

# UC Santa Barbara

## UC Santa Barbara Previously Published Works

### Title

Gratitude Is Morally Sensitive.

### Permalink

<https://escholarship.org/uc/item/5n441445>

### Journal

Personality and Social Psychology Bulletin, 50(3)

### Authors

Zhou, Yubo

Nussberger, Anne-Marie

Yu, Hongbo

### Publication Date


2024-03-01

### DOI

10.1177/01461672221092273

Peer reviewed

# Gratitude Is Morally Sensitive

Hongbo Yu<sup>1</sup> , Yubo Zhou<sup>1,2</sup>, and Anne-Marie Nussberger<sup>3</sup>

Personality and Social  
Psychology Bulletin  
2024, Vol. 50(3) 406–422  
© The Author(s) 2022



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/01461672221092273  
journals.sagepub.com/home/pspb



## Abstract

Helping acts, however well intended and beneficial, sometimes involve immoral means or immoral helpers. Here, we explore whether help recipients consider moral evaluations in their appraisals of gratitude, a possibility that has been neglected by existing accounts of gratitude. Participants felt less grateful and more uneasy when offered immoral help (Study 1,  $N = 150$ ), and when offered morally neutral help by an immoral helper (Study 2,  $N = 172$ ). In response to immoral help or helpers, participants were less likely to accept the help and less willing to strengthen their relationship with the helper even when they accepted it. Study 3 ( $N = 276$ ) showed that recipients who felt grateful when offered immoral help were perceived as less likable, less moral, and less suitable as close relationship partners than those who felt uneasy by observers. Our results demonstrate that gratitude is morally sensitive and suggest this might be socially adaptive.

## Keywords

gratitude, morality, appraisal, socially adaptive

Received November 2, 2021; revision accepted March 18, 2022

## Introduction

In 2019, a scandal in the American higher education system hit headlines all over the world: Wealthy parents were found guilty of bribing admission officials to get their children into prestigious universities across the country.<sup>1</sup> If you were one of the children, would you feel grateful for your parents' efforts to secure your admission to a top university?

According to a classic appraisal model of gratitude (Tesser et al., 1968), gratitude has three cognitive antecedents—a Helper's intrinsic intention to help the Recipient, their cost in providing the help, and the benefit to the Recipient ("Tesser Model" hereafter). All these antecedents seem to apply in the above example: First, the Helpers (the parents) help the Recipient (the children) not by accident or out of strategic considerations, but out of intrinsic motivation. Second, bribing incurred financial and reputational cost to the parents. Third, admittance to a prestigious university was certainly a valuable outcome from the children's perspectives. Still, intuitively one might hope that any feelings of gratitude on the side of the children were attenuated in view of the immoral ways by which their parents sought to secure their admission. The moral status of the helping behavior and the Helper is, however, not part of the Tesser Model.

A more recent model regarding the moral status of gratitude posits that gratitude functions as a "moral barometer," signaling to the subject that they receive a "provision of a benefit by another moral agent that enhances one's well-being" (McCullough et al., 2001). The authors make a distinction between *local* and *absolute* perceptions of morality,

where the former refers to the perspective of the recipient of the help and the latter refers to the perspective of an impartial observer. Importantly, the perceptions in these two senses can diverge: While a merchant who knowingly sells illegal weapons to a criminal may perceive their purchaser's business as moral and therefore feel grateful, an impartial observer will likely perceive that as immoral in the absolute sense. The authors argue that what matters for gratitude is morality in the local sense—as long as the recipient perceives a helping act/agent as moral, there would be no problem for them to feel genuine gratitude. An implicit logic behind this conclusion is that gratitude should be correlated with the morality of a helping act/agent as perceived by the recipient. However, most empirical research on gratitude to date has assumed helping behaviors and Helpers to be moral (for reviews, see Emmons & McCullough, 2004; Gulliford et al., 2013; Roberts, 2015). As illustrated by the above example, this may not always be the case in real-world helping instances. Here, we propose and provide evidence for an additional antecedent in the appraisal model of gratitude, namely, the moral evaluation of the helping behavior and the Helper.

<sup>1</sup>University of California, Santa Barbara, USA

<sup>2</sup>University of Pennsylvania, Philadelphia, USA

<sup>3</sup>Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, Germany

### Corresponding Author:

Hongbo Yu, Department of Psychological & Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106, USA.  
Email: hongbo.yu@psych.ucsb.edu

From a functionalist perspective, gratitude is adaptive in the sense that it signals to the Recipient the presence of a responsive social partner and motivates the Recipient to maintain and strengthen the relationship with this high-quality social partner (Algoe, 2012; Algoe et al., 2013). To identify a high-quality, responsive social partner, focusing solely on the objective benefit provided is not enough; the motives or intentions behind the help and the manner in which the benefit is provided matter. These appraisals are reflected in the Tesser Model—knowing why the Helper helps (i.e., intention) and how much they are willing to invest in the help (i.e., cost) is critical for beneficiaries to infer the responsiveness of the Helper as a social partner (Algoe & Haidt, 2009). However, helping behaviors rarely take place in a social vacuum; the Helper's stable traits as a person should also be informative about how responsive they are as a social partner. Indeed, research in moral psychology has consistently demonstrated that moral character is a reliable predictor of whether someone is a valuable or a threatening social partner and is more influential in impression-development than other traits, such as sociability (Brambilla et al., 2021; Goodwin et al., 2014). Here, we hypothesize that a Recipient evaluates the moral status of the Helper and the help they offer. Perceptions of moral status regarding the helping act and actor will inform inferences about the intention to help (e.g., benevolent vs. strategic), which in turn will modulate the Recipient's gratitude response. We further hypothesize that morally sensitive gratitude leads beneficiaries to perceive Helpers as more likable when they behave morally, thereby playing a functional role in conferring socially adaptive values (Sznycer, 2018).

Gratitude is typically referred to as a moral emotion or moral sentiment, primarily because it usually leads to prosocial and approaching behaviors, motivations, and tendencies (Algoe, 2012; DeSteno et al., 2010; Emmons & McCullough, 2004; Ma et al., 2017; McCullough et al., 2001; Tsang, 2006; Watkins, 2013). For example, it has been consistently demonstrated that help/gift recipients react prosocially toward the Helper, third-party individuals, the environment, and the future self (Beeler-Duden & Vaish, 2020; Dickens & DeSteno, 2016; Kates & DeSteno, 2020; McCullough et al., 2001). Gratitude also motivates the grateful person to care more about the well-being of the Helper and strengthen the relationship to the Helper (Algoe, 2012; Bartlett et al., 2012; Lambert et al., 2010). These prosocial and relationship-building tendencies dissociate propositional gratitude from other related emotions, such as indebtedness and propositional gratitude (or appreciation; Watkins et al., 2006). Gratitude also enhances Recipients' communal (as opposed to exchange) relationship to the Helper (Bartlett et al., 2012; Lambert et al., 2010), which is characterized by noncontingent support, goodwill, and willingness to maintain long-term relationships (Clark & Mills, 2011; Manela, 2016). These behavioral and motivational tendencies can thus provide an independent measure of gratitude, in addition to self-reported gratitude.

We carried out three vignette-based studies to investigate the role of moral status of helping acts and actors in appraisals of gratitude, and the potential social adaptive value of exhibiting such moral sensitivity. In the first two studies, we manipulated the moral status of the proposed helping act (Study 1) and of the helping actor (Study 2), and examined how these factors modulate (a) gratitude and related social emotions, (b) beneficiaries' willingness to accept the proposed help, and (c) beneficiaries' motivations and attitudes toward benefactors. In Study 3, we manipulated the beneficiary's emotion when facing a morally problematic help and measured observers' perceptions of and preference for the Recipient. In all of our studies, we included both an American sample and a Chinese sample. While we focus on observations that generalize across our culturally distinct samples, we also discuss cultural specificities observed in our data. In these studies, we report all measures, manipulations, and exclusions. Sample size was determined before any data analysis.

## Studies 1 and 2

### Method

All data, materials, preregistration documents, and analysis codes for all of our studies can be accessed at the Open Science Framework page (<https://osf.io/64sqn>). All the studies were approved by the Human Subject Committee of the first author's institution (Protocol No. 2-21-0743). American participants recruited through Prolific were paid at a rate of US\$8/hour; Chinese participants recruited from Credamo were paid at a rate of ¥15/hour (standard in this context).

### Participants

**American samples.** For Study 1, we preregistered to recruit 65 participants via the online platform Prolific Academic (<https://www.prolific.co/>). Specifically, results from a pilot data set not reported here (collected from a Chinese university using a different set of vignettes) indicated that the correlation between the difference in moral judgment of helping behavior (i.e., morally neutral > morally problematic) and the difference in gratitude is .44. To detect such an effect with 95% power, we need a total sample of 57. To account for potential exclusion of participants based on their performance in attention check questions, 65 participants who self-reported as American residents and currently lived in the United States were recruited. Among them, four were excluded because they failed attention checks, leaving a final sample of  $N = 61$  (35 females;  $M_{\text{age}} = 28.2$  years,  $SD = 8.9$ , age range = 18–55; 41.7% with a bachelor's or higher degree; 65.0% self-identified as White/Caucasian). For Study 2, 85 participants who self-reported as American residents and currently lived in the United States were recruited from Prolific, among whom seven were excluded due to failing attention checks, leaving a final sample of

$N = 78$  (35 females;  $M_{\text{age}} = 31.7$  years,  $SD = 11.6$ , age range = 18–65; 59.0% with a bachelor's or higher degree; 68.0% self-identified as White/Caucasian). Study 2 was exploratory in nature and the sample size was determined so that it would be comparable with Study 1 and affordable by the available research resource.

**Chinese samples.** For Study 1, to be consistent with the American sample, we preregistered to recruit a sample of 70 participants via the online platform Credamo (<https://www.credamo.com/>). Due to a technical error, 92 Chinese participants were recruited. Among them three were excluded due to failure in the attention checks, leaving a final sample of  $N = 89$  (53 females;  $M_{\text{age}} = 28.2$  years,  $SD = 4.9$ , age range = 20–46; 100% with a bachelor's or higher degree, 98.9% self-identified as Han Chinese). Only including the first 70 participants in the analysis led to an identical pattern of results as using the final sample. We therefore only reported the results based on the full final sample here. For Study 2, we preregistered a sample of 90 participants to be comparable with the American sample. Ninety-six Chinese participants were recruited from Credamo, among whom two were excluded due to failure in the attention checks, leaving a final sample of  $N = 94$  (54 females;  $M_{\text{age}} = 32.7$  years,  $SD = 5.7$ , age range = 20–52; 97.9% with a bachelor's or higher degree, 94.4% self-identified as Han Chinese).

**Experimental design and procedure.** In Study 1, we presented participants with two vignettes, each describing an interpersonal helping situation involving two figures, a Helper and a Recipient. We used a within-subjects manipulation of the described helping act's moral status (morally neutral vs. morally problematic) according to which participants saw two versions of each vignette. For example, the morally problematic versions of one vignette described a Helper helping the Recipient to secure a medical appointment by canceling another patient's appointment, whereas the morally neutral version of the vignette described a Helper helping the Recipient to secure a medical appointment without harming someone else (please see Supplementary Methods for the full materials). Study 2 followed a similar design, but the manipulation of moral status related to the person of the Helper. For example, the morally problematic version described a Helper who was unfaithful to their romantic partner, whereas the morally neutral version described the Helper as being good at learning foreign languages (i.e., a morally neutral trait). In both studies, participants were asked to imagine themselves as the Recipient in the vignettes. The participants read each vignette exactly once, but in different conditions. The mapping between vignettes and experimental conditions was counterbalanced across participants. Although similar in their overall narrative, the vignettes for the Chinese and American samples were tailored to their respective cultures and presented in the respective languages. Because we were not interested in the effect of vignette, we

included vignette as a random intercept in our regression analyses. After reading each vignette, the participants reported their feelings of gratitude, uneasiness, and indebtedness (among a few other filler terms) that were measured on a series of explanatory and control variables (see "Measurements" section).

