



Cite this article: Chan CF, Lim HC, Lau FY, Ip W, Lui CFS, Tam KY, Feldman G. 2025 Revisiting the effects of helper intentions on gratitude and indebtedness: Replication and extensions Registered Report of Tsang (2006). *R. Soc. Open Sci.* **12**: 250508.

<https://doi.org/10.1098/rsos.250508>

Received: 17 March 2025

Accepted: 18 March 2025

Subject Category:

Psychology and cognitive neuroscience

Subject Areas:

Psychology

Keywords:

gratitude, indebtedness, intentions, reciprocity, affect, judgement and decision-making, replication, registered report

Author for correspondence:

Gilad Feldman

e-mail: gfeldman@hku.hk

[†]Hiu Ching Lim, Fung Yee Lau, Wing Ip and Chak Fong Shannon Lui are joint second authors.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7790589>.

Revisiting the effects of helper intentions on gratitude and indebtedness: Replication and extensions Registered Report of Tsang (2006)

Chi Fung Chan¹, Hiu Ching Lim^{1,†}, Fung Yee Lau^{1,†}, Wing Ip^{1,†}, Chak Fong Shannon Lui^{1,†}, Katy Yuen Yan Tam³ and Gilad Feldman²

¹Department of Psychology, and ²Department of Psychology, University of Hong Kong, Hong Kong

³University of Toronto, Toronto, Ontario, Canada

KYYT, 0000-0003-2197-8705; GF, 0000-0003-2812-6599

Gratitude and indebtedness are common emotions in response to a favour, yet research suggests that they are experienced differently depending on the situation. Tsang (Tsang JA. 2006 The effects of helper intention on gratitude and indebtedness. *Motiv. Emot.* **30**, 198–204. ([doi:10.1007/s11031-006-9031-z](https://doi.org/10.1007/s11031-006-9031-z))), found that gratitude for a favour depended on perceived helper intention, whereas indebtedness did not. Perceived benevolent helper intentions were associated with higher gratitude from beneficiaries compared to selfish ones, yet had no associations with indebtedness. In a registered report with a United States Prolific student sample ($n = 759$), we conducted a replication and extensions of studies 2 and 3 from Tsang, 2006. In the original studies, Tsang found support for the impact of the helper's intention on gratitude (study 2: $\eta_p^2 = 0.20$ [0.08, 0.32]; study 3: $\eta_p^2 = 0.14$ [0.03, 0.26]), but not for indebtedness (study 2: $\eta_p^2 = 0.01$ [0.00, 0.08]; study 3: $\eta_p^2 = 0.00$ [0.00, 0.03]). In our replications, we found support for the impact of helper's intention on gratitude (study 2: $\eta_p^2 = 0.33$ [0.28, 0.37]; study 3: $\eta_p^2 = 0.16$ [0.12, 0.20]), and—as expected—no support for an effect on indebtedness (study 2: $\eta_p^2 = 0.00$ [0.00, 0.01]; study 3: $\eta_p^2 = 0.01$ [0.00, 0.01]). We concluded a successful replication, that helping intent was more strongly associated with gratitude than with indebtedness. Extending

the replication, we found evidence for the impact of helper intention on perceived expectations for reciprocity ($d = 1.51$ [1.31, 1.71]), and reciprocity inclination ($d = 0.66$ [0.48, 0.84]), and for opposite associations of perceived reciprocity expectations with gratitude ($r = -0.28$ [-0.35, -0.22]) and indebtedness ($r = 0.17$ [0.10, 0.24]). Materials, data and code are available on: <https://osf.io/ghfy4/>. This registered report has been officially endorsed by the Peer Community in Registered Reports: <https://doi.org/10.24072/pci.rr.100788>.

1. Background

Gratitude and indebtedness are common emotions in response to receiving help, but studies suggest that they are experienced differently depending on the situation. Tsang [1] showed that helper intentions were associated with feelings of gratitude, yet less so for indebtedness: people reported feeling more grateful when the helper's intentions were perceived as being less selfish, with weaker to no effects of the helpers' intentions on feelings of indebtedness.

We conducted a close replication and extension of Tsang [1] with two main goals. Our first goal was to conduct an independent replication of the impact of the helper's intentions, comparing gratitude and indebtedness. Our second goal was to examine extensions, aiming to enrich our understanding of how differences in helper intentions impact reciprocation.

We begin by introducing the literature on gratitude and indebtedness and the impact of helpers' intentions on these emotions. We then discuss our motivations for the current replication and review Tsang [1] as our chosen article for replication. Finally, we outline the replication and extension hypotheses, study design and methods.

1.1. The effect of helper intention on gratitude and indebtedness

Gratitude and indebtedness are common reactions to receiving help, with these emotions varying across situations [2–4]. Consider, for example, how a student would respond to a classmate who has volunteered to help with homework depends on perceptions of selfish-ulterior intent. The student's gratitude and indebtedness may depend on whether the act of helping seemed to have been purely benevolent to help another, or rather based on selfish-ulterior intent. These two emotions have often been equated in the early literature and yet evidence showing that these emotions are elicited in different situations suggested the need to differentiate between them [1,5,6].

Gratitude is commonly defined as a positive emotion arising from the appreciation of an action by another person that is desirable and valuable to oneself [7]. McCullough *et al.* [8] argued that it is associated with the prosocial and voluntary nature of the act, as well as the cost incurred and benefits received. Therefore, gratitude may depend on evaluations of the helpers' costs, altruistic intentions and the value of the favour to the person being helped [9–12]. This is in line with the cognitive perspective that gratitude is defined as the product of the cognition that one has been the beneficiary of others' goodwill [13].

Greenberg [5] defined indebtedness as a beneficiary's feeling of obligation to repay the benefactor following norms of reciprocity [14], so as to restore equality in a social exchange [15]. In this context, the favour does not necessarily have to be altruistic. Peng *et al.* [16] suggested that it is the cost of the favour, rather than its intent, that determines the inequality of social exchange, thus affecting feelings of indebtedness.

1.1.1. Relationship and differences between gratitude and indebtedness

Algoe *et al.* [17] posited that gratitude is a positive emotion, whereas indebtedness is a negative emotion. Gratitude leads people to thank their benefactor, whereas indebtedness leads people to try and return the favour. This aligns with the work by Frijda [18] showing that distinct emotions manifest distinct action tendencies—a consequential urge to carry out certain expressive behaviours. It also echoes the research by Gray *et al.* [3] indicating that gratitude is associated with prosocial motivations, whereas indebtedness is associated with avoidance motivations. This was later explained by the broaden-and-build theory by Fredrickson [19], that gratitude, as one of the positive emotions,

serves to broaden one's thoughts and actions to reciprocate, whereas indebtedness is associated with a relatively narrower tit-for-tat reciprocity.

