 3) The aversion to harm eventually causes individuals violating moral norm for protecting the benefactor. It is noteworthy that this is not the only way that gratitude exerts influence on moral violation. For example, in Study 3 the direct effect of gratitude on moral violation was also significant. But, importantly, the significant serial multiple mediation effect verifies the involvement of relationship-building tendency in gratitude-induced violation, which is a strong support for the social function of gratitude being relationship promotion. With the mediation analyses, we delineate the underlying psychological mechanisms of gratitude-induced moral violation.

In the current studies, participants in both the gratitude and control conditions obtained benefits from A, and A bore the cost of helping. So, the difference in participants' moral violation between the conditions was not due to the need for equity restoration (e.g., return a favour), but on account of the intent of A (e.g., [Algoe et al., 2008](#); [Peng et al., 2018](#)). A's intent was kind in the gratitude condition, whereas it was utilitarian (Studies 1A and 1B) or ambiguous (Studies 2A, 2B, 3) in the control condition. Therefore, compared to the control condition, participants in the gratitude condition were more inclined to violate moral norm for A, probably because they perceived the kindness and sought a high-quality relationship with A, not because they felt the need to repay.

Although grateful individuals violate moral norm to protect their benefactors (Studies 1A, 1B, 2A, 2B and 3), they reject to engage in moral violation when the violation results in harm to others (Study 4). A possible explanation is that grateful individuals are uncertain about the thoughts of the benefactor. Though the benefactor is treated unfairly by the allocator, the benefactor may still prefer to avoid harming the allocator's benefits. The uncertainty of the benefactor's thoughts prevents grateful individuals from exerting extra punishment on the allocator. Rather than decline the existence of gratitude-induced harm, our findings provide preliminary evidence that gratitude-induced harm may be different from gratitude-induced protection.

Besides gratitude, other covariant emotions (e.g., guilt) were also different between the gratitude and control conditions (Tables S2, S3, S4, and S5), which is a common problem for studies on social emotions (e.g., [Wagner et al., 2011](#)). To test whether gratitude was responsible for the observed effects of the current studies, we involved other emotions into the statistical analyses as covariates. Firstly, we examined whether the difference in lying or punishment remained between the gratitude and control conditions, when unhappiness, indebtedness, anger, guilt, and shame were entered into the analyses as covariates (see the SI-11). Controlling the effects of these emotions did not change the statistical significance of the previous findings in Studies 1A, 2A, 2B, 3, and 4. Secondly, we investigated whether other emotions exerted significant effects on lying or punishment through relationship-building tendency and harm avoidance in sequence (other emotion → relationship-building tendency → harm avoidance → lying or punishment) (see Tables S13, S14, S15, S16, S17, and S18). We found significant serial multiple mediation effects when some no-gratitude emotions (e.g., unhappiness and anger) were entered into the model as the independent variable in Studies 1A, 1B, 2A, 2B, and 3. However, these significant effects disappeared when gratitude was entered into the serial multiple mediation analysis as a covariate in Studies 1A, 2A, 2B and 3. The serial multiple mediation effect remained, when gratitude was entered as the independent variable and other emotions were

involved as covariates in Studies 1A, 1B, 2A, 2B and 3. These findings advocate that gratitude was the key emotion responsible for the moral violation and the serial multiple mediation effect we observed.

A large number of studies have attempted to improve individuals' health and well-being through gratitude interventions (see a review, [Wood et al., 2010](#)). Despite the efficiency of the gratitude interventions ([Davis et al., 2016](#)), recently some researcher warned that blindly promoting gratitude could be problematic in some situations. [Ng et al. \(2017\)](#) found that gratitude facilitated conformity in a colour judgment task and a material consumption task. They argued that gratitude may lead individuals to follow social conventions, regardless of what these conventions are ([Ng et al., 2017](#)). [Algoe and Zhaoyang \(2016\)](#) revealed that individuals in the expressed gratitude condition with unresponsive partners had stronger negative emotions than individuals in the control condition with unresponsive partners. It offers a caveat that artificially injecting gratitude (an expressed gratitude practice) into a romantic relationship where the partner is unresponsive may backfire ([Algoe & Zhaoyang, 2016](#)). [Wood et al. \(2016\)](#) proposed gratitude may be harmful in an objectively abusive relationship, with the victim feeling grateful to the abuser. In the same line, our finding suggest that gratitude promotes individuals to violate the moral norms that go against their benefactors' benefits, which may disrupt social order. We hope these findings are relevant for clinical psychologists' use to consider how and with whom gratitude interventions should be used.