### Measurements

**Self-reported emotions.** As our primary dependent variable, we measured three emotions related to being the Recipient of help, namely, feeling grateful, indebted, and uneasy. In addition, participants reported their emotions for three filler items (i.e., disgust, outrageous, and ashamed). Participants indicated the extent to which they felt these emotions on rating scales using a slider (0 = *not at all*, 100 = *extremely*).

**Willingness to accept the help.** We assessed participants' willingness to accept the proposed help as a behavioral manifestation of the participants' attitude toward the proposed help. In particular, we asked how likely they would accept or reject the proposed help on a 4-point scale (1 = *definitely reject*, 2 = *more likely to reject than accept*, 3 = *more likely to accept than reject*, 4 = *definitely accept*). To better understand the reasons underlying participants' choices, we asked them to briefly describe why they tended to accept or reject the help. Two independent coders evaluated the free response data along three dimensions. For Study 1, the three dimensions were whether participants mentioned (a) potential harm to others and feelings of guilt, (b) violation of moral norms, and (c) benefits to the Recipient (i.e., the participants themselves). For Study 2, the three dimensions were whether the participants mentioned (a) morally good traits of the Helper, (b) morally bad traits of the Helper, and (c) benefits to the Recipient. The interrater reliability (IRR) was satisfactory ( $>.8$ ). In the cases where the two raters disagreed, the first author (H.Y.), who was not one of the two coders, read the free responses and made final decisions.

**Moral evaluations of the helping behavior and the helper.** We measured participants' moral evaluations of the helping behavior by using a single item, "Overall, how morally right or wrong is [Helper's name]'s suggested help?" (from  $-50 = \textit{totally wrong}$  to  $50 = \textit{totally right}$ ). This was intended as a manipulation check for Study 1. We expected that participants to judge the helping behavior in the morally problematic condition to be significantly less morally right than that in the morally neutral condition. We also assessed participants' perceptions of the Helper's moral traits. Moral traits were measured using the three items: "moral," "kind," and "trustworthy" (cf. Yu et al., 2021). Participants indicated how much they thought the Helper could be described by those traits on analogue scales (0 = *not at all*, 100 = *extremely*). The internal reliability of the moral traits was high across conditions for both studies ( $\alpha_s > .83$  for Study 1,  $\alpha_s > .89$  for Study 2). We therefore averaged across the three items

measures to obtain a composite measure of the Helper's moral character. This was used as a mediator in the mediation analysis (see "Mediation Analysis" section). The moral trait evaluation also functioned as a manipulation check for Study 2. We expected that the participants judged the Helper in the morally problematic condition to be significantly less moral than the Helper in the morally neutral condition.

**Antecedents of gratitude in the Tesser Model.** We measured the three antecedents of gratitude proposed in the Tesser Model as control variables (Tesser et al., 1968). These included the Helper's effort in providing the help ("How much effort do you think [Helper's name] would have to invest if you accepted their offer to help?" 0 = *no effort at all*, 100 = *huge effort*), the value of the help to the Recipient ("To what extent do you think you would benefit from [the Helper's name] suggested help if you accepted it?" 0 = *not at all*, 100 = *substantially*), and the Helper's intention to help ("To what extent do you think [Helper's name] has a good intention to help you?" 0 = *not at all* to 100 = *extremely*).

**Gratitude-related behavioral tendencies.** We measured three behavioral tendencies: pressure to repay ("Imagine that you accepted [Helper's name]'s help. How obligated or pressured would you feel to pay something back to [Helper's name]?" 0 = *not at all*, 50 = *somewhat*, 100 = *extremely*), caring about the Helper's well-being ("Would you care more or less about [Helper's name]'s well-being after he suggests the way that he could help?" 0 = *care a lot less*, 50 = *no change*, 100 = *care a lot more*), and strengthening the relationship with the Helper ("How likely would you be to strengthen your relationship with [Helper's name]?" 0 = *not at all likely*, 50 = *somewhat likely*, 100 = *extremely likely*).

## Results

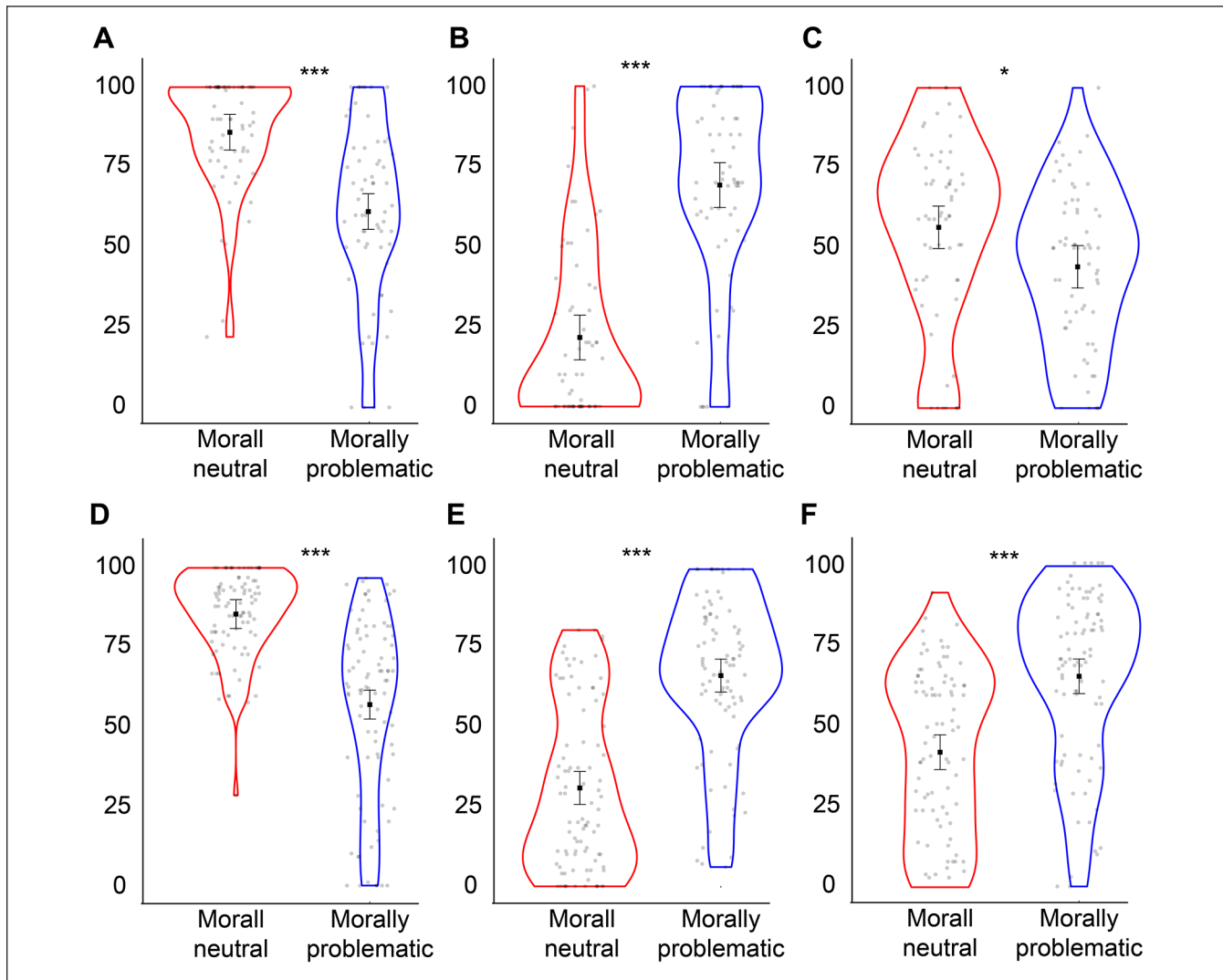
**Manipulation check.** For Study 1, we examined participants' moral evaluations of the helping behavior. As expected, participants judged the helping behavior in the morally problematic condition ( $-21.0 \pm 2.0$ ) to be significantly less morally right than that in the morally neutral condition ( $25.4 \pm 1.8$ ;  $B = -46.4 \pm 2.6$ , 95% confidence interval [CI] =  $[-51.5, -41.3]$ ,  $t = -17.78$ ,  $p < .001$ ). This pattern was observed in both the Chinese sample and the American sample, and there was no significant country-by-condition interaction ( $B = 3.4 \pm 5.3$ , 95% CI =  $[-7.0, 13.7]$ ,  $t = 0.63$ ,  $p = .526$ ; Supplementary Figure S1). We further examined the effects of our experimental manipulation and participant cohort (Chinese vs. American) on the antecedents in the Tesser Model (Supplementary Figure S1). Participants perceived that the Helper had better intention ( $B = 22.5 \pm 2.3$ , 95% CI =  $[18.0, 27.0]$ ,  $t = 9.80$ ,  $p < .001$ ) and that the help was more beneficial ( $B = 14.1 \pm 3.1$ , 95% CI =  $[8.1, 20.2]$ ,  $t = 4.54$ ,  $p < .001$ ) in the morally neutral condition than in the morally problematic condition. These were true for both the Chinese and the American sample. Participants also

perceived that the Helper exerted more effort in the morally neutral condition than in the morally problematic condition ( $B = 10.5 \pm 2.8$ , 95% CI =  $[5.0, 15.9]$ ,  $t = 3.78$ ,  $p < .001$ ), although the country-by-condition interaction was significant ( $B = 14.1 \pm 4.4$ , 95% CI =  $[5.6, 22.6]$ ,  $t = 3.24$ ,  $p = .001$ ). While the American participants' effort judgment did not differ between conditions ( $B = -4.8 \pm 2.9$ , 95% CI =  $[-10.7, 0.9]$ ,  $t = -1.66$ ,  $p = .102$ ), the Chinese participants perceived that the help was significantly more effortful in the morally neutral condition ( $B = 10.5 \pm 2.9$ , 95% CI =  $[4.6, 16.3]$ ,  $t = 3.51$ ,  $p < .001$ ).