Researchers have tried to distinguish the two emotions in different ways. In terms of their causes, Watkins *et al.* [6] distinguished them by manipulating helpers' expectations of reciprocity, finding that higher expectations resulted in decreased gratitude yet increased indebtedness. In line with the difference in action tendencies between gratitude and indebtedness, they also found that participants were more likely to express willingness to return the favour if the benefactor communicated weaker reciprocation expectations (reciprocation negatively associated with reciprocation expectations). However, this predicted outcome is arguably counter to combining the theoretical paradigms by Tsang [1] and Algoe *et al.* [17], which together suggest that beneficiaries would be less likely to reciprocate if benefactors held weaker reciprocation expectations (reciprocation positively associated with reciprocation expectations).

Emmons & Crumpler [20] proposed that gratitude is an interpersonal emotion that enhances relational wellbeing, with Mathews & Green [15] arguing that indebtedness is more of a self-focused emotion. Therefore, self-focused people felt less commitment and closeness towards the benefactor. This supported the conclusion by Algoe *et al.* [17] that gratitude enhances relationships, whereas indebtedness only maintains relationships. As shown in the research that distinguishes them, they differ in terms of causes and effects.

Furthermore, there is some research which indicates that these two emotions play different functions in sociality. For example, accumulated literature suggested that gratitude contains a relation-oriented function to promote intimate bonds (e.g. [4,21–24]), whereas indebtedness contains an exchange-oriented function (e.g. [4,25,26]). These functional differences may explain why helpers' intentions are influential to one's gratitude and indebtedness.

1.1.2. Perceived helper intention

The differences between gratitude and indebtedness can be reflected in perceived helpers' intentions. Ames *et al.* [27] found that when beneficiaries perceived helpers' intention as caring they experienced more positive feelings towards the helper (e.g. happiness and gratitude). Alternatively, perceiving helping intent as manipulative or deceitful triggered negative affect (e.g. indebtedness and anger).

Tsang [1] further examined the effects of helper intentions on feelings of gratitude and indebtedness. The result was partially consistent with Watkins *et al.* [6], with participants feeling more grateful for favours offered with benevolent intentions, compared to favours perceived as having selfish intentions which did not seem to affect indebtedness. However, more recent findings by Welsh *et al.* [28] found helpers' motives (prosocial versus self-interested) do influence individuals' levels of indebtedness. They argued that favours with self-interested motives induced less indebtedness than those with prosocial motives, contradicting the work by Tsang [1]. If both findings hold in independent well-powered pre-registered replications, then more work is needed to account for these mixed results.

1.2. Choice of study for replication: Tsang (2006)

We embarked on a well-powered close replication and extension registered report of Tsang [1]. We aimed to revisit the phenomenon to examine the reproducibility and replicability of the findings with an independent pre-registered well-powered replication and extension. This follows the recent growing recognition of the importance of reproducibility and replicability in psychological science (e.g. [29,30]).

We chose the study by Tsang [1] based on several factors: its profound academic impact, the absence of direct replications and the realignment in the literature initiated by the article. The article has had an impact on scholarly research in the area of social psychology, and at the time of writing (February 2025), there were 377 Google Scholar citations of the article with many impactful follow-up theoretical and empirical articles. One example is the work by Algoe *et al.* [17] on how gratitude and indebtedness affect romantic relationships. They concluded that gratitude improves interpersonal relationship quality, whereas indebtedness exerts no detectable influence. Based on the findings by Tsang [1], they established a link between the nature of these two emotions and such interpersonal outcomes. That is, the dependence of gratitude on helper's intentions entails that the focus is on the helper's positive feelings and favourable mental states (e.g. being generous/caring more for the beneficiary). Meanwhile, indebtedness, being less sensitive to helper's intentions, focuses on the benefit and thus triggers

reciprocity only as a dutiful exchange. Therefore, Tsang [1] contributed to the development of the field by elaborating on the contrast between the two emotions. Her work deepened our understanding of their distinct role in different aspects of life, such as interpersonal relationships. To the best of our knowledge, there are currently no published direct replications of this study.

Despite its impact, the departure of the theory and findings by Tsang [1] from previous research necessitates independent replications to help ensure the realignment is built on solid ground, examines the robustness of the findings and clarifies possible directions for resolution. Prior to Tsang, many studies did not distinguish between the two emotions and instead measured them as one single construct [5,12,31]. The more recent body of research has mostly aligned with the notion that gratitude and indebtedness are distinct emotions.

In addition, the target article presented a theoretical model that predicted no effects for the impact of intent on indebtedness. In their findings, they also reported failing to find a signal in support of rejecting the null hypothesis of finding no differences for indebtedness between the benevolent and selfish intent conditions and built on that to conclude no effects. However, null hypothesis significance testing methods are not well suited for testing and quantifying support for a null hypothesis. We felt it important to revisit the theoretical model by reframing the null hypothesis to differences in effects between gratitude and indebtedness, to rerun the studies with well-powered samples and to add additional analyses that address the null hypothesis issue to gain deeper insights into the phenomenon.

1.3. Overview of our replication and extensions

The empirical work by Tsang [1] consisted of three studies, and in the current replication, we focused on studies 2 and 3, which we ran in a single data collection, with the study order randomized to address order effects and to allow us to examine potential consistency between studies. We extended study 3 to test whether helper intentions were associated with reciprocity inclination. In the extension, we proposed and tested five hypotheses, summarized in table 1. We expected to find support for an association between helper intention, gratitude and reciprocation inclination (see table 4 for extension experimental design).

1.4. Original hypotheses and findings in the target article

Tsang [1] examined how perceived helper intentions are associated with gratitude and indebtedness experienced by the beneficiary. The core hypothesis was that benevolent (versus selfish) intentions were more strongly associated with gratitude than with indebtedness. We focused our replication on the studies 2 and 3 by Tsang [1], given that study 3 contains all the essential experimental designs of study 1 with an extra condition of ambiguous helper intention for investigation. We briefly outline the studies below.

Study 2 examined the effect of perceived helper intentions on levels of gratitude and indebtedness experienced by the beneficiary in real-life situations. It was conducted using an undergraduate sample at Baylor University, asking participants to recall and write about an experience in which someone offered them a valuable favour, randomly assigning participants to either recall a benevolent or a selfish helper. Study 3 had similar research questions to study 2 yet had different experimental designs. Rather than instructing participants to recall favours to them, it presented participants with a scenario describing benefactors' helping intent, either benevolent or selfish, randomly assigned. It added an extra condition of ambiguous intent as a control condition, allowing participants to make their own inferences regarding the helper's intentions.

We provided a summary of the hypotheses and their corresponding findings in table 1 (see the electronic supplementary material, Analysis of the original article subsection for further details). The target article had many hypotheses and many associated analyses, and we therefore pre-registered that our replication criteria will focus on the following. In our replication of study 2, our focus was on the comparison of hypotheses 2 and 3: 'impact of intent (benevolent > selfish) on gratitude is stronger than on indebtedness'. In our replication of study 3, our focus was on the comparison of hypotheses 7b/c and 8b/c: 'impact of intent (benevolent > ambiguous > selfish) on gratitude is stronger than on indebtedness'.

Given the two studies, we pre-registered our overall strategy to conclude a successful replication if the findings of the two studies are aligned with a signal in the same direction as the target article by

Table 1. Summary of effect size and CI of the target article.