One may wonder the decision to lie in Studies 1A and 1B is immoral or not. We think it is an open question. From the perspective of utilitarianism, which proposes that one should take action to produce the greatest amount of benefits for the largest number of people ([Casebeer, 2003](#)), lying in Studies 1A and 1B may be immoral. Lying to the boss means that one takes action to benefit their benefactors who skip off work, while telling the truth means that one take action to maintain the order of the company, which is conducive to the development of the company and may benefit all members of the company in a long run. Sure, deontologist, virtue theorists, and people who hold moral theories different from utilitarianism are likely to disagree with this judgment. Similar to our studies, [Lupoli et al. \(2017\)](#) also investigated whether individuals lie to prevent harm to others and found that compassion increases lying for others' benefit. They considered lying or not in such a condition as a morally complex decision which presents a conflict between the norm of honesty and the norm of harm avoidance. We think judging whether lying for others is immoral or not is beyond the scope of our studies. We only propose that lying (even for others) can be regarded as a violation of the norm of honesty ([Lupoli et al., 2017](#)).

There were several limitations in our studies. The findings about gratitude-related harm in Study 4 have not been replicated in a laboratory experiment. The limitations of the hypothetical paradigm prevent us from drawing a strong conclusion based on the findings we have ([FeldmanHall et al., 2012](#)). The Study 4 is an interesting start. Future studies on gratitude-related harm are needed.

We planned to use the data of the relationship-building tendency and relational utility to investigate the associations among the gratitude, relationship-building tendency, relational utility, intent of the benefactor, costs undertaken by the benefactor for helping the beneficiary, and benefits the beneficiary receiving from the benefactor, and to report them in another independent article. A comment from an anonymous reviewer inspired us to test whether relationship-building tendency correlates with relational utility, whether gratitude affects relationship-building tendency, and whether relationship-building tendency plays a role in gratitude-induced moral violation. These results were finally involved in the current article. Here we kindly remind that the analyses about the relationship-building tendency and relational utility were post-hoc to some extent.

We note that we conducted the serial multiple mediation analysis more than once in each study for testing different psychological factors as the second mediator. This caused alpha inflation, which might need to be controlled. The significant results of the serial multiple mediation

in our studies would not survive after the Bonferroni correction. If we found the significant effect of the serial multiple mediation in only one study, we would consider this finding as a type I error. However, we observed that the relationship-building tendency and harm avoidance (but not any other psychological factor) marginally or significantly (when the Bonferroni correction was not conducted) mediated the effect of gratitude on moral violation across Studies 1A, 1B, 2A, 2B and 3. The findings suggested that the effect of the serial multiple mediation probably existed, but not strong enough to pass the Bonferroni correction in the current article.

Our studies focused on testing whether gratitude leads to violation of the norms of honesty and justice. Besides honesty and justice, there are other widely accepted moral standards, such as authority, sanctity and so on (Graham et al., 2013). Whether grateful individuals are willing to violate other moral standards for their benefactors merits further enquiry.

In conclusion, the current work demonstrates that gratitude encourages individuals to enhance personal relationships with their benefactors through benefitting the benefactors at the cost of violating moral norms of honesty and justice. Relationship-building tendency towards the benefactors and harm avoidance (aversion) on behalf of the benefactors play crucial roles in the effect of gratitude on moral violation. The findings highlight the social function of gratitude to be relationship promotion and deepen our understanding of the relationship between gratitude and moral violation, and its underlying psychological and computational mechanisms.

## Acknowledgements

This work was supported by the National Key Research and Development Program of China (2017YFC0803402), the National Natural Science Foundation of China (NSFC) (31871094), the Major Project of National Social Science Foundation (19ZDA363), the Beijing Municipal Science and Technology Commission (Z151100003915122), and the National Program for Support of Top-notch Young Professionals.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jesp.2020.104048>.