For Study 2, participants judged the Helper in the morally problematic condition ( $53.7 \pm 1.6$ ) to be significantly less moral than the Helper in the morally neutral condition ( $80.2 \pm 1.1$ ;  $B = -26.5 \pm 1.8$ , 95% CI =  $[-30.0, -22.9]$ ,  $t = -14.71$ ,  $p < .001$ ). This pattern remained significant after controlling for participants' moral judgment of the helping behavior ( $B = -23.6 \pm 1.7$ , 95% CI =  $[-26.9, -20.2]$ ,  $t = -13.70$ ,  $p < .001$ ). This effect was observed both in the Chinese ( $B = -20.1 \pm 2.3$ , 95% CI =  $[-24.5, -15.6]$ ,  $t = -8.79$ ,  $p < .001$ ) and in the American sample ( $B = -28.2 \pm 2.5$ , 95% CI =  $[-33.1, -23.3]$ ,  $t = -11.09$ ,  $p < .001$ ), although the effect was stronger in the American sample than in the Chinese sample (country-by-condition interaction:  $B = -8.2 \pm 3.4$ , 95% CI =  $[-14.7, -1.6]$ ,  $t = -2.41$ ,  $p = .017$ ; Supplementary Figure S2). Like in Study 1, we examined the effects of our experimental manipulation and participants cohort on the antecedents in the Tesser Model (Supplementary Figure S2). Participants perceived that the Helper had better intention ( $B = 15.5 \pm 2.7$ , 95% CI =  $[10.2, 20.8]$ ,  $t = 5.72$ ,  $p < .001$ ) and that the help was more beneficial in the morally neutral condition than in the morally problematic condition ( $B = 8.9 \pm 2.5$ , 95% CI =  $[4.1, 13.8]$ ,  $t = 3.64$ ,  $p < .001$ ). These effects were observed both in the Chinese sample and in the American sample. There was no significant country-by-condition interaction (intention:  $B = -0.5 \pm 4.0$ , 95% CI =  $[-8.4, 7.4]$ ,  $t = -0.13$ ,  $p = .900$ ; benefit:  $B = -3.5 \pm 3.6$ , 95% CI =  $[-10.7, 3.6]$ ,  $t = -0.97$ ,  $p = .335$ ). For perceived Helper effort, neither the main effect of condition ( $B = -2.6 \pm 2.2$ , 95% CI =  $[-6.8, 1.7]$ ,  $t = -1.19$ ,  $p = .237$ ) nor the country-by-condition interaction ( $B = -2.6 \pm 3.2$ , 95% CI =  $[-8.9, 3.7]$ ,  $t = -0.82$ ,  $p = .414$ ) was significant.

In sum, given that some of the antecedents in the Tesser Model were different between conditions, in the following analysis, we reported both the analysis that statistically controlled for these variables and the analysis that did not control these variables (see below). Controlling these variables did not significantly impact our key results, suggesting that our manipulations of the moral status of the helping behavior and the Helper cannot be fully explained by the Tesser Model.

**Gratitude appraisals and help-reception-related emotions.** For Study 1, we first examined the effect of experimental condition on gratitude and other emotions related to help-reception (namely, uneasiness and indebtedness). As preregistered, we



**Figure 1.** Self-reported emotions as a function of condition in Study 1: (A) Gratitude (American), (B) uneasiness (American), (C) indebtedness (American), (D) gratitude (Chinese), (E) uneasiness (Chinese), (F) indebtedness (Chinese).

Note. Gratitude was lower in the morally problematic condition than the morally neutral condition for both the American (A) and Chinese sample (D).

Conversely, uneasiness was significantly higher in the morally problematic condition than the morally neutral condition for both the American (B) and Chinese sample (E). Indebtedness exhibited opposite patterns in the American (C) and the Chinese samples (F). Error bars show the 95% confidence intervals.

estimated a linear mixed effect model for each of the above emotions based on the American-Chinese combined sample. In these models, experimental condition was included as the predictor of interest, participants' demographics (i.e., age, gender, studentship, highest level of education, socioeconomic status) were included as covariates of no interest, and country, participant, and vignette version were included as random intercepts. Participants expressed less gratitude in the morally problematic condition ( $58.6 \pm 2.2$ ) than in the morally neutral condition ( $86.2 \pm 1.2$ ;  $B = -27.6 \pm 2.4$ , 95% CI =  $[-32.3, -22.9]$ ,  $t = -11.48$ ,  $p < .001$ ; Figure 1A and D). Consistent with these observations, participants expressed more uneasiness in the morally problematic condition ( $68.0 \pm 2.1$ ) than in the morally neutral condition ( $26.5 \pm 2.2$ ;  $B =$

$41.5 \pm 3.0$ , 95% CI =  $[35.7, 47.4]$ ,  $t = 13.89$ ,  $p < .001$ ; Figure 1B and E).

Indebtedness exhibited a similar pattern as uneasiness in the combined sample ( $B = 8.9 \pm 3.2$ , 95% CI =  $[2.7, 15.1]$ ,  $t = 2.81$ ,  $p = .005$ ). However, inspecting the patterns of the Chinese and American samples separately (Figure 1C and F), it seemed that the helping act's moral status was inverted in the American sample. To formally test this observation, we ran another regression model including additional fixed effects for country and its interaction with condition. This model, which showed a significantly better fit with a Bayesian information criterion (BIC) of 2,847 compared with the model without these additional fixed effects (BIC = 2,879), indicated that the main effect of condition was

qualified by a significant interaction with country ( $B = -35.1 \pm 5.9$ , 95% CI =  $[-46.6, -23.5]$ ,  $t = -5.95$ ,  $p < .001$ ). Specifically, for American participants, the morally neutral help elicited more indebtedness than the morally problematic help ( $B = 12.3 \pm 4.7$ , 95% CI =  $[3.2, 21.4]$ ,  $t = 2.62$ ,  $p = .011$ ), a pattern similar to gratitude. In contrast, Chinese participants showed the opposite pattern, expressing more indebtedness in the morally problematic condition than the morally neutral condition ( $B = 23.2 \pm 3.6$ , 95% CI =  $[16.1, 30.2]$ ,  $t = 6.48$ ,  $p < .001$ ). We note that this cross-culture comparison analysis was not reported in the preregistration and therefore was exploratory in nature.

As a next step, we sought to demonstrate that the effect of the helping behavior's moral status on gratitude (as well as uneasiness and indebtedness) is independent from the traditionally considered antecedents proposed in the Tesser Model (Tesser et al., 1968). To this end, we ran another set of regression models that statistically controlled for perceived Helper intention and effort as well as benefit to the Recipient. The effect of experimental condition on gratitude remained significant in this model ( $B = -14.9 \pm 2.1$ , 95% CI =  $[-19.0, -10.8]$ ,  $t = -7.06$ ,  $p < .001$ ), indicating that the helping behavior's moral status has an impact on gratitude over and above the antecedents of gratitude in the Tesser Model. Adding these control variables did not change the pattern for uneasiness ( $B = 32.9 \pm 3.1$ , 95% CI =  $[26.9, 39.0]$ ,  $t = 10.65$ ,  $p < .001$ ), nor for indebtedness, where the interaction between country and condition remained significant ( $B = -32.8 \pm 6.0$ , 95% CI =  $[-44.8, -21.3]$ ,  $t = -5.46$ ,  $p < .001$ ).

For Study 2, we ran a similar set of regression models as in Study 1. Participants expressed less gratitude toward the Helper with a morally problematic trait ( $70.3 \pm 1.6$ ) than the Helper with a morally neutral trait ( $84.9 \pm 1.2$ ;  $B = -14.6 \pm 1.6$ , 95% CI =  $[-17.8, -11.4]$ ,  $t = -9.00$ ,  $p < .001$ ; Figure 2A and D). Also paralleling the results of Study 1, participants expressed more uneasiness toward a morally problematic Helper ( $45.5 \pm 2.0$ ) than a morally neutral Helper ( $23.7 \pm 1.8$ ;  $B = 21.8 \pm 2.5$ , 95% CI =  $[16.9, 26.7]$ ,  $t = 8.70$ ,  $p < .001$ ; Figure 2B and E). Different to Study 1, participants across both national samples felt less indebted in the morally problematic condition ( $41.7 \pm 2.0$ ) than in the morally neutral condition ( $52.3 \pm 1.9$ ;  $B = -10.6 \pm 2.4$ , 95% CI =  $[-15.3, -6.0]$ ,  $t = -4.49$ ,  $p < .001$ ; Figure 2C and F). Including the antecedents of gratitude in the Tesser Model and moral evaluation of the proposed helping behavior did not change these results ( $|ts| > 2.48$ ,  $ps < .014$ ), indicating the independent contribution of Helper's moral character to Recipient's help reception-related emotion.

**Willingness to accept the help.** How did the moral status of the proposed help affect Recipients' willingness to accept it? Figure 3A and B display the distribution of "willingness to accept the help" as a function of experimental condition for the American and the Chinese participants. For statistical

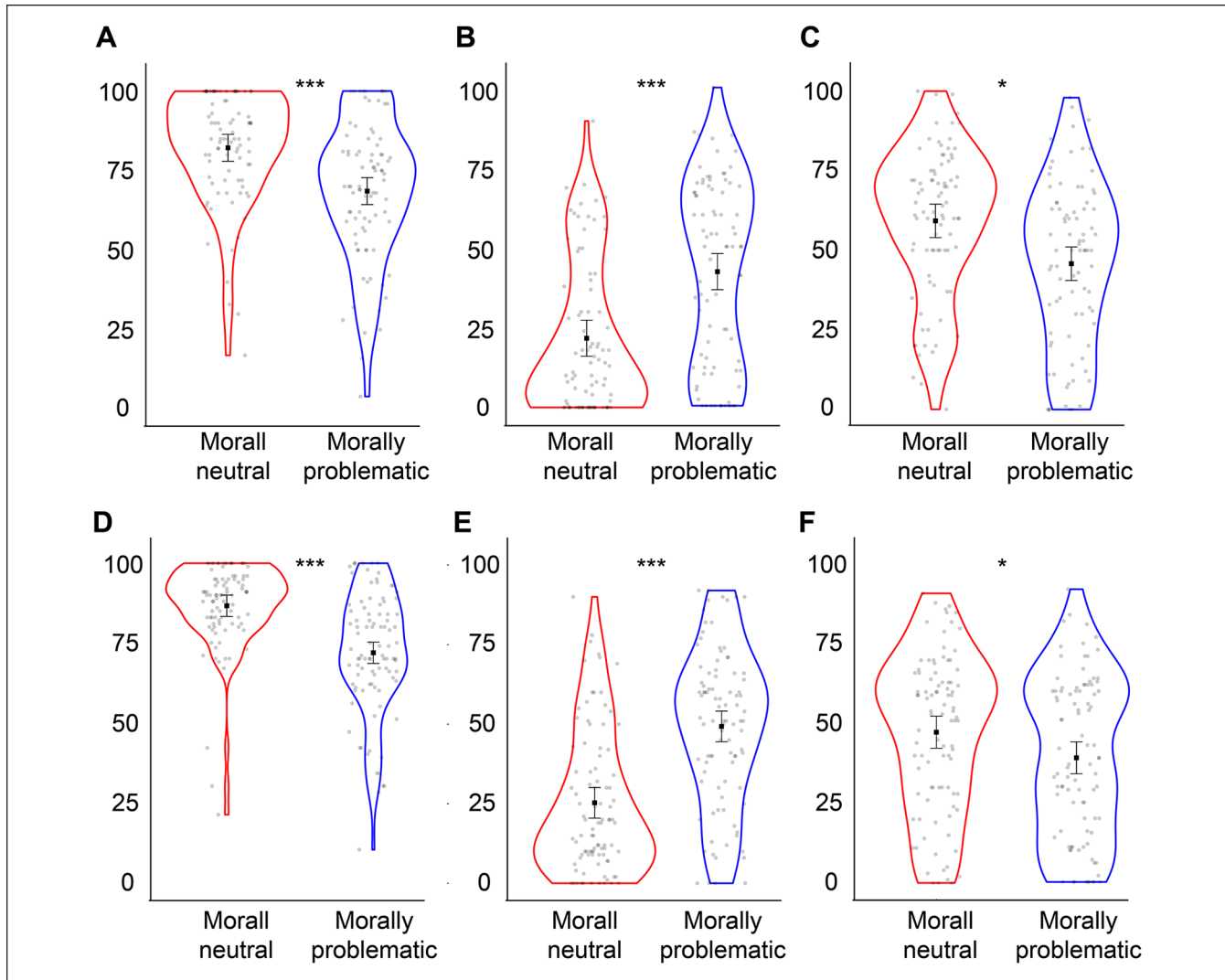
analysis, we collapsed "more likely to accept than reject" and "definitely accept" into one choice ("accept" = 1), and "more likely to reject than accept" and "definitely reject" into another choice ("reject" = 0; cf. Hutcherson et al., 2015). Similar patterns were obtained when we used the original scores (see Supplementary Results).