Study	Hypothesis	Hypotheses	<i>p</i>	Effect size	90% CI	95% CI
2	1a	Gratitude is associated with indebtedness across conditions (selfish-ulterior and benevolent combined)	<0.001*	$r = 0.57$	/	[0.41, 0.69]
	1b	Gratitude is associated with indebtedness in the selfish-ulterior condition	<0.001*	$r = 0.61$	/	[0.39, 0.76]
	1c	Gratitude is [not] associated with indebtedness in benevolent condition. (Reframed from the target article's null hypothesis)	>0.20	$r = 0.20$	/	[−0.10, 0.47]
	1b + 1c	Combined: gratitude is more strongly associated with indebtedness in the selfish-ulterior condition than in the benevolent condition. (Reframed from the target article's effect/no-effect)		$0.61 > 0.20$		
	2	Benevolent favours result in more gratitude than selfish-ulterior favours, even after controlling the magnitude of favour	<0.001*	$\eta_p^2 = 0.2$	[0.08, 0.32]	/
	3	Benevolent favours [do not] result in more indebtedness than selfish-ulterior favours, even after controlling the magnitude of favour. (Reframed from the target article's null hypothesis)	>0.20	$\eta_p^2 = 0.01$	[0.00, 0.08]	/
	2 + 3	Combined: impact of intent on gratitude (benevolent > selfish) is stronger than on indebtedness. (Reframed from the target article's effect/no-effect)		$0.2 > 0.01$		
	4 (2r)	(Regression complementary analysis). Benevolent favours result in more gratitude than selfish-ulterior favours, even after controlling the magnitude of favour	<0.001* (i) <0.01 (ii) <0.01	$R^2 = 0.73$ $\beta = 0.32$ $\beta = 0.62$	/	[0.61, 0.81]
	5 (3r)	(Regression complementary analysis). Benevolent favours result in more indebtedness than selfish-ulterior favours, even after controlling the magnitude of favour. (Reframed from the target article's null hypothesis)	<0.001* (i) >0.20 (ii) <0.001	$R^2 = 0.26$ $\beta = 0.13$ $\beta = 0.42$	/	[0.10, 0.41]
	4 + 5 (2r + 3r)	(Regression complementary analysis). The differences between benevolent and selfish-ulterior are stronger for gratitude (H4) than for indebtedness (H5), even after controlling the magnitude of favour		$0.32 > 0.13$		
3	6	Gratitude is associated with indebtedness in the ambiguous condition	<0.05*	$r = 0.42$	/	[0.06, 0.68]
	7a	Gratitude is different between the three conditions (benevolent, ulterior, and ambiguous)	<0.01*	$\eta_p^2 = 0.14$	[0.03, 0.26]	/
	7b	Gratitude is higher in the benevolent condition compared to the ambiguous condition	<0.05*	$d = 0.55$	/	[0.02, 1.08]
	7c	Gratitude is [not] higher in the ambiguous condition compared to the selfish-ulterior condition. (Reframed from the target article's null result)	0.07	$d = 0.49$	/	[−0.04, 1.01]
	8a	Indebtedness is [not] different between the three conditions (benevolent, selfish, and ambiguous. (Reframed from the target article's null hypothesis)	>0.20	$\eta_p^2 = 0.00$	[0.00, 0.03]	/
	8b	Indebtedness is [not] higher in the benevolent condition compared to the ambiguous condition. (Reframed from the target article's null hypothesis)	>0.20	$d = 0.13$	/	[−0.39, 0.64]
	8c	Indebtedness is [not] higher in the ambiguous condition compared to the selfish-ulterior condition. (Reframed from the target article's null hypothesis)	>0.20	$d = 0.03$	/	[−0.49, 0.55]
	7b/c + 8b/c	Combined: impact of intent on gratitude (benevolent > ambiguous > selfish) is stronger than on indebtedness. (Reframed from the target article's effect/no-effect)		$0.55 > 0.13$ $ −0.49 > −0.03 $		
	9	Ratings of helper intentions are associated with gratitude in the ambiguous condition	<0.05*	$r = −0.40$	/	[−0.67, −0.04]
	10	Ratings of helper intentions are [not] associated with indebtedness in the ambiguous condition. (Reframed from the target article's null hypothesis)	>0.20	$r = 0.00$	/	[−0.37, 0.37]
	9 + 10	Combined: Ratings of helper intentions are more strongly associated with gratitude than indebtedness in the ambiguous condition. (Reframed from the target article's effect/no-effect)		$0.40 > 0.00$		
Extensions						
3	11a	Competing hypotheses: benevolent helping is perceived as involving <i>lower</i> expectations for reciprocation than selfish helping				
	11b	Competing hypotheses: benevolent helping is perceived as involving <i>higher</i> expectations for reciprocation than selfish helping				
	12a	Competing hypotheses: benevolent helping leads to <i>lower</i> intent to reciprocate than selfish helping				
	12b	Competing hypotheses: benevolent helping leads to <i>higher</i> intent to reciprocate than selfish helping				

(Continued.)

Table 1. (Continued.)

Study	Hypothesis	Hypotheses	<i>p</i>	Effect size	90% CI	95% CI
13		Lower expectation for reciprocity is associated with a stronger intent to reciprocate				
14		Higher expectation for reciprocity is associated with less gratitude				
15		Higher expectation for reciprocity is associated with more indebtedness				

Note: all calculations are corrected to two decimal places if possible. Effect = Cohen's *d* or partial eta squared. CI = confidence interval.

The asterisk refers to hypotheses supported with $p < 0.05$. Bolded hypotheses are the core hypotheses which will be used to test the replicability of the target article. Hypotheses 4, 5, and 4 + 5, are re-analyses of the hypotheses 2r, 3r, and 2r + 3r. Hypotheses 1c, 3, 8a, 8b, and 8c were originally null hypotheses, yet we reframed those to a testable alternative to the null, with indication of the null hypothesis in brackets (e.g. '[not]'). Similarly, the combined hypotheses 1b+1c, 2+3, and 7b/c + 8b/c reframed the null hypotheses from 1c, 3, 8a, 8b, and 8c to a testable hypothesis expecting stronger effects for gratitude compared to indebtedness.

Tsang [1], mixed results if only one of two is supported and failed replication if we fail to find support for both studies.

1.5. Exploratory extension: effect of helper intentions on reciprocity expectations and inclination

We aimed to extend study 3 by examining the impact of helper intentions (benevolent versus selfish-ulterior versus ambiguous) on reciprocity using two measures: (i) perceived reciprocity expectations, and (ii) inclination to reciprocate.

We built our extension on the findings by Watkins *et al.* [6], who, like Tsang [1], argued that gratitude and indebtedness are distinct, but went further to argue that expectations for reciprocity would increase indebtedness but decrease gratitude. Tying these findings together with the experimental paradigm of Tsang [1], we aimed to examine the associations between perceived expectations for reciprocity and the inclination to reciprocate. If perceived expectations to reciprocate are *positively* associated with inclination to reciprocate then it would, according to Watkins *et al.* [6], be associated with *increased* indebtedness and *decreased* gratitude. However, if perceived expectations to reciprocate are *negatively* associated with inclination to reciprocate, then it would, according to Watkins *et al.* [6], be associated with *decreased* indebtedness and *increased* gratitude. Therefore, if we were to try and tie the two sets of findings together then the more theory consistent association seems to be that inclination to reciprocate is negatively associated with expectations to reciprocate and therefore higher gratitude than indebtedness.