## References

- Agler, R., & De Boeck, P. (2017). On the interpretation and use of mediation: Multiple perspectives on mediation analysis. *Frontiers in Psychology, 8*, 1–11.
- Algoe, S. B. (2012). Find, remind, and bind: The functions of gratitude in everyday relationships. *Social and Personality Psychology Compass, 6*, 455–469.
- Algoe, S. B., Haidt, J., & Gable, S. L. (2008). Beyond reciprocity: Gratitude and relationships in everyday life. *Emotion, 8*, 425–429.
- Algoe, S. B., & Zhaoyang, R. (2016). Positive psychology in context: Effects of expressing gratitude in ongoing relationships depend on perceptions of enactor responsiveness. *Journal of Positive Psychology, 11*, 399–415.
- Bartlett, L., & DeSteno, D. (2006). Gratitude and prosocial behavior. *Psychological Science, 17*, 319–325.
- Berns, G. S., Bell, E., Capra, C. M., Prietula, M. J., Moore, S., Anderson, B., ... Atran, S. (2012). The price of your soul: Neural evidence for the non-utilitarian representation of sacred values. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*, 754–762.
- Casebeer, W. D. (2003). Moral cognition and its neural constituents. *Nature Reviews Neuroscience, 4*, 841–846.
- Cicero, M. T. (1851). *The orations of Marcus Tullius Cicero*. Vol. III. London, UK: George Bell & Sons (C. D. Younge, Trans.).
- Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology, 104*, 216–235.
- 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 of the United States of America, 111*, 17320–17325.
- Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*, 353–380.
- Cushman, F., Gray, K., Gaffey, A., & Mendes, W. B. (2012). Simulating murder: The aversion to harmful action. *Emotion, 12*, 2–7.
- Davis, D. E., Choe, E., Meyers, J., Wade, N., Varjas, K., Gifford, A., ... Worthington, E. L. (2016). Thankful for the little things: A meta-analysis of gratitude interventions. *Journal of Counseling Psychology, 63*, 20–31.
- de Hooge, I. E., Nelissen, R. M. A., Breugelmans, S. M., & Zeelenberg, M. (2011). What is moral about guilt? Acting “prosocially” at the disadvantage of others. *Journal of Personality and Social Psychology, 100*, 462–473.
- Decety, J., & Cowell, J. M. (2018). Interpersonal harm aversion as a necessary foundation for morality: A developmental neuroscience perspective. *Development and Psychopathology, 30*, 153–164.
- DeSteno, D., Bartlett, M. Y., Baumann, J., Williams, L. A., & Dickens, L. (2010). Gratitude as moral sentiment: Emotion-guided cooperation in economic exchange. *Emotion, 10*, 289–293.
- DeSteno, D., Duong, F., Lim, D., & Kates, S. (2019). The grateful don't cheat: Gratitude as a fount of virtue. *Psychological Science, 30*, 979–988.
- Desteno, D., Li, Y., Dickens, L., & Lerner, J. S. (2014). Gratitude: A tool for reducing Economic impatience. *Psychological Science, 25*, 1262–1267.
- Dickens, L., & DeSteno, D. (2016). The grateful are patient: Heightened daily gratitude is associated with attenuated temporal discounting. *Emotion, 16*, 421–425.
- Farrell, S., & Lewandowsky, S. (2010). Computational models as aids to better reasoning in psychology. *Current Directions in Psychological Science, 19*, 329–335.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavioral Research Methods, 39*, 175–191.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*, 137–140.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics, 114*, 817–868.
- FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition, 123*, 434–441.
- Forster, D. E., Pedersen, E. J., Smith, A., McCullough, M. E., & Lieberman, D. (2017). Benefit valuation predicts gratitude. *Evolution and Human Behavior, 38*, 18–26.
- Fox, G. R., Kaplan, J., Damasio, H., & Damasio, A. (2015). Neural correlates of gratitude. *Frontiers in Psychology, 6*, 1–11.
- Gao, X., Yu, H., Sáez, I., Blue, P. R., Zhu, L., Hsu, M., & Zhou, X. (2018). Distinguishing neural correlates of context-dependent advantageous- and disadvantageous-inequity aversion. *Proceedings of the National Academy of Sciences, 115*, E7680–E7689.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in experimental social psychology*. Vol. 47. *Advances in experimental social psychology* (pp. 55–130). Elsevier.
- 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*, 366–385.
- Greene, J. D. (2014). *Moral tribes: Emotion, reason, and the gap between us and them*. London: Penguin.
- Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 852–870). Oxford: Oxford University Press.
- Kuhlman, D. M., & Marshello, A. F. (1975). Individual differences in game motivation as moderators of preprogrammed strategy effects in prisoner's dilemma. *Journal of Personality and Social Psychology, 32*, 922–931.
- Luo, Y., Héту, S., Lohrenz, T., Hula, A., Dayan, P., Ramey, S. L., ... Ramey, C. (2018). Early childhood investment impacts social decision-making four decades later. *Nature Communications, 9*, 4705.
- Lupoli, M. J., Jampol, L., & Oveis, C. (2017). Lying because we care: Compassion increases prosocial lying. *Journal of Experimental Psychology: General, 146*, 1026–1042.
- Ma, L. K., Tunney, R. J., & Ferguson, E. (2017). Does gratitude enhance prosociality? A meta-analytic review. *Psychological Bulletin, 143*, 601–635.
- Ma, Q., Shen, Q., Xu, Q., Li, D., Shu, L., & Weber, B. (2011). Empathic responses to others' gains and losses: An electrophysiological investigation. *NeuroImage, 54*, 2472–2480.
- McCullough, M., & Kilpatrick, S. (2001). Is gratitude a moral affect? *Psychological Bulletin, 127*, 249–266.
- McCullough, R. A. E. M. E. (2004). *The psychology of gratitude*. Oxford University Press.
- Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences, 11*, 319–321.
- Nelissen, R. (2014). Relational utility as a moderator of guilt in social interactions. *Journal of Personality and Social Psychology, 106*, 257–271.
- Nelissen, R. M. A., & Zeelenberg, M. (2009). Moral emotions as determinants of third-party punishment: Anger, guilt, and the functions of altruistic sanctions. *Judgment and Decision making, 4*, 543–553.
- Ng, J. W. X., Tong, E. M. W., Sim, D. L. Y., Teo, S. W. Y., Loy, X., & Giesbrecht, T. (2017). Gratitude facilitates private conformity: A test of the social alignment hypothesis. *Emotion, 17*, 379–387.
- Ohtsubo, Y., & Yagi, A. (2015). Relationship value promotes costly apology-making: Testing the valuable relationships hypothesis from the perpetrator's perspective. *Evolution and Human Behavior, 36*, 232–239.
- Peng, C., Nelissen, R. M. A., & Zeelenberg, M. (2018). Reconsidering the roles of gratitude and indebtedness in social exchange. *Cognition and Emotion, 32*, 760–772.
- Perera, P., Canic, E., & Ludvig, E. A. (2016). Cruel to be kind but not cruel for cash: Harm aversion in the dictator game. *Psychonomic Bulletin & Review, 23*, 893–898.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–891.
- Qu, C., Météreau, E., Butera, L., Villeval, M. C., & Dreher, J. C. (2019). Neurocomputational mechanisms at play when weighing concerns for extrinsic rewards, moral values, and social image. *PLoS Biology, 17*, Article e3000283.
- Sáez, I., Zhu, L., Set, E., Kayser, A., & Hsu, M. (2015). Dopamine modulates egalitarian