For Study 1, we ran a generalized mixed effect model, including condition as the predictor of interest and participants' demographics (i.e., age, gender, highest level of education, socioeconomic status, studentship) as fixed effect covariates. We also included country, participant, and vignette version as random intercepts. We found that participants were less likely to accept the help when it was morally problematic ( $B = -3.86 \pm 0.42$ , 95% CI =  $[-4.69, -3.04]$ ,  $z = -9.21$ ,  $p < .001$ ). Including the Tesser Model's antecedents of gratitude as additional covariates did not change the pattern of results ( $B = -3.70 \pm 0.47$ , 95% CI =  $[-4.61, -2.78]$ ,  $z = -7.91$ ,  $p < .001$ ).

To better understand the reasons behind participants' willingness to accept, we added fixed effects for mentions of the three rationale dimensions extracted from participants' free responses (i.e., harming others and guilt, violation of moral norms, benefits to oneself) and for participants' demographics (Figure 3C). We found that mentions of the first two dimensions predicted rejections of help (harm and guilt:  $B = -1.11 \pm 0.44$ , 95% CI =  $[-1.96, -0.23]$ ,  $z = -2.49$ ,  $p = .013$ ; norm violation:  $B = -1.48 \pm 0.47$ , 95% CI =  $[-2.40, -0.55]$ ,  $z = -3.12$ ,  $p = .002$ ). In contrast, mentions of benefits to oneself increased the willingness to accept the help ( $B = 1.72 \pm 0.46$ , 95% CI =  $[0.82, 2.62]$ ,  $z = 3.74$ ,  $p < .001$ ). Including the antecedents of gratitude in the Tesser Model as additional covariates did not change the patterns of results ( $|z| > 2.16$ ,  $ps < .03$ ). Together, these results suggest that the helping behavior's moral status does not only modulate Recipients' gratitude appraisals but also their willingness to accept the help on offer.

Next, we ran the same analyses for Study 2, where the Helper's moral status varied across conditions (Figure 4A and B). Here, we found that participants were less likely to accept the help when it was proposed by a morally problematic Helper ( $B = -1.56 \pm 0.38$ , 95% CI =  $[-2.30, -0.81]$ ,  $z = -4.12$ ,  $p < .001$ ). However, including the antecedents of gratitude in the Tesser Model as additional fixed effect covariates rendered the effect of condition nonsignificant ( $B = -0.42 \pm 0.45$ , 95% CI =  $[-4.64, -2.77]$ ,  $z = -0.93$ ,  $p = .353$ ). This might indicate that Helpers' moral traits might have a less impactful and less independent effect on Recipients' decision as to whether to accept the Helper's proposed help, over and above the other antecedents of gratitude.

Similar to Study 1, we further explored how participants' self-reported concerns were related to their hypothetical decision of accepting or rejecting the help. As a reminder, in this case the three rationales included allusions to the Helper's moral traits, the Helper's immoral traits, and benefits to oneself (Figure 4C). We found that considerations of



**Figure 2.** Self-reported emotions as a function of condition in Study 2: (A) Gratitude (American), (B) uneasiness (American), (C) indebtedness (American), (D) gratitude (Chinese), (E) uneasiness (Chinese), (F) indebtedness (Chinese).

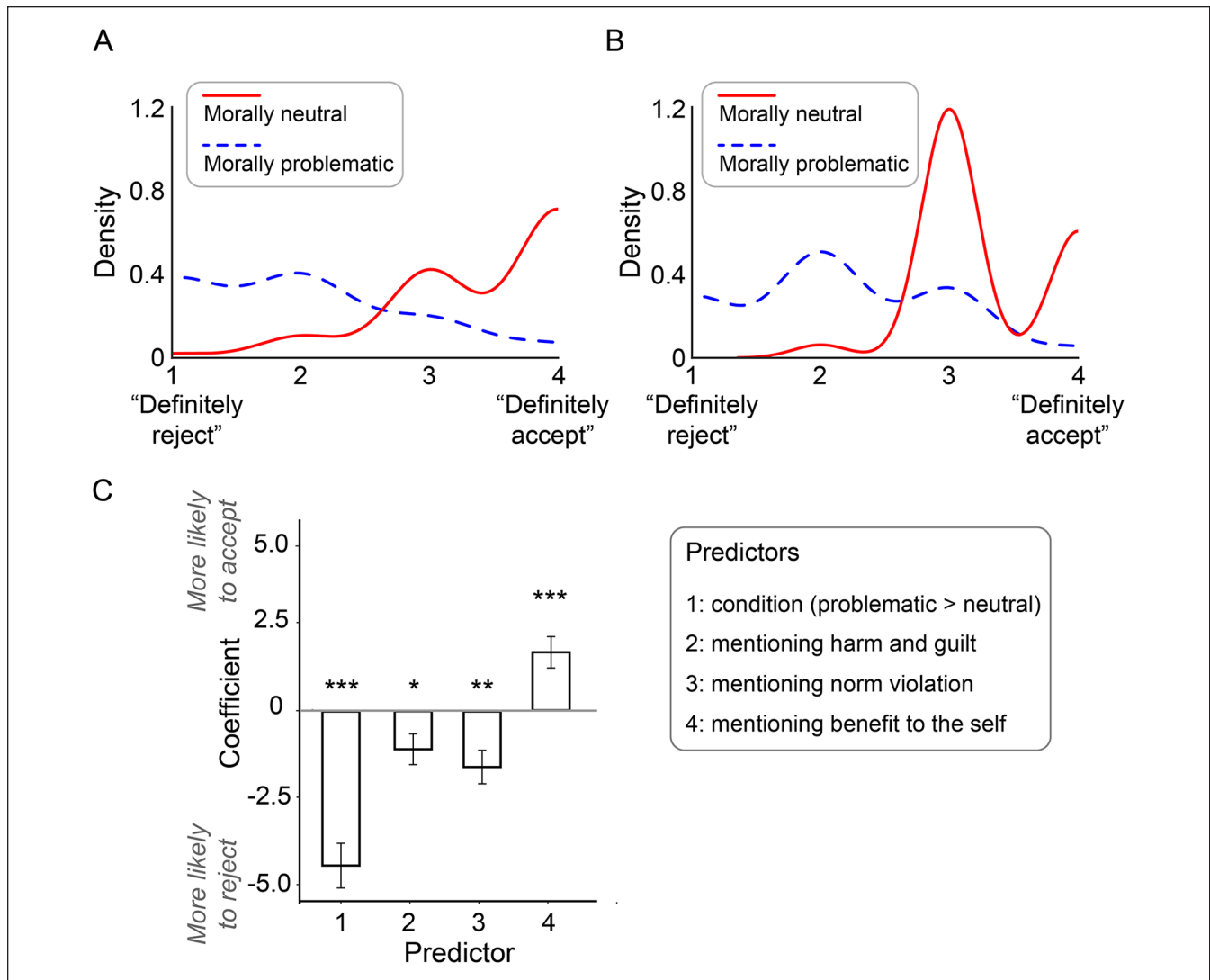
Note. Gratitude was lower in the morally problematic condition than the morally neutral condition for both the American (A) and Chinese sample (D). Conversely, uneasiness was significantly higher in the morally problematic condition than the morally neutral condition for both the American (B) and Chinese sample (E). Indebtedness exhibited a similar pattern as gratitude for both the American sample (C) and Chinese (F) sample. Error bars show the 95% confidence intervals.

the Helper's moral traits and of the benefits to oneself made participants more likely to accept the help (moral traits:  $B = 2.34 \pm 0.76$ , 95% CI = [0.85, 3.82],  $z = 3.08$ ,  $p = .002$ ; benefits to the self:  $B = 2.51 \pm 0.56$ , 95% CI = [1.40, 3.62],  $z = 4.44$ ,  $p < .001$ ). In contrast, participants tended to reject the help if they mentioned the Helper's immoral traits ( $B = -2.21 \pm 0.67$ , 95% CI = [-3.92, -0.90],  $z = -3.30$ ,  $p < .001$ ). Further including other antecedents of gratitude (i.e., effort, benefit, motive) and the moral evaluation of the proposed helping behavior as covariates did not substantially change the pattern of results ( $|z| > 1.95$ ,  $ps < .052$ ).

**Gratitude-related behavioral tendencies.** We next examined how the moral status of helping behavior and the Helper

modulated gratitude-related behavioral motivations or tendencies. For Study 1, using similar regression models as above, we found that participants were less likely to strengthen their relationship with the Helper ( $B = -6.92 \pm 1.85$ , 95% CI = [-10.63, -3.40],  $t = -3.75$ ,  $p < .001$ ) and cared less about the Helper's well-being ( $B = -3.40 \pm 1.58$ , 95% CI = [-6.50, -0.34],  $t = -2.15$ ,  $p = .033$ ) when the proposed help was morally problematic than morally neutral. However, they felt similar level of pressure to repay the Helper immediately ( $B = 1.39 \pm 2.45$ , 95% CI = [3.12, 13.48],  $t = 0.57$ ,  $p = .57$ ). Including the Tesser Model's antecedents as additional covariates did not change the patterns of the first two effects ( $|ts| > 2.25$ ,  $ps < .025$ ) and made the third effect (i.e., pressure to repay)