Findings in the literature about the associations between gratitude and reciprocity have so far been mixed. For example, a seminal study by Bartlett & DeSteno [2] illustrated that gratitude is positively associated with reciprocity, whereas Peng *et al.* [16] failed to replicate Bartlett & DeSteno [2] and did not find any support for links with reciprocity for both gratitude and indebtedness. Therefore, our extension could be thought of as a conceptual replication of the Bartlett & DeSteno [2] and Peng *et al.* [16] directions to try and determine whether reciprocity might play a role, using an empirical design from a different study. To the best of our knowledge, there has been no research examining the impact of helper intention on reciprocation inclination.

In summary, our extension ties and contrasts the predictions by Tsang [1] and Watkins *et al.* [6] and by Bartlett & DeSteno [2] and Peng *et al.* [16] to examine (i) the associations between helper intentions and expectations for reciprocity, and (ii) the relationship between reciprocity inclination, gratitude and indebtedness.

1.6. Pre-registration and Open Science

We provided all materials, data and code on: <https://osf.io/ghfy4/>. This registered report was submitted to *Royal Society Open Science* (RSOS) following peer review and recommendation for stage 2 acceptance at the *Peer Community In* (PCI) *Registered Reports* platform. Full details of the peer review and recommendation of the paper at PCI Registered Reports may be found at the links below. After submission to the journal, the article received no additional external peer review but was accepted on the basis of the Editor's recommendation according to the RSOS PCI Registered Reports' policy (<https://royalsocietypublishing.org/rsos/registered-reports#PCIRR>). Stage 1 recommendation and

review history: Chen [32]; <https://rr.peercommunityin.org/articles/rec?id=373> / <https://osf.io/uyfvq/> (our frozen pre-registration version of the entire stage 1 packet: <https://osf.io/ka2sv/>). Stage 2 recommendation and review history: Chen [33]; <https://doi.org/10.24072/pci.rr.100788>. All measures, manipulations and exclusions conducted for this investigation are reported, and data collection was completed before conducting the data analyses. The project was part of a large mass replications and extensions project, which received ethics approval from the University of Hong Kong (no. EA210265). This registered report was written based on the registered report template by Feldman [34].

2. Method

2.1. Power and sensitivity analyses

We first calculated effect sizes and conducted a power analysis based on the effects reported in the target article. Effect size and confidence intervals (CIs) were calculated with R (version: 4.1.2; [35]) with the help of a guide by Jané *et al.* [36], and power analyses were then conducted with a combination of R and GPOWER (version 3.1.9.6; [37]) for the factors that the authors found support for in the target article (i.e. flagged as significant results). Rounding up to the highest minimum sample size required for both studies, we concluded that the minimum required sample size was 264 participants in total. This calculation was based on the effect size of $d = 0.55$, with power of 0.95, $\alpha = 0.05$ and allocation ratio of 1 : 1. We provide more information regarding these calculations in the electronic supplementary material, Power analysis of the original study effect to assess the required sample for replication subsection. To allow for a comparison, the target article's study 2 had 92 participants, and study 3 had 86 participants.

Given the likelihood that the original effects are overestimated, we used the suggested Simonsohn [38] rule of thumb, even if meant for other designs, and multiplied the estimated required sample of 264 by 2.5 to result in 660. We then aimed for an even larger sample size of 800.

In stage 1, we reported a sensitivity analysis for a sample of 750, expecting that some participants would not complete the survey, which was close to our final sample of 759 we report below. We found that it allowed for the detection of $d = 0.24$ for independent *t*-test two conditions contrasts for the study 2 design and $f = 0.14$ for a three conditions ANOVA for the study 3 design and $d = 0.29$ for contrasts between conditions with $n = 250$ (all 95% power, $\alpha = 5\%$, one-tail), which are typically considered medium effects in social psychology research [36]. These are much smaller effects than those reported in the target article.

2.2. Participants

We recruited United States (US) American students online through Prolific, with a final sample of 759 participants ($M_{\text{age}} = 30.47$, s.d. = 11.02; 297 males, 439 females; 18 other; five did not disclose). We note that 907 persons began the survey but 148 did not proceed beyond the consent and verifications and therefore had no data to analyse (explained in 'Procedure'). We did not pre-register any additional exclusion criteria. We summarized a comparison of the target article sample and the replication samples in table 2.

We targeted US American students using Prolific's filters. We restricted the location to the US using 'standard sample', we set it to 'nationality: United States', 'country of birth: United States', 'student status: yes', 'minimum approval rate: 90%', 'maximum approval rate: 100%', 'minimum previous submissions: 50', 'maximum previous submissions: 10000' (Prolific's upper limit), 'place of most time spent before turning 18: United States'.

2.3. Design

We ran the two studies together in a single unified data collection. The display of scenarios and conditions was counterbalanced using the randomizer 'evenly present' function in QUALTRICS. Scenarios were presented in random order and participants were randomly and evenly assigned to the different conditions. This method was previously tested successfully in many of the replications and extensions conducted by our team (e.g. [39,40]) and is especially powerful in addressing concerns about the target sample (naivety, attentiveness, etc.) when some studies from the same target article replicate successfully whereas others do not, as well as in the potential in drawing inferences about the

Table 2. Differences and similarities between the target article and our replication.

	Tsang [1]	US Prolific workers
Sample size	Study 2: 92 Study 3: 86	759
Geographical origin	Undergraduates studying at Baylor University	US American Prolific students
Gender	Study 2: 16 males, 76 females Study 3: 13 males, 49 females, 24 did not disclose	297 males, 439 females, 18 others and five did not disclose
Median age (years)	Unreported	27
Average age (years)	Unreported	30.47
S.D. age (years)	Unreported	11.02
Age range (years)	Unreported	18–85
Medium (location)	Study 2: University laboratory cubicles Study 3: Unreported	Online
Compensation	Receiving extra course credit for their participation	Nominal payment
Year	2006 or earlier	2024

links between the different studies and consistency in participants’ responding to similar psychology paradigms.

2.3.1. Replication

We summarized the experimental design in tables 3 and 4. The replication of study 2 was a between-subject experiment with two conditions (perceived helper intention: benevolent versus selfish-ulterior condition). The replication of study 3 was a between-subject experiment with three conditions (perceived helper intention: benevolent, ambiguous and selfish-ulterior). We provided all measures in the QUALTRICS export in the Open Science Framework (OSF) folder.

2.4. Procedure

We reached out to the author of the target article and are very grateful for the materials she provided which were very helpful in our reconstruction of the studies.