- behavior in humans. *Current Biology*, 25, 912–919.
- Tesser, A., Gatewood, R., & Driver, M. (1968). Some determinants of gratitude. *Journal of Personality and Social Psychology*, 9, 233–236.
- Tsang, J. A. (2006a). Gratitude and prosocial behaviour: An experimental test of gratitude. *Cognition and Emotion*, 20, 138–148.
- Tsang, J. A. (2006b). The effects of helper intention on gratitude and indebtedness. *Motivation and Emotion*, 30, 198–204.
- van Den Haag, E. (1980). Punishment as a device for controlling the crime rate. *Rutgers L. Rev.* 33, 706–720.
- Vayness, J., Duong, F., & DeSteno, D. (2019). Gratitude increases third-party punishment. *Cognition and Emotion*, 1–8.
- Villarroya, O., & Hilferty, J. (2013). The neuroimaging of sacred values. *Annals of the New York Academy of Sciences*, 1299, 25–35.
- Vitaglione, G. D., & Barnett, M. A. (2003). Assessing a new dimension of empathy: Empathic anger as a predictor of helping and punishing desires. *Motivation and Emotion*, 27, 301–325.
- Wagner, U., N'Diaye, K., Ethofer, T., & Vuilleumier, P. (2011). Guilt-specific processing in the prefrontal cortex. *Cerebral Cortex*, 21, 2461–2470.
- Watkins, P. C., Scheer, J., Ovnicek, M., & Kolts, R. (2006). The debt of gratitude: Dissociating gratitude and indebtedness. *Cognition and Emotion*, 20, 217–241.
- Wood, A. M., Brown, G. D. A., & Maltby, J. (2011). Thanks, but I'm used to better: A relative rank model of gratitude. *Emotion*, 11, 175–180.
- Wood, A. M., Emmons, R. A., Algoe, S. B., Froh, J. J., Lambert, N. M., & Watkins, P. (2016). A dark side of gratitude? Distinguishing between beneficial gratitude and its harmful impostors for the positive clinical psychology of gratitude and well-being. *The Wiley handbook of positive clinical psychology* (pp. 137–151).
- Wood, A. M., Froh, J. J., & Geraghty, A. W. A. (2010). Gratitude and well-being: A review and theoretical integration. *Clinical Psychology Review*, 30, 890–905.
- Yu, H., Cai, Q., Shen, B., Gao, X., & Zhou, X. (2017). Neural substrates and social consequences of interpersonal gratitude: Intention matters. *Emotion*, 17, 589–601.
- Yu, H., Gao, X., Zhou, Y., & Zhou, X. (2018). Decomposing gratitude: Representation and integration of cognitive antecedents of gratitude in the brain. *Journal of Neuroscience*, 38, 4886–4898.
- Zhang, Y., Chen, Z. J., & Ni, S. (2020). The security of being grateful: Gratitude promotes risk aversion in decision-making. *The Journal of Positive Psychology*, 15, 285–291.
- Zhu, R., Jin, T., Shen, X., Zhang, S., Mai, X., & Liu, C. (2017). Relational utility affects self-punishment in direct and indirect reciprocity situations. *Social Psychology*, 48, 19–27.