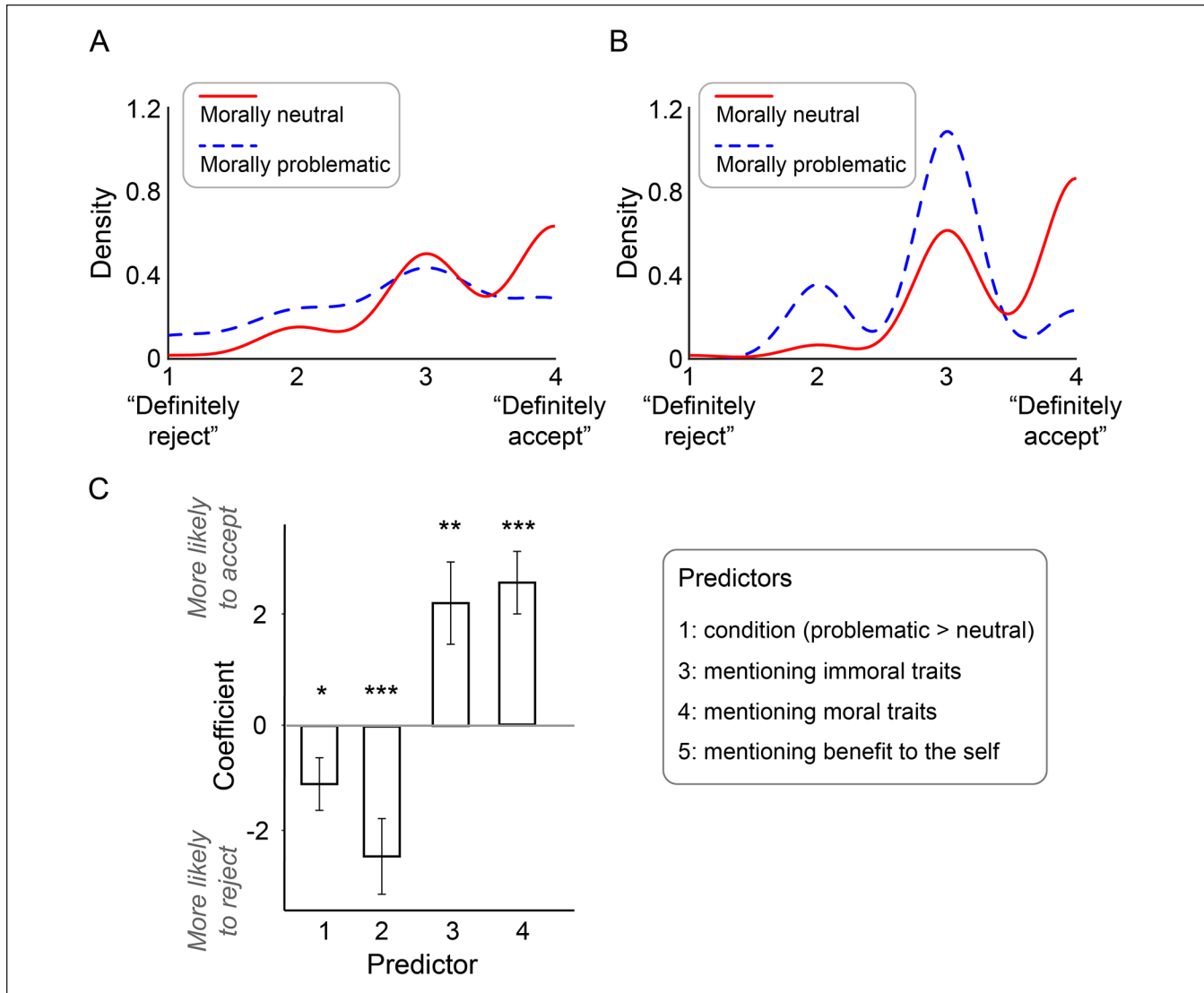
**Figure 3.** Willingness to accept the proposed help as a function of condition and self-reported rationales (Study 1): (A) Willingness to accept (American), (B) willingness to accept (Chinese), (C) willingness to accept as a function of self-reported rationales and condition. Note. Participants were less willing to accept the help, both in the American (A) and in the Chinese (B) samples, when the proposed helping behavior was morally problematic than when it was morally neutral. (C) Participants' free responses about their rationale of accepting or rejecting the proposed help suggested that concerns about potential harm to others and guilt, and violation of moral norms, made the participants more likely to reject the help, whereas considerations of the benefits to oneself and one's own urgent needs made the participants more likely to accept the help. Error bars represent s.e.m.

stronger ( $B = 8.37 \pm 2.65$ , 95% CI = [3.12, 13.48],  $t = 3.17$ ,  $p = .002$ ).

For Study 2, similarly, participants were less likely to strengthen the relationship with ( $B = -20.40 \pm 1.98$ , 95% CI = [-24.29, -16.50],  $t = -10.29$ ,  $p < .001$ ) and cared less about the well-being of ( $B = -11.65 \pm 1.65$ , 95% CI = [-14.89, -8.41],  $t = -7.06$ ,  $p < .001$ ) a morally problematic Helper compared with a morally neutral Helper. Including the Tesser Model's antecedents and the helping behavior's moral status as additional covariates did not change the pattern of results ( $|t| > 2.79$ ,  $ps < .006$ ). However, the pressure to pay back was not significantly different between the

morally neutral Helper condition and morally problematic Helper condition ( $B = -0.53 \pm 1.83$ , 95% CI = [-4.13, 3.07],  $t = -0.29$ ,  $p = .773$ ).

**Individual difference analysis.** To further demonstrate the influence of moral judgment of helping behavior and Helper on Recipient's gratitude and subsequent motivations, we adopted an individual difference approach as pre-registered. Specifically, we calculated difference scores of the key dependent variables (i.e., gratitude, uneasiness, willingness to accept the help, care about the Helper's well-being, and the willingness to improve the relationship with



**Figure 4.** Willingness to accept the proposed help as a function of the Helper's moral status (Study 2): (A) Willingness to accept (American), (B) willingness to accept (Chinese), (C) willingness to accept as a function of self-reported rationales and condition. Note. Participants were more willing to reject the help, both in the American (A) and in the Chinese (B) samples, when the Helper had a morally problematic trait than when the Helper had a morally neutral trait. (C) Participants' free responses regarding their rationale of accepting or rejecting the proposed help suggested that concerns about the Helper's morally problematic trait made them more likely to reject the proposed help, whereas reference to the morally good trait of the Helper and the benefits to oneself made the participants more likely to accept the proposed help. Error bars represent s.e.m.

the Helper), the explanatory variables (i.e., moral judgment of the helping behavior for Study 1, moral evaluation of the Helper for Study 2), and control variables (the three antecedents in the Tesser Model for both studies and the moral judgment of the helping behavior for Study 2) between the two conditions. Then, we computed Spearman partial correlations between the difference scores of the dependent variables and the difference scores of the explanatory variables, controlling for the difference scores of the control variables and demographic variables (i.e., age, gender, education, subjective socioeconomic status). All the correlations were in the predicted direction and statistically

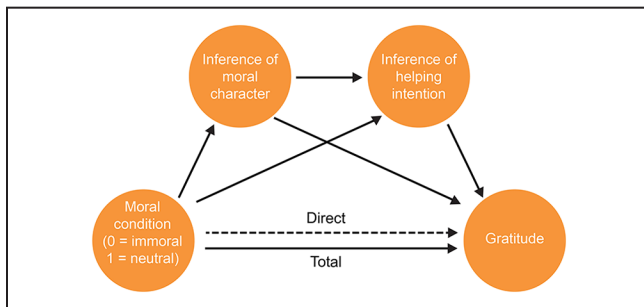
significant (see Table 1), buttressing our condition-wise analysis reported above.

**Mediation analysis.** As summarized in the "Introduction" section, the Tesser Model identifies three help-centered antecedents of gratitude, namely, the Helper's intention and cost in providing the help, and the benefits to the Recipient (Tesser et al., 1968). None of these antecedents are directly concerned with the Helper's moral character. This is in stark contrast with recent findings from moral psychology showing that moral character inference is rapid and exerts influences on other aspects of person perception (Brambilla et al.,

**Table 1.** Results of the Individual Difference Analysis.

Morality effect on key DVs	Difference in moral judgment of helping behavior (Study 1) N = 150	Difference in moral evaluation of Helper (Study 2) N = 172
Gratitude	0.34	0.28
Uneasiness	-0.51	-0.33
Willingness to accept the help	0.45	0.24
Care about well-being	0.30	0.55
Improving relationship	0.26	0.56

Note. The difference scores were calculated as follows: morally neutral – morally problematic. Numbers are correlation coefficients (Spearman's  $\rho$ ). All results are significant at  $p < .005$  level.

**Figure 5.** Structure of the mediation models.

2021; Uhlmann et al., 2015). Could inferences about the Helper's moral character underlie any or all of the help-centered antecedents of gratitude? Here, we tested one such possibility, namely, that inferences about the Helper's moral character may inform inferences about the Helper's intention in providing the help (e.g., altruistic vs. strategic), which may in turn modulate gratitude. This would be in line with evidence that judgments of intention and causality behind an agent's act are imbued with inferences about the agent's moral character (Alicke, 2000; Siegel et al., 2017; Uhlmann et al., 2015). As an exploratory analysis not formally preregistered, we tested two serial mediation models with the data from Studies 1 and 2 to directly test this possibility. In these models, condition (morally neutral = 1, morally problematic/immoral = 0) was included as independent variable, inference of the Helper's moral character and inference of the Helper's intention were included as two sequential mediators, and self-reported gratitude was the dependent variable (Figure 5). In addition, we included as covariates perceived effort to the Helper and benefit to the Recipient, moral judgment of the proposed help, and demographic variables. We used SPSS macro PROCESS (Model 6) with 5,000 bias-corrected bootstraps to estimate these mediation models (Hayes, 2013).

Table 2 shows the mediation results based on the data of Studies 1 and 2. For both studies, the serial mediation term was significant, indicating that morally problematic helping behaviors or a morally problematic character trait of the

Helper triggered a bad moral impression of the Helper. This bad moral impression was related to the inference about less benevolent intentions behind the proposed help, which in turn was related to reduced gratitude. We note that the mediation analysis was not conclusive with regard to the causal relationships among the variables and the specific models we tested were only one of several possible models. For example, one could hypothesize that the perception of Helper's moral character may have an impact on the perceived benefit or value of the help, or the Helper's effort in providing the help. We explored these possibilities by replacing "Inference of helping intention" in the model illustrated in Figure 5 with perceived benefit of the help and Helper's effort. The serial mediation effects in these alternative models were less consistent as the ones reported in Table 2 (for detail, please see Supplementary Table S1 and Table S2).

## Discussion

Studies 1 and 2 demonstrated that Recipients were sensitive to the moral status of the Helper and the help they provided. Morally problematic helping behavior or morally neutral helping behavior offered by a morally problematic Helper elicited less gratitude and more uneasiness. Moral status of the Helper and the helping behavior did not only affect the three self-reported emotions of interest (i.e., gratitude, indebtedness, uneasiness), but it also affected the beneficiary's willingness to accept the proposed help and their gratitude-related behavioral tendencies. The mediation analysis suggested that evaluations of the benefactor's moral character may function as another antecedent of gratitude in addition to those proposed in the Tesser Model and inform inferences about the intention underlying the helping behavior.

Why is gratitude hampered when the help or Helper is immoral? The welfare tradeoff ratio (WTR) theory offers additional explanatory perspectives for how various antecedents influence gratitude (Cosmides & Tooby, 2015; Forster et al., 2017, 2021; Sznycer, 2018). Essentially, this line of research consistently suggests that the change in how a Recipient values the Helper (as a consequence of receiving the help) predicts gratitude. Our findings extend this work by

**Table 2.** Results of the Serial Mediation Via Moral Character and Intention.

Terms	Study 1	Study 2
	$B \pm SE$ (95% CI)	$B \pm SE$ (95% CI)
Mediation of moral character	$2.82 \pm 1.18$ [0.84, 5.42]	$5.59 \pm 1.38$ [2.99, 8.35]
Mediation of intention	$-0.05 \pm 0.49$ [-1.04, 0.99]	$-0.74 \pm 0.62$ [-2.18, 0.29]
Serial mediation	$0.70 \pm 0.33$ [0.14, 1.44]	$3.05 \pm 0.94$ [1.30, 5.00]
Direct effect	$3.02 \pm 2.62$ [-2.05, 8.09]	$0.21 \pm 1.80$ [-3.33, 3.75]
Total effect	$6.49 \pm 2.79$ [0.99, 11.99]	$8.11 \pm 1.70$ [4.75, 11.47]

Note. CI = confidence interval.

showing that moral evaluations of the help and the Helper modulate Recipient's valuation update process over and above the material benefit they receive.