Participants first indicated their consent, with four questions confirming their eligibility, understanding and agreement with study terms, which they had to answer with a ‘yes’ and the required responses in order to proceed to the study. Three of the four questions also served as attention checks, with a randomized display order of the options (yes, no and not sure)—(i) ‘are you able to pay close attention to the details provided and carefully answer questions that follow?’; (ii) ‘do you understand the study outline and are willing to participate in a survey with brief writing and comprehension checks?’; and (iii) ‘are you a native English speaker born, raised, and currently located in the US?’. Failing any of the three attention questions meant that the participants did not indicate consent and therefore could not embark on the study. These were followed by writing or copy-pasting a statement indicating that they understand and agree to the terms, which participants had to enter correctly in order to proceed, with as many attempts as needed. Upon completion of these steps, participants proceeded to begin the survey.

Following consent and qualification questions, participants completed two studies, a replication of study 2 and a replication of study 3 from the target article, in random order. In contrast to the original article where the two studies were conducted separately, we combined the two studies into a single data collection.

In the replication of study 2, participants recalled an experience in the past year in which they felt that ‘someone else had caused, and was controlling, what was happening in the situation’, and ‘the positive consequences of this other person’s actions were important to you’. Participants in the benevolent condition recalled a situation in which ‘the other person was doing something good for you for unselfish reasons’. Participants in the selfish-ulterior condition recalled someone having done something good for them for selfish reasons.

Table 3. Study 2: replication and extension experimental design.

Independent variable (IV):	Benevolent condition	Selfish-ulterior condition
Helper intentions (between-subject)	Asked to recall a situation that someone has done something good for a benevolent reason	Asked to recall a situation that someone has done something good for a selfish reason
Dependent variables (DV)	DV1: Gratitude ‘Please choose the number by each adjective to indicate the extent to which you feel each of the following emotions right now in reaction to thinking about the past situation:’ emotion adjectives include <i>grateful</i>, <i>thankful</i>, and <i>appreciative</i>. Scale: 1 = <i>Would feel very little of this emotion</i>, 7 = <i>Would feel a lot of this emotion</i> (source: [8]) DV2: Indebtedness ‘Please choose the number by each adjective to indicate the extent to which you feel each of the following emotions right now in reaction to thinking about the past situation:’ emotion adjectives include <i>indebted</i> and <i>obligated</i>. Scale: 1 = <i>Would feel very little of this emotion</i>, 7 = <i>Would feel a lot of this emotion</i> (source: [5]) DV3: Other emotions ‘Please choose the number by each adjective to indicate the extent to which you feel each of the following emotions right now in reaction to thinking about the past situation:’ emotion adjectives include <i>calm</i>, <i>pleased</i>, <i>resentful</i>, <i>upset</i>, and <i>annoyed</i>. Scale: 1 = <i>Would feel very little of this emotion</i>, 7 = <i>Would feel a lot of this emotion</i> DV4 (manipulation check): Perceived helpers’ motivations ‘Please rate the other person’s motivations on the following scale:’ Scale: 1 = <i>Very concerned about me</i>, 7 = <i>Motivated mostly by selfish reasons</i> DV5 (extension): Perceived expectations for reciprocity ‘Please rate your understanding of the other person’s expectations of you to reciprocate.’ Scale: 1 = <i>No expectations to reciprocate</i>, 7 = <i>Very high expectations to reciprocate</i>	
Covariate (C)	Magnitude of the favour ‘How big of a favour do you think the other person did for you?’ and ‘how costly (in terms of money, time, effort, etc.) do you think this situation was for the person who did something good for you?’ Scale: 1 = <i>A very small favour</i>, 7 = <i>A very big favour</i>	
Comprehension checks (CC)	(1) What type of helping behaviour are you asked to recall? (2) Whose helping behaviour are you asked to recall?	

Note: DV3 was found in the study materials provided by the author but not reported in the target article. Comprehension check questions were newly designed for this replication and extension study and are not from the target article.

We used comprehension checks to ensure that participants read and understood the instructions, with multiple choice questions that participants had to answer correctly in order for them to proceed to the task. These questions were as follows: ‘what type of helping behaviour are you asked to recall?’ and ‘whose helping behaviour are you asked to recall?’. Then, participants recalled the described experience and rated their thoughts and emotions that they were feeling in that experience. After that, they proceeded to type the details of the situation in the given box. We then asked about their current emotions in response to the recalled experience with the seven-point gratitude and indebtedness scales used in the target article. They also rated the helpers’ intention and the magnitude of the favour in the experience.

In the replication of study 3, participants were randomly assigned to read one of the three scenarios, namely benevolent, selfish-ulterior and ambiguous. We instructed them to imagine themselves in the scenario. We used the scenarios from the target article, followed by comprehension checks, including questions about what favour was offered in the scenario and why the benefactor offered the favour. The remaining dependent measures, including gratitude and indebtedness scales, favour magnitude scale and helper intention scale, were identical to the items provided in study 2.

Finally, they moved on to the extension. We asked about their inclination to reciprocate. We picked three items from the thought/action readiness items [6] relevant to reciprocation. Justifications were included in the Measures section below.

Table 4. Study 3: Replication and extension experimental design.

Independent variable (IV):	<u>Benevolent condition</u>	<u>Selfish-ulterior condition</u>	<u>Ambiguous condition</u>
Helper intentions (between-subject)	'You can tell that your friend is really concerned about you and wants to help you out, so you say yes'	'You know that your friend is really doing you this favour in order to borrow your car next weekend, but you really need those textbooks, so you say yes'	'You really need those textbooks, so you say yes. The next weekend that same friend asks you if they can borrow your car to run some errands'
Dependent variables (DV)	DV1: <u>Gratitude</u> 'Please choose the number by each adjective to indicate the extent to which you would feel each of these emotions in the scenario you just read.' emotion adjectives include: <i>grateful</i>, <i>thankful</i>, and <i>appreciative</i>. Scale: 1 = <i>Would feel very little of this emotion</i>, 7 = <i>Would feel a lot of this emotion</i> (source: [8]) DV2: <u>Indebtedness</u> 'Please choose the number by each adjective to indicate the extent to which you would feel each of these emotions in the scenario you just read.' emotion adjectives include: <i>indebted</i> and <i>obligated</i>. Scale: 1 = <i>Would feel very little of this emotion</i>, 7 = <i>Would feel a lot of this emotion</i> (source: [5]) DV3: <u>Other emotions</u> 'Please choose the number by each adjective to indicate the extent to which you would feel each of these emotions in the scenario you just read.' emotion adjectives include <i>calm</i>, <i>pleased</i>, <i>resentful</i>, <i>upset</i>, and <i>annoyed</i>. Scale: 1 = <i>Would feel very little of this emotion</i>, 7 = <i>Would feel a lot of this emotion</i> DV4 (manipulation check): <u>Perceived helpers' motivations</u> 'Please rate the friend's motivations on the following scale.' Scale: 1 = <i>Very concerned about me</i>, 7 = <i>Motivated mostly by selfish reasons</i> DV5: <u>Loaning experience</u> 'Have you ever had a friend loan you money for textbooks?' Choice: <i>yes</i> or <i>no</i> DV6 (extension): <u>Perceived expectations for reciprocity</u> 'Please rate your understanding of the other person's expectations of you to reciprocate.' Scale: 1 = <i>No expectations to reciprocate</i>, 7 = <i>Very high expectations to reciprocate</i> DV7 (extension): <u>Reciprocity inclination</u> 'To what extent would you have the urge to act in the following ways?' Items include: (i) <i>I would feel like helping my friend in return</i>, (ii) <i>I would feel like giving my friend a gift in return</i>, and (iii) <i>I would feel like doing something for my friend in return</i>. Scale: 1 = <i>Slight urge</i>, 7 = <i>Very strong urge</i> (source: [6])		
Covariate (C)	<u>Magnitude of the favour</u> 'How much of a favour do you think the friend did by giving money for the textbooks?' Scale: 1 = <i>A very small favour</i>, 7 = <i>A very big favour</i>		
Comprehension checks (CC)	Specific CC questions include: (Q1) 'How much money did the friend offer to give to help pay for the textbooks?' (Q2) 'What was the favour offered in the scenario?' (Q3) 'According to the text: why is your friend offering to help you?'		