The results of the Chinese sample and the American sample were largely consistent, with the most pronounced exception of self-reported indebtedness and effort in Study 1. Although multiple factors could contribute to the between-group difference in indebtedness (e.g., the two groups of participants did not read exactly the same vignettes), one interesting possibility is that the English term "indebtedness" and its Chinese counterpart (*ren-qing zhai*, literally meaning "debt of human feelings/relationships"; cf. Cheung et al., 1996; Gao et al., 2021; Yan, 1996) that we used in the Chinese questionnaire referred to slightly different facets of social and emotional experience. As for perceived effort, the cultural difference was unlikely due to the differences in the vignettes because in all the vignettes the Helper explicitly stated that the help was effortless for them. One speculation is that while American participants might have assumed others to be moral by default, Chinese participants might have made the opposite assumption (Liu et al., 2019; Rand et al., 2012). Therefore, for the American participants, offering an immoral help would go against the Helper's default mindset and require more mental effort, whereas for the Chinese participants the opposite might be true. Future research is needed to more formally examine these cross-cultural variations in responses to (unwanted) help.

### Study 3

Studies 1 and 2 highlighted that gratitude is morally sensitive. In Study 3, we further sought to address potential adaptive values of exhibiting such a morally sensitive pattern, thereby approximating an explanation of why such a sensitivity pattern might have evolved. From one perspective, gratitude might be seen as a fundamental virtue that should not be conditional on the moral status of helping behaviors and/or actors. For instance, in Chinese culture, gratitude—especially gratitude to one's parents (i.e., filial piety)—is regarded as a fundamental virtue (Ivanhoe, 2004). Analogously, the opposite of gratitude—ingratitude—is commonly seen as a vice. In light of this, one might expect that experiencing (and expressing) gratitude is a moral

obligation, regardless of the moral status of the helping behavior or Helper. On the contrary, according to the recent theorization of "witnessing effect" of emotions (Algoe et al., 2020), feeling and expressing gratitude has a communicative function in group contexts as it signals the Recipient's moral endorsement of the helping behavior (McCullough et al., 2001). Then, feeling grateful in response to morally problematic helping behavior may signal moral insensitivity or even a lack of moral integrity of the Recipient to observers or send false signal regarding the Helper's quality as a social partner to one's group. In Study 3, we dissociated these possibilities by examining participants' perceptions of and preferences for Recipients of morally problematic help who expressed different emotions.

### Method

#### Participants

**American sample.** We assumed a small-to-medium main effect ( $f^2 = 0.1$ ) of condition in a linear regression. To detect such effect with 95% power, we need a total sample of at least 132 participants. We therefore preregistered a sample size of 150 to account for participants exclusion due to failure in attention checks. As preregistered, we recruited 150 participants who self-reported as American residents and currently lived in the United States from Prolific. After excluding 18 participants who failed attention checks, our final sample had 132 participants (60 females;  $M_{\text{age}} = 30.6$  years,  $SD = 10.3$ , age range = 18–59; 47.7% with a bachelor's or higher degree, 58.3% self-identified as White/Caucasian, 24.2% were students).

**Chinese sample.** Aiming for a comparable sample size, we preregistered a sampling goal of 150 Chinese participants, a sample size comparable with the American sample. One hundred seventy-eight Chinese participants were recruited from Credamo, among whom 34 were excluded due to failure in the attention checks, leaving 144 participants in the final data set (78 females;  $M_{\text{age}} = 28.1$  years,  $SD = 5.23$ , age range = 20–50; 96.5% with a bachelor's or higher degree, 97.9% self-reported as Han Chinese, 34.0% were students).

**Experimental design and procedure.** In Study 3, we manipulated the emotion felt by a Recipient receiving morally

problematic help in a three-level between-subject design (grateful, uneasy, no emotion). We used two different hypothetical vignettes for this purpose (see Supplementary Materials for details) yielding a total of six possible scenarios (2 vignettes  $\times$  3 emotions). Each participant was randomly assigned one of those six scenarios. Because we were not interested in the effect of vignette, we included vignette as a random slope in our regression analyses. The sample size for each condition was comparable (Gratitude: 40 American, 51 Chinese; Uneasiness: 44 American, 50 Chinese; No emotion: 48 American, 43 Chinese). These cell sizes were all above the suggested threshold of 20 observations (Simmons et al., 2011) and roughly met the more stringent recommendation of 50 observations (Simmons et al., 2013).

### Measurements

**Manipulation and attention checks.** To ensure that our manipulation was effective and attended to, we asked participants to select what emotion the Recipient felt upon being offered help (options: gratitude, uneasiness, shame, no particular emotion). In line with our preregistration, we excluded participants whose choice was not in line with the assigned between-subjects condition from our data analysis. We also asked participants to judge how moral or immoral the proposed help was ( $-50 = \text{extremely immoral}$ ,  $0 = \text{neither moral nor immoral}$ ,  $50 = \text{extremely moral}$ ).

**Moral evaluations of the Recipient.** We assessed participants' (i.e., observers') inferences about three moral traits of the Recipient: "moral," "kind," and "trustworthy" (cf. Yu et al., 2021). We included two competence traits as filler items (i.e., "competent" and "intelligent"). Participants indicated how much they thought the Recipient could be described by those traits on analogue scales ( $0 = \text{not at all}$ ,  $100 = \text{extremely}$ ). The internal reliability of the moral traits ( $\alpha > .85$ ) was good across conditions. We therefore averaged across those measures to obtain a composite measure of inferred moral trait.

**Likeability.** We measured how much the participants liked the Recipient, by asking, "How much do you like [Recipient's name]?" ( $0 = \text{not at all}$ ,  $100 = \text{very much}$ ).

**Suitability as close relationship partners.** We measured how much participants perceived the Recipient as suitable for being a close relationship partner, including spouse, friend, and parent (cf. Everett et al., 2018). We also included a few non-close relationship terms as fillers (i.e., boss, political leader). Participants indicated how good the Recipient would be for each of the above roles on analogue scales ( $0 = \text{not good at all}$ ,  $100 = \text{extremely good}$ ). The internal reliability of the suitability as close relationship partners was good across conditions ( $\alpha > .87$ ). We averaged across those measures to obtain a composite measure of suitability as close relationship partners.

## Results

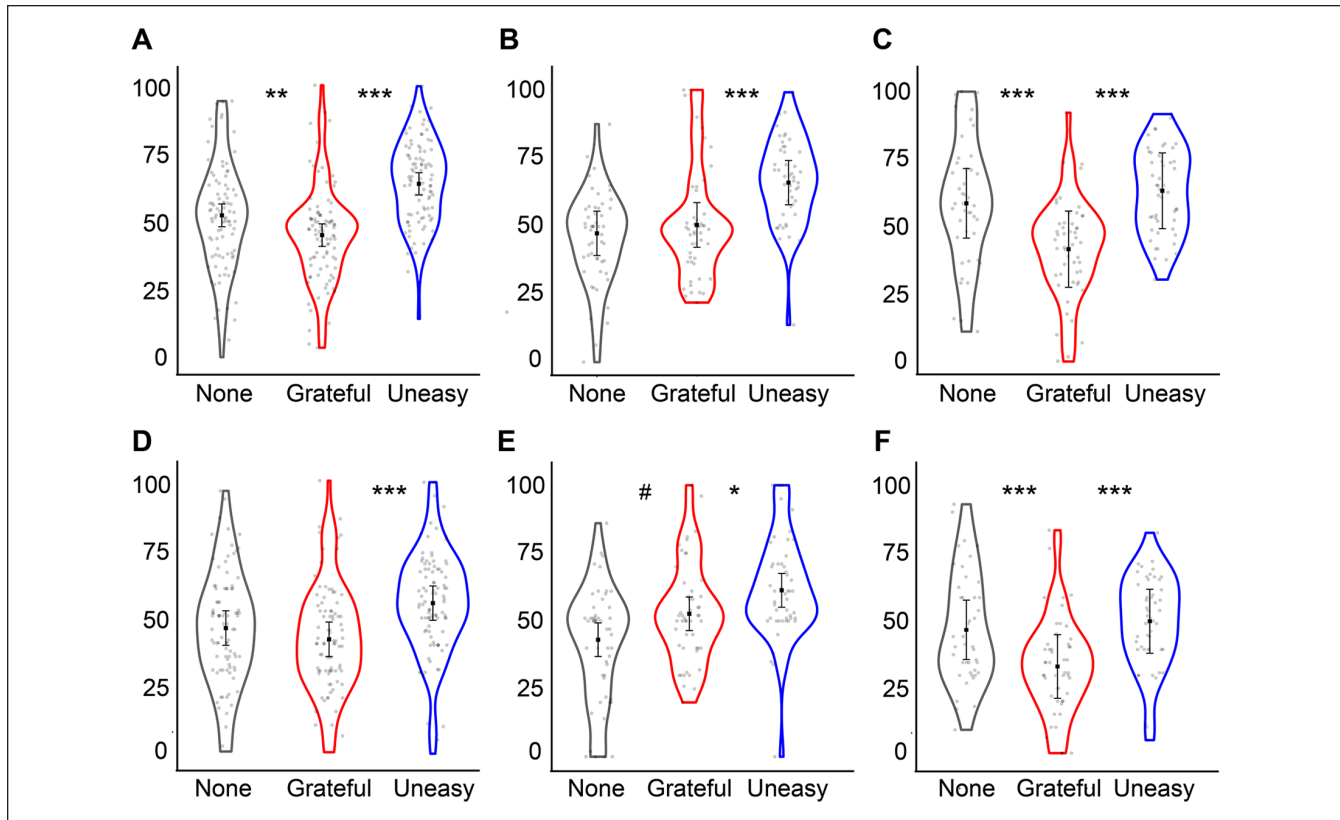
**Manipulation checks.** As intended, participants judged the proposed help as immoral on average ( $M \pm SE = -23.50 \pm 1.31$ ),  $t(275) = -17.91, p < .001$ . This moral evaluation was not modulated by experimental condition ( $F = 0.31, p = .731$ ), participants country ( $F = 2.95, p = .226$ ), or country-by-condition interaction ( $F = 0.69, p = .500$ ; Supplementary Figure S3).

**Moral impression and likeability.** To examine how Recipient's emotional responses to morally problematic help influenced observers' moral impression of them, we ran a linear mixed effect regression model. We included condition as the predictor of interest and demographic variables (e.g., age, gender, highest level of education, studentship, socioeconomic status) as covariates of no interest. Vignette and country were included as random intercepts. The Recipient who felt grateful when offered a morally problematic help was perceived as less moral than the Recipient who felt uneasy ( $B = -19.3 \pm 2.5, 95\% \text{ CI} = [-24.2, -14.3], t = -7.58, p < .001$ ; Figure 6A) or felt no particular emotion ( $B = -7.5 \pm 2.6, 95\% \text{ CI} = [-12.5, -2.3], t = -2.86, p = .005$ ; Figure 6A).

We applied the same analysis to the likeability rating. Participants liked the Recipient who felt grateful when offered a morally problematic help less than the Recipient who felt uneasy ( $B = -13.7 \pm 2.9, 95\% \text{ CI} = [-19.4, -8.1], t = -4.77, p < .001$ ; Figure 6D). However, likeability ratings did not differ significantly between the grateful and the no-emotion condition ( $B = -3.3 \pm 3.0, 95\% \text{ CI} = [-9.2, 2.3], t = -1.13, p = .262$ ; Figure 6D).

Inspecting the results of the American and the Chinese samples separately, it seemed that these two groups of participants exhibited opposite patterns in this latter appraisal: While American participants liked the grateful Recipient slightly more than the Recipient experiencing no particular emotion ( $B = 7.9 \pm 4.1, 95\% \text{ CI} = [0.0, 15.9], t = 1.93, p = .055$ ; Figure 6E), Chinese participants showed the reverse pattern ( $B = -14.2 \pm 4.0, 95\% \text{ CI} = [-21.6, -6.2], t = -3.54, p < .001$ ; Figure 6F). To formally examine this differential pattern, we ran another model where we additionally included country and country-by-condition interaction as fixed effects. This model had a lower BIC (2,432) than the original model (2,447), indicating a better fit. Based on this model, the country-by-condition interaction was significant,  $F(2, 262) = 7.87, p < .001$ , confirming the existence of differential patterns across participant samples. The main effect of condition remained significant,  $F(2, 263) = 12.46, p < .001$ . We note that these between-group comparisons were not formally preregistered and were exploratory.

**Suitability as a close relationship partner.** We ran a similar model as above to examine how the Recipient's emotion amid receiving morally problematic help influenced how observers perceived them as suitable for a close relationship partner (i.e., spouse, parent, friend). The Recipient who felt uneasy was judged as more suitable as a close relationship

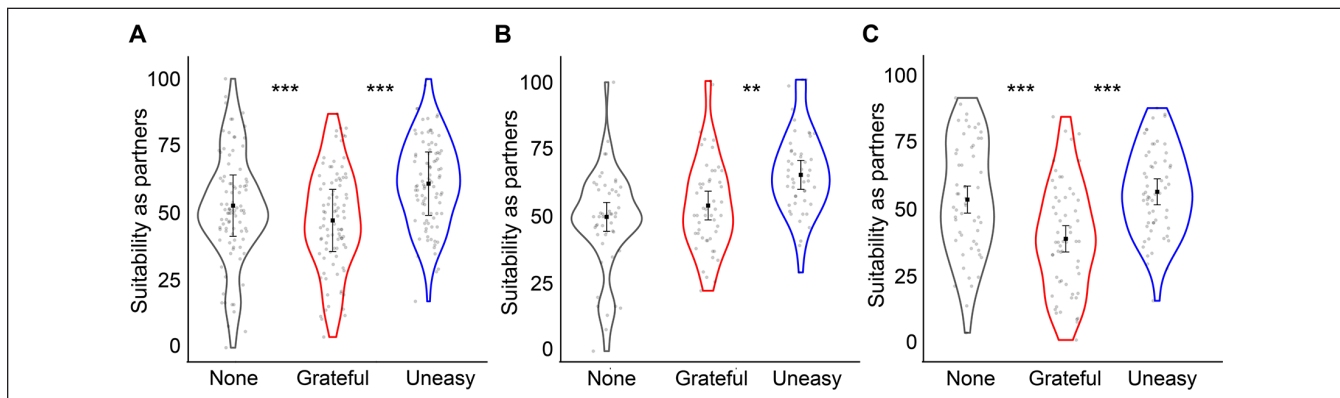


**Figure 6.** Moral impression and likeability as a function of condition: (A) Moral traits (combined), (B) moral traits (American), (C) moral traits (Chinese), (D) likeability (combined), (E) likeability (American), (F) likeability (Chinese).

Note. (A–C) Moral impression of the Recipient as a function of the Recipient’s emotional responses to the proposed immoral help; (D–F) Likeability of the Recipient as a function of the Recipient’s emotional responses to the proposed immoral help.

Note. Error bars show the 95% confidence intervals.

#.05 <  $p$  < .1. \* $p$  < .05. \*\*\* $p$  < .001.



**Figure 7.** Perceived suitability of a close relationship partner as a function of condition: (A) the combined sample, (B) the American sample, and (C) the Chinese sample.

Note. Error bars show the 95% confidence intervals.

\*\* $p$  < .01. \*\*\* $p$  < .001.

partner than the Recipient who felt grateful ( $B = 14.5 \pm 2.7$ , 95% CI = [9.3, 19.7],  $t = 5.44$ ,  $p < .001$ ; Figure 7A) or no particular emotion ( $B = 8.2 \pm 2.7$ , 95% CI = [3.0, 13.5],  $t = 3.02$ ,  $p = .003$ ; Figure 7A). The Recipient who felt no

particular emotion was also judged as more suitable as a close relationship partner than the Recipient who felt grateful ( $B = 6.36 \pm 2.8$ , 95% CI = [0.9, 11.6],  $t = 2.31$ ,  $p = .022$ ; Figure 7A).

To further explore the partner suitability data, in a separate model we included country and country-by-condition interaction as two additional fixed effects. This model had a smaller BIC (2,410) than the model without the two terms (2,421). The country-by-condition interaction effect was significant,  $F(2, 263) = 5.96, p = .003$ . Specifically, for the Chinese sample (Figure 7C), the Recipient who felt grateful was perceived as less suitable for a close relationship partner than the Recipient who felt uneasy ( $B = -16.4 \pm 3.5, 95\% \text{ CI} = [-23.2, -9.5], t = -4.64, p < .001$ ) or had no particular emotion ( $B = -15.1 \pm 3.7, 95\% \text{ CI} = [-22.2, -7.7], t = -4.03, p < .001$ ). For the American sample (Figure 7B), the Recipient who felt grateful was perceived as less suitable for a close relationship partner than the Recipient who felt uneasy ( $B = -12.1 \pm 3.9, 95\% \text{ CI} = [-19.7, -4.6], t = -3.12, p = .002$ ) but no less than the Recipient who had no particular emotion ( $B = 2.6 \pm 3.9, 95\% \text{ CI} = [-4.9, 10.1], t = 0.67, p = .503$ ).

## Discussion

Study 3 suggested that when offered morally problematic help, the Recipient who felt uneasy, relative to grateful, was perceived as more moral, likable, and suitable as a close relationship partner. This suggests that feeling and expressing gratitude when receiving help is not perceived as an unconditional moral virtue. Feeling grateful for a morally problematic help implies that the Recipient consents to the morally problematic help (otherwise they would be more likely to feel uneasy, as Study 1 showed). Such “morally insensitive gratitude” signals to an observer the Recipient’s own moral preference or value and may trigger inferences about an immoral character. More generally, having a morally sensitive gratitude disposition (i.e., feeling grateful when it is morally appropriate and not when it is inappropriate) indicates moral maturity and reliability (Kristjánsson, 2015; Morgan et al., 2017). It therefore makes more sense to have such a person in one’s close relationships (Barclay & Willer, 2007). Taken together, these findings highlight the importance of emotion as information in social and moral life (Anderson et al., 2021; Van Kleef et al., 2010).

## General Discussion

In this article, we demonstrated that the moral status of helping behaviors and Helpers influenced Recipient’s gratitude and gratitude-related behavioral tendencies. Complementary to previous research on the antecedents of gratitude, which has predominantly focused on nonmoral attributes of helping behaviors (McCullough et al., 2001; Naito et al., 2005; Tesser et al., 1968; Tsang, 2007; Wood et al., 2008), our results highlight the importance of moral attributions concerning not only helping behaviors but also the person of the Helper in appraisals of gratitude. Consistent with the moral inference literature (e.g., Brambilla et al., 2021) and the

person-centered approach in moral psychology (Uhlmann et al., 2015), we found that inferences about the Helper’s moral character inform inferences about the Helper’s intentions in offering the help, which in turn has an impact on gratitude. These results underscore the primacy of morality in the appraisal of gratitude, providing empirical evidence for the “moral barometer” function of gratitude (McCullough et al., 2001).

What is the relation between gratitude and morality? This can be answered from a proximal and a distal perspective. The proximal perspective, largely along the line of appraisal theory of emotion, is concerned with how the cognitive processes of moral evaluations modulate or give rise to the cognitive processes of gratitude. Our results reveal two possible pathways with which this could happen: Moral evaluations of the Helper could influence inferences about the Helper’s intention (Table 2) and the perceived benefit or value of the help (Supplementary Table S1). When the Helper is perceived to behave morally problematic, the Recipient has more ground to question the true motive behind the help, and the help may be seen as a cost rather than benefit in the eyes of the Recipient, according to the “moral contagion” literature (Rozin et al., 1994; Rozin, Lowery, et al., 1999). The distal relation is concerned with the social function of gratitude. According to the “find-remind-bind” model (Algoe, 2012), a crucial social function of gratitude is to remind the Recipient of a high-quality social partner. If the Helper is morally problematic, then their quality as a social partner is compromised. Therefore, gratitude is no longer needed to signal to the Recipient the existence of a high-quality social partner or to motivate the Recipient to strengthen the relationship with the Helper. Our findings that Recipients care less about the Helper’s well-being and are less likely to strengthen the relationship with the Helper in the case of morally problematic helping behavior and Helper provide support for this interpretation.

In the present research, we only compared morally problematic help and Helper with morally neutral ones. Because we only used a limited number of vignettes, it is difficult to empirically examine how our results would be affected by the degree of moral badness of the helping behavior or the Helper. Future studies using more fine-grained manipulations of moral valence of helping behaviors and agents (e.g., building on Siegel et al., 2017; Yu et al., 2021) may provide a useful framework for addressing this question. Importantly, to have a more balanced understanding of the relationship between morality and gratitude, it is crucial to investigate the cases on the other end of the moral spectrum, namely, the extremely moral and selfless help and Helper. Here, we would also like to make a distinction between mild and extreme (supererogatory) moral cases. Most of the helping situations we encounter in everyday life and the paradigmatic cases studied in most previous research on gratitude fall under the mildly moral category (Marsh, 2019), such as a stranger holding the door for us. Less commonly considered are cases where a helper

voluntarily makes unexpectedly high self-sacrifice to help others, such as voluntary organ donors and effective altruists. The sparse empirical research on this topic has yielded mixed results, but it seems clear that the relationship between morality and gratitude is more complicated than linear or even monotonic in the extremely moral or supererogatory domain (MacFarquhar, 2016; O'Brien et al., 2014; Pleasant & Barclay, 2018). We speculate that access to the Helper's intention or motive behind extremely moral help may be a key to understanding gratitude and other emotions in return.

Two additional variables may be relevant—the social closeness between the Helper and the Recipient and the expectedness of the help (Forster et al., 2017). As mentioned above, extremely self-sacrificial behaviors are unexpected and may trigger systematic inferences of Helper's intentions or motives behind such acts. Because altruistic acts can bring social adaptive values to the agent (Barclay & Willer, 2007), it is possible that such acts are strategic or performative, even more so when they are unexpectedly extreme. One potential response of Recipients then might be to experience indebtedness and uneasiness rather than gratitude (Gao et al., 2021). However, in close relationship context, or other situations where the Recipient is certain that a Helper's extremely self-sacrificial behaviors are not driven by strategic motives, extreme acts may lead to enhanced gratitude. Of course, social closeness and expectedness of extreme self-sacrifice are inherently linked—such behaviors are more likely to happen within close relationships. Future research incorporating quantitative operationalization of social closeness and expectedness of help is needed to empirically test these hypotheses (cf. Earp et al., 2021). Addressing these questions is an important avenue for future research, whereas pursuing this avenue would go beyond the scope of the present work.