Note: DV3 and DV5 were found in the materials provided by the author but not reported in the target article. CC-Q1 was extracted from the target article materials. CC-Q2 and Q3 were newly designed for this replication and extension study.

2.5. Manipulations of helper intentions

2.5.1. Study 2

We manipulated the perceived helper intentions using a recall task (table 3). Participants in the benevolent condition were expected to rate the helper's motivations as less selfish in the manipulation checks.

2.5.2. Study 3

We manipulated the helper intentions in a vignette according to the condition assigned (i.e. benevolent, ambiguous and selfish-ulterior conditions; table 4). Participants were expected to rate the helper as less selfish in the benevolent condition compared to the selfish-ulterior condition in the manipulation checks.

2.6. Measures

2.6.1. Replication

2.6.1.1. Emotional responses on gratitude and indebtedness

We adopted the gratitude and indebtedness scales used in the target article. Specifically, the gratitude scale consisted of the emotional adjectives 'grateful', 'thankful' and 'appreciative' (study 2: $\alpha = 0.97$; study 3: $\alpha = 0.91$) and the indebtedness scale consisted of 'indebted' and 'obligated' (study 2: $\alpha = 0.72$; study 3: $\alpha = 0.65$). Both were on a seven-point Likert scale (1 = *would feel very little of this emotion*; to 7 = *would feel a lot of this emotion*). We took an average for these adjectives to get an overall score of gratitude and indebtedness individually.

We also found five emotions, i.e. 'calm', 'pleased', 'resentful', 'upset' and 'annoyed', which were on the scale from the target article's materials provided by the author, but they were not included in the analysis of gratitude and indebtedness. We followed the target article and added them to the data collection.

2.6.1.2. Helper intention (manipulation check)

We adopted the helper intention scale from the target article. Participants rated the helper's intention in the situation from 1 = *very concerned about me*; to 7 = *motivated mostly by selfish reasons*.

2.6.1.3. Magnitude of the favour

We adopted the favour magnitude scale from the original material. Items were as follows: 'how big of a favor do you think the other person did for you?', and 'how costly (in terms of money, time, effort, etc.) do you think this situation was for the person who did something good for you?' ($\alpha = 0.71$). Participants rated the magnitude of the favour from 1 = *a very small favour*; to 7 = *a very big favour*.

2.6.2. Extensions

2.6.2.1. Perceived expectations for reciprocity

We asked subjects to rate their perceived reciprocity expectations of the benefactor from 1 = *no expectations to reciprocate*; to 7 = *very high expectations to reciprocate*. We note that although the target article did not set off to manipulate expectations, study 3 did vary expectations for reciprocity between the conditions with specific mention of such expectations in the selfish condition. As one reviewer noted, this measure could be considered a manipulation check examining the impact of making that expectation explicit in one of the conditions.

Table 5. Comparison of the target article to our replication.

	Target article	Replication	Reasons for change
Study design	Participants completed the studies with pen and paper in the laboratories	Participants completed the studies with an online survey	To reach more and a wider variety of participants; to conduct the studies with lower cost and higher efficiency
Sample characteristics	Sample size: study 2: 92; study 3: 86 Geographical origin: Undergraduates studying at Baylor University	$n = 759$ Students at the online research platform Prolific	Generalizability of results by including a larger more diverse sample of participants
Procedure	Items on gratitude and indebtedness were not randomized	Items on gratitude and indebtedness were randomized	To reduce the order effect
	Study 2 and study 3 were conducted separately	The replication of study 2 and the replication of study 3 were conducted with the same participants in one setting	Potentially explore consistency in participants' answers across the two studies (whether an answer in one study is predictive of an answer in the other study) and order/decline effects
		The order of the replications of study 2 and study 3 was randomized	To explore and address potential order effects Allows subsequent separate analysis on participants who took study 2/3 as their first presented study
Procedure	No comprehension check	We added comprehension checks for replication of study 2 and study 3	To ensure that the participants read and understood the instructions and scenarios
Conditions	No change	No change	NA

2.6.2.2. Inclination to reciprocate

We picked three items from the thought/action readiness items [6] to measure the inclination to reciprocate ($\alpha = 0.83$): 'I would feel like helping my friend in return', 'I would feel like giving my friend a gift in return' and 'I would feel like doing something for my friend in return'.

We chose these items based on their relevance to the reciprocation inclination, excluding items about affect (e.g. 'I would feel like thinking positive thoughts or happy memories about my friend') and those irrelevant to reciprocity (e.g. 'I would feel like ignoring my friend'), focusing on those about actual reciprocation. To better suit our study goal of reciprocity, we slightly modified the items by adding the words 'in return' at the end of the sentences (e.g. I would feel like helping my friend *in return*). It was originally a five-point Likert scale about the inclination to have certain thoughts and actions. We changed it to seven-point (1 = *slight urge*;¹ to 7 = *very strong urge*) to align this measure with the other measures in the study so as to not confuse participants in shifting scale ranges. We took an average for these items to get an overall score of reciprocity inclination.

¹Stage 2 note: we implemented this scale with 1 = 'slight urge' given that this was the scale used in the target article. For future research, researchers might consider the option of offering participants the option of 'no urge', either as an added 0 value or to replace the lowest value of 1.

Table 6. Classification of the replication, based on LeBel *et al.* [42].

Design facet	Replication	Details of deviation
Effect/hypothesis	Same	
IV construct	Same	
DV construct	Same	
IV operationalization	Same	
DV operationalization	Same	
Population (e.g. age)	Similar	Target article: the study recruited students from Baylor University in the United States Replication: we targeted students on the online research platform Prolific
IV stimuli	Similar	Target article: two groups of subjects were recruited to receive stimuli from study 2 and study 3, respectively Replication: the same participants answered both the replication of study 2 and the replication of study 3 Target article: items not randomized Replication: items randomized
DV stimuli	Same	
Procedural details	Similar	Target article: one comprehension check Replication: one extra comprehension check was added
Physical settings	Different	Target article: participants completed the studies with pen and paper in a laboratory setting Replication: participants completed the studies online, recruited through Prolific
Contextual variables	Different	Different time and context
Replication classification	Close replication	

2.7. Evaluation criteria for replication findings

There were 16 effect sizes calculated from the target study (see [table 1](#)). We compared the replication effects with the corresponding original effects calculated from the target article using the criteria set by LeBel *et al.* [41] (see the electronic supplementary material, Replication evaluation).

2.8. Replication closeness evaluation and deviations

We deviated from the target article in a few aspects, summarized in [table 5](#). We evaluated the classification of the replications using the criteria by LeBel *et al.* [42], summarized in [table 6](#) (see also the electronic supplementary material, Replication closeness evaluation). We summarized the replication as a ‘close’ replication.

2.9. Data analysis strategy

2.9.1. Replication: as in the target article

In both the replication of study 2 and the replication of study 3, to mirror the target article’s analyses, we first ran (Pearson’s) correlation tests to examine the associations between gratitude and indebtedness across conditions and then in the separate benevolent and selfish helper intention conditions.

In study 2, we used ANCOVAs to examine the effect of helper intention (benevolent versus selfish) on gratitude and indebtedness, with the magnitude of favour as the covariate. We supplemented those

Table 7. Studies 2 and 3 replication and extensions: descriptives.

Replication study and factors	Benevolent	Selfish/ulterior	Ambiguous
Replication of Study 2	(<i>n</i> = 381)	(<i>n</i> = 378)	
Gratitude	6.54 [0.85]	4.05 [1.92]	
Indebtedness	3.91 [1.80]	3.49 [1.84]	
Perceived helper intention	1.91 [1.39]	5.28 [1.46]	
Magnitude of the favour	4.71 [1.44]	3.53 [1.59]	
Perceived expectations for reciprocity (extension)	2.14 [1.71]	4.78 [1.99]	
Replication of Study 3	(<i>n</i> = 251)	(<i>n</i> = 254)	(<i>n</i> = 254)
Gratitude	6.59 [0.79]	5.48 [1.24]	6.14 [1.09]
Indebtedness	5.09 [1.42]	5.09 [1.58]	5.09 [1.43]
Perceived helper intention	1.71 [1.17]	4.81 [1.36]	3.02 [1.59]
Magnitude of the favour	6.27 [0.99]	5.67 [1.32]	6.16 [1.09]
Perceived expectations for reciprocity (extension)	3.65 [1.79]	5.89 [1.18]	4.83 [1.83]
Reciprocity inclination (extension)	6.44 [0.84]	5.76 [1.2]	6.11 [0.98]

Note: mean [s.d.] (condition sample size).

with regression analyses using the same factors which served a similar purpose to the ANCOVA and merely meant to mirror that target article's analyses and reported effects.

In study 3, we used one-way ANOVAs to examine the impact of helper intention (benevolent versus selfish-ulterior versus ambiguous) on gratitude and indebtedness. After that, we conducted planned comparisons to examine the differences in emotions between helper intention conditions.

2.9.2. Replication: extension analyses

In both studies in the target article, the comparison between gratitude and indebtedness was done by comparing signals, in which support was found for intent as affecting gratitude but no support for affecting indebtedness. We reframed this to a comparison of the effects of the two dependent variables. To complement the original analyses, we conducted extension analyses of a two-way mixed ANOVA, with helper intent conditions as a between-subject factor (benevolent versus selfish-ulterior in study 2, and benevolent versus selfish-ulterior versus ambiguous in study 3), emotion type as a within-subject factor (gratitude versus indebtedness), and emotion ratings as the dependent variable.

2.9.3. Extensions

We conducted independent samples' Welch's *t*-tests (two-tailed) to examine the differences in perceived expectations for reciprocity and reciprocity inclination respectively between the benevolent and selfish-ulterior conditions. Then, we used correlation tests (Pearson's) to examine the association between perceived reciprocity expectations and reciprocity inclination with the two emotions.

2.9.4. Order effects and outliers and exclusions

Following our stage 1 pre-registration plan, we did not classify any exclusions or outliers.

One deviation from the target article was that all participants completed all scenarios in random order. We consider this to be a stronger design with many advantages, yet one disadvantage is that answers to one scenario may bias participants' answers to the following scenarios. We pre-registered that if we fail to find support for our hypotheses that we would run exploratory analyses for the failed study by (i) focusing on the participants that completed that study first and examine order as a moderator ($\alpha = 0.005$), and (ii) excluding those who failed the manipulation checks ($\alpha = 0.001$). We

Table 8. Replication: Summary of statistical tests and results interpretation.

Hypothesis	Statistical tests	Target article			Replication			Interpretation
		<i>p</i>	Effect size	CI	<i>p</i>	Effect size	CI	
1a	Pearson correlation	<0.001*	<i>r</i> = 0.57	[0.41, 0.69]	<0.001	<i>r</i> = 0.34	[0.28, 0.40]	Signal—inconsistent, smaller
1b	Pearson correlation	<0.001*	<i>r</i> = 0.61	[0.39, 0.76]	<0.001	<i>r</i> = 0.53	[0.46, 0.60]	Signal—inconsistent, smaller
1c	Pearson correlation	>0.20	<i>r</i> = 0.20	[−0.10, 0.47]	= 0.70	<i>r</i> = 0.02	[−0.08, 0.12]	No signal—consistent
2	ANCOVA	<0.001*	η_p^2 = 0.2	[0.08, 0.32]	<0.001	η_p^2 = 0.33	[0.28, 0.37]	Signal—inconsistent, larger
3	ANCOVA	>0.20	η_p^2 = 0.01	[0.00, 0.08]	= 0.54	η_p^2 < 0.001	[0.00, 0.01]	No signal—consistent
2 + 3	Mixed ANOVA (extension)	/	/	/	<0.001	η_p^2 = 0.23	[0.19, 0.27]	Fully supported
4	Linear regression	<0.0001*	<i>R</i> ² = 0.73	[0.61, 0.81]	<0.001	<i>R</i> ² = 0.53	[0.48, 0.58]	Signal—inconsistent, smaller
		(i) <0.01	β = 0.32		(i) <0.001	β = −0.51		
		(ii) <0.01	β = 0.62		(ii) <0.001	β = 0.37		
5	Linear regression	<0.001*	<i>R</i> ² = 0.26	[0.10, 0.41]	<0.0001	<i>R</i> ² = 0.14	[0.10, 0.19]	Signal—inconsistent, smaller
		(i) >0.20	β = 0.13		(i) = 0.540	β = 0.02		
		(ii) <0.001	β = 0.42		(ii) <0.001	β = 0.38		
6	Pearson correlation	<0.05*	<i>r</i> = 0.42	[0.06, 0.68]	<0.001	<i>r</i> = 0.23	[0.11, 0.34]	Signal—inconsistent, smaller
7a	One-way ANOVA	<0.01*	η_p^2 = 0.14	[0.03, 0.26]	<0.001	η_p^2 = 0.16	[0.12, 0.20]	Signal—consistent
7b	Independent <i>t</i> -test (two-tailed)	<0.05*	<i>d</i> = 0.55	[0.02, 1.08]	<0.001	<i>d</i> = 0.42	[0.25, 0.60]	Signal—consistent
7c	Independent <i>t</i> -test (two-tailed)	0.07	<i>d</i> = 0.49	[−0.04, 1.01]	<0.001	<i>d</i> = 0.63	[0.45, 0.81]	Signal—inconsistent, positive effect
7b/c + 8b/c	Mixed ANOVA (extension)	/	/	/	<0.001	η_p^2 = 0.07	[0.05, 0.10]	Fully supported
8a	One-way ANOVA	>0.20	η_p^2 = 0.00	[0.00, 0.03]	= 1.00	η_p^2 < 0.01	[0.00, 1.00]	No signal—consistent

(Continued.)