To conclude, the present studies addressed a series of questions regarding how the moral status of helping behaviors and Helpers shape Recipients' gratitude and gratitude-related behavioral tendencies. The results consistently demonstrate that morally problematic help or morally neutral help from a morally problematic Helper elicit less gratitude and more uneasiness, accompanied with less gratitude-specific behavioral tendencies. This pattern, consistent with the notion of the primacy of morality in social cognition, calls for the inclusion of moral evaluation as a cognitive antecedent in the appraisal model of gratitude.

### Acknowledgments

The authors thank Dr. Xiaolin Zhou's team and Dr. Molly Crockett's team for their insightful discussions about the design and significance of the research, and the two anonymous reviewers for their constructive comments on an early version of the paper.

### Declaration of Conflicting Interests

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

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research is partly funded by a University of California Santa Barbara Academic Senate Faculty Research Grant awarded to H.Y.

### ORCID iD

Hongbo Yu  <https://orcid.org/0000-0002-3384-7772>

### Supplemental Material

Supplemental material is available online with this article.

### Note

1. Kate Taylor, Richard A. O'Connell Jr., and Jennifer Medina (2019). Lori Loughlin and 15 others face new charges in college admissions scandal. *The New York Times*.

### References

- Algoe, S. B. (2012). Find, remind, and bind: The functions of gratitude in everyday relationships. *Social and Personality Psychology Compass*, 6(6), 455–469. <https://doi.org/10.1111/j.1751-9004.2012.00439.x>
- Algoe, S. B., Dwyer, P. C., Young, A., & Oveis, C. (2020). A new perspective on the social functions of emotions: Gratitude and the witnessing effect. *Journal of Personality and Social Psychology*, 119(1), 40–74.
- Algoe, S. B., Fredrickson, B. L., & Gable, S. L. (2013). The social functions of the emotion of gratitude via expression. *Emotion*, 13(4), 605–609. <https://doi.org/10.1037/a0032701>
- Algoe, S. B., & Haidt, J. (2009). Witnessing excellence in action: The “other-praising” emotions of elevation, gratitude, and admiration. *Journal of Positive Psychology*, 4(2), 105–127. <https://doi.org/10.1080/17439760802650519>
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, 126(4), 556–574.
- Anderson, R. A., Kamtekar, R., Nichols, S., & Pizarro, D. A. (2021). “False positive” emotions, responsibility, and moral character. *Cognition*, 214, 104770.
- Barclay, P., & Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proceedings of the Royal Society B: Biological Sciences*, 274(1610), 749–753.
- Bartlett, M. Y., Condon, P., Cruz, J., Baumann, J., & Desteno, D. (2012). Gratitude: Prompting behaviours that build relationships. *Cognition and Emotion*, 26(1), 2–13. <https://doi.org/10.1080/02699931.2011.561297>
- Beeler-Duden, S., & Vaish, A. (2020). Paying it forward: The development and underlying mechanisms of upstream reciprocity. *Journal of Experimental Child Psychology*, 192, 104785.
- Brambilla, M., Sacchi, S., Rusconi, P., & Goodwin, G. P. (2021). The primacy of morality in impression development: Theory, research, and future directions. In *Advances in experimental social psychology* (Vol. 64, pp. 187–262). Elsevier.
- Cheung, F. M., Leung, K., Fan, R. M., Song, W.-Z., Zhang, J.-X., & Zhang, J.-P. (1996). Development of the Chinese Personality Assessment Inventory. *Journal of Cross-Cultural Psychology*, 27(2), 181–199.



- Clark, M. S., & Mills, J. R. (2011). A theory of communal (and exchange) relationships. *Handbook of Theories of Social Psychology*, 2, 232–250.
- Cosmides, L., & Tooby, J. (2015). Adaptations for reasoning about social exchange. *The Handbook of Evolutionary Psychology*, 2, 625–668.
- 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(2), 289–293. <https://doi.org/10.1037/a0017883>
- Dickens, L., & DeSteno, D. (2016). The grateful are patient: Heightened daily gratitude is associated with attenuated temporal discounting. *Emotion*, 16(4), 421–425. <https://doi.org/10.1037/emo0000176>
- Earp, B. D., McLoughlin, K. L., Monrad, J. T., Clark, M. S., & Crockett, M. J. (2021). How social relationships shape moral wrongness judgments. *Nature Communications*, 12(1), 1–13.
- Emmons, R. A., & McCullough, M. E. (2004). *The psychology of gratitude (series in affective science)*. Oxford University Press.
- Everett, J. A. C., Faber, N. S., Savulescu, J., & Crockett, M. J. (2018). The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *Journal of Experimental Social Psychology*, 79, 200–216.
- Forster, D. E., Pedersen, E. J., McCullough, M., & Lieberman, D. (2021). *Evaluating benefits, costs, and social value as predictors of gratitude*. <https://psyarxiv.com/ckedr>
- Forster, D. E., Pedersen, E. J., Smith, A., McCullough, M. E., & Lieberman, D. (2017). Benefit valuation predicts gratitude. *Evolution and Human Behavior*, 38(1), 18–26. <https://doi.org/10.1016/j.evolhumbehav.2016.06.003>
- Gao, X., Jolly, E., Yu, H., Liu, H., Zhou, X., & Chang, L. J. (2021). The hidden cost of receiving favors: A theory of indebtedness. *BioRxiv*. <https://doi.org/10.1101/2020.02.03.926295>.
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106(1), 148–168.
- Gulliford, L., Morgan, B., & Kristjánsson, K. (2013). Recent work on the concept of gratitude in philosophy and psychology. *Journal of Value Inquiry*, 47(3), 285–317. <https://doi.org/10.1007/s10790-013-9387-8>
- Hayes, A. F. (2013). *Methodology in the social sciences. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.
- Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A neuro-computational model of altruistic choice and its implications. *Neuron*, 87(2), 451–462.
- Ivanhoe, P. J. (2004). Filial piety as a virtue. In A. Chan (Ed.), *Filial piety in Chinese thought and history* (pp. 198–211). Routledge.
- Kates, S., & DeSteno, D. (2020). Gratitude reduces consumption of depleting resources. *Emotion*, 21, 1119–1123.
- Kristjánsson, K. (2015). An Aristotelian virtue of gratitude. *Topoi*, 34(2), 499–511. <https://doi.org/10.1007/s11245-013-9213-8>
- Lambert, N. M., Clark, M. S., Durtschi, J., Fincham, F. D., & Graham, S. M. (2010). Benefits of expressing gratitude: Expressing gratitude to a partner changes one's view of the relationship. *Psychological Science*, 21(4), 574–580.
- Liu, S. S., Morris, M. W., Talhelm, T., & Yang, Q. (2019). Ingroup vigilance in collectivistic cultures. *Proceedings of the National Academy of Sciences*, 116(29), 14538–14546.
- Ma, L. K., Tunney, R. J., & Ferguson, E. (2017). Does gratitude enhance prosociality? A meta-analytic review. *Psychological Bulletin*, 143(6), 601–635.
- MacFarquhar, L. (2016). *Strangers drowning: Impossible idealism, drastic choices, and the urge to help*. Penguin.
- Manela, T. (2016). Negative feelings of gratitude. *Journal of Value Inquiry*, 50(1), 129–140. <https://doi.org/10.1007/s10790-015-9501-1>
- Marsh, A. A. (2019). The caring continuum: Evolved hormonal and proximal mechanisms explain prosocial and antisocial extremes. *Annual Review of Psychology*, 70, 347–371.
- McCullough, M. E., Kilpatrick, S. D., Emmons, R. A., & Larson, D. B. (2001). Is gratitude a moral affect? *Psychological Bulletin*, 127(2), 249–266.
- Morgan, B., Gulliford, L., & Kristjánsson, K. (2017). A new approach to measuring moral virtues: The multi-component gratitude measure. *Personality and Individual Differences*, 107, 179–189. <https://doi.org/10.1016/j.paid.2016.11.044>
- Naito, T., Wangwan, J., & Tani, M. (2005). Gratitude in university students in Japan and Thailand. *Journal of Cross-cultural Psychology*, 36(2), 247–263.
- O'Brien, G. M., Donaghue, N., Walker, I., & Wood, C. A. (2014). Deservingness and gratitude in the context of heart transplantation. *Qualitative Health Research*, 24(12), 1635–1647.
- Pleasant, A., & Barclay, P. (2018). Why hate the good guy? Antisocial punishment of high cooperators is greater when people compete to be chosen. *Psychological Science*, 29(6), 868–876.
- Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489(7416), 427–430.
- Roberts, R. C. (2015). The normative and the empirical in the study of gratitude. *Res Philosophica*, 92(4), 883–914. <https://doi.org/10.11612/resphil.2015.92.4.9>
- Rozin, P., Markwith, M., & McCauley, C. (1994). Sensitivity to indirect contacts with other persons: AIDS aversion as a composite of aversion to strangers, infection, moral taint, and misfortune. *Journal of Abnormal Psychology*, 103(3), 495.
- Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: a mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76(4), 574.
- Siegel, J. Z., Crockett, M. J., & Dolan, R. J. (2017). Inferences about moral character moderate the impact of consequences on blame and praise. *Cognition*, 167, 201–211. <https://doi.org/10.1016/j.cognition.2017.05.004>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013, January 17–19). *Life after p-hacking*. Meeting of the Society for Personality and Social Psychology, New Orleans, LA, United States.
- Sznycer, D. (2018). Forms and functions of the self-conscious emotions. *Trends in Cognitive Sciences*, 23, 143–157.
- Tesser, A., Gatewood, R., & Driver, M. (1968). Some determinants of gratitude. *Journal of Personality and Social Psychology*, 9(3), 233–236. <https://doi.org/10.1037/h0025905>
- Tsang, J. A. (2006). The effects of helper intention on gratitude and indebtedness. *Motivation and Emotion*, 30(3), 198–204. <https://doi.org/10.1007/s11031-006-9031-z>

- Tsang, J. A. (2007). Gratitude for small and large favors: A behavioral test. *Journal of Positive Psychology, 2*(3), 157–167. <https://doi.org/10.1080/17439760701229019>
- Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science, 10*(1), 72–81. <https://doi.org/10.1177/1745691614556679>
- Van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2010). An interpersonal approach to emotion in social decision making: The emotions as social information model. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 42, pp. 45–96). Elsevier.
- Watkins, P. C. (2013). *Gratitude and the good life: Toward a psychology of appreciation*. Springer.
- Watkins, P. C., Scheer, J., Ovnicek, M., & Kolts, R. (2006). The debt of gratitude: Dissociating gratitude and indebtedness. *Cognition and Emotion, 20*(2), 217–241. <https://doi.org/10.1080/02699930500172291>
- Wood, A. M., Maltby, J., Gillett, R., Linley, P. A., & Joseph, S. (2008). The role of gratitude in the development of social support, stress, and depression: Two longitudinal studies. *Journal of Research in Personality, 42*(4), 854–871.
- Yan, Y. (1996). *The flow of gifts: Reciprocity and social networks in a Chinese village*. Stanford University Press.
- Yu, H., Siegel, J. Z., Clithero, J. A., & Crockett, M. J. (2021). How peer influence shapes value computation in moral decision-making. *Cognition, 211*, 104641.