Table 8. (Continued.)

Hypothesis	Statistical tests	Target article			Replication			Interpretation
8b	Independent <i>t</i> -test (two-tailed)	>0.20	$d = 0.13$	$[-0.39, 0.64]$	$= 0.994$	$d = 0.00$	$[-0.17, 0.18]$	No signal—consistent
8c	Independent <i>t</i> -test (two-tailed)	>0.20	$d = 0.03$	$[-0.49, 0.55]$	$= 0.969$	$d = 0.00$	$[-0.18, 0.16]$	No signal—consistent
9	Pearson correlation	$<0.05^*$	$r = -0.40$	$[-0.67, -0.04]$	<0.001	$r = -0.50$	$[-0.59, -0.40]$	Signal—consistent
10	Pearson correlation	>0.20	$r = 0.00$	$[-0.37, 0.37]$	$= 0.80$	$r = -0.21$	$[-0.14, 0.11]$	No signal—consistent

Note: see [table 1](#) for hypotheses. For partial eta-squared, we report 90% CIs instead of 95% in order not to include zero with the *p*-value falling below 0.05 [\[43\]](#) Lakens, 2014. The interpretation of replication outcome is based on an evaluation criteria by LeBel *et al.* [\[41\]](#).

Table 9. Extensions: summary of statistical tests.

Hypothesis	Stat. tests	d.f.	<i>p</i>	Effect size	CI
11	Independent <i>t</i> -test (two-tailed)	433	<0.001	<i>d</i> = 1.51	[1.31, 1.71]
12	Independent <i>t</i> -test (two-tailed)	455	<0.001	<i>d</i> = 0.66	[0.48, 0.84]
13	Pearson correlation	757	=0.08	<i>r</i> = −0.06	[−0.13, 0.01]
14	Pearson correlation	757	<0.001	<i>r</i> = −0.28	[−0.35, −0.22]
15	Pearson correlation	757	<0.001	<i>r</i> = 0.17	[0.10, 0.24]

Note: see [table 1](#) for all hypotheses. CI = 95% confidence intervals.

concluded a successful replication, and so according to the pre-registration did not plan for additional order analyses, yet to address a request by a reviewer in stage 2 to help readers better understand the possible impact of order on the findings, we conducted an exploratory analysis of the data focusing on the findings when studies were presented first. We provided Rmarkdown code employing a filter that allows the analysis to run on the full high-power sample or on the subset where studies were presented first, included in our OSF. We compared the set of results and concluded the findings to be highly consistent with no major changes.

3. Results

We summarized descriptive statistics in [table 7](#) and statistical tests in [tables 8](#) and [9](#). Our analyses were all performed with R (version: 4.1.2), and we used ggstatsplot [44] and jamovi [45] jmv package for our analyses and figures.

3.1. Replication

3.1.1. Study 2

3.1.1.1. Manipulation check: helper intention

We conducted an independent samples *t*-test (Welch's, two-tailed) and found that participants in the benevolent condition rated the helper's motivations as less selfish ($n = 381$, $M = 1.91$, s.d. = 1.39) than in the selfish-ulterior condition ($n = 378$, $M = 5.28$, s.d. = 1.46; $M_d = -3.37$; $t_{754} = -33$, $p < 0.001$; $d = -2.40$, 95% CI [−2.21, −2.58]).

We also found that participants rated the magnitude of favour as larger in the benevolent condition ($n = 381$; $M = 4.71$, s.d. = 1.44) than in the selfish-ulterior condition ($n = 378$; $M = 3.53$, s.d. = 1.59; $M_d = 1.18$; $t_{748} = 11$, $p < 0.001$; $d = 0.80$, 95% CI [0.65, 0.95]).

3.1.1.2. Associations between gratitude and indebtedness

We conducted Pearson's correlation tests and found support for a positive association between gratitude and indebtedness, both across conditions (hypothesis 1a: $r_{757} = 0.34$, 95% CI [0.28, 0.40], $p < 0.001$), and in the selfish-ulterior condition (hypothesis 1b: $r_{376} = 0.53$, 95% CI [0.46, 0.60], $p < 0.001$; [figure 1](#)), and as expected we failed to find support for an association in the benevolent condition (hypothesis 1c reframed from a null hypothesis: $r_{379} = 0.02$, 95% CI [−0.08, 0.12], $p = 0.7$; [figure 2](#)).

3.1.1.3. Core hypothesis: impact of helper intent (benevolent > selfish) on gratitude is stronger than on indebtedness

We conducted ANCOVAs, with the rated magnitude of favour as a covariate and found support for differences in gratitude between the benevolent condition and the selfish-ulterior condition (H2: $F_{1, 756}$

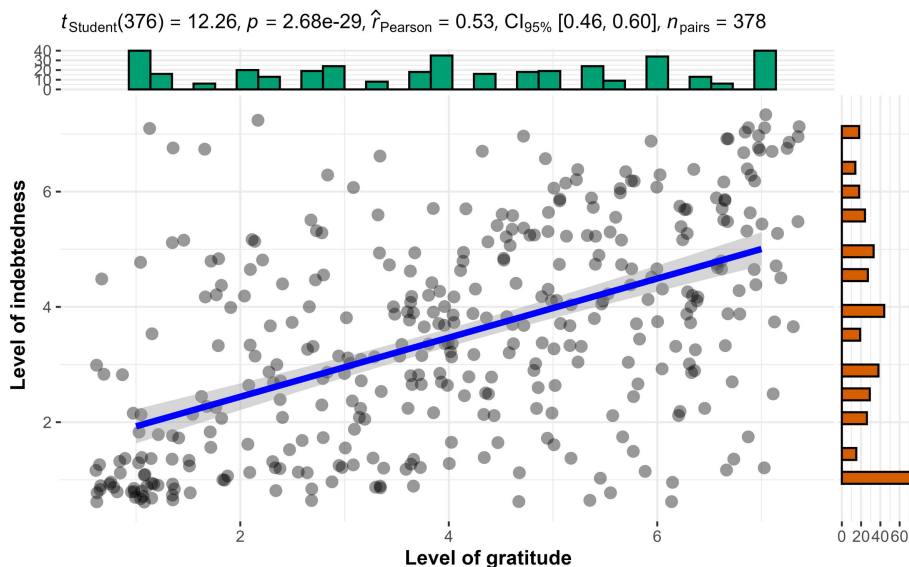


Figure 1. Study 2 selfish-ulterior condition (H1b): association between gratitude and indebtedness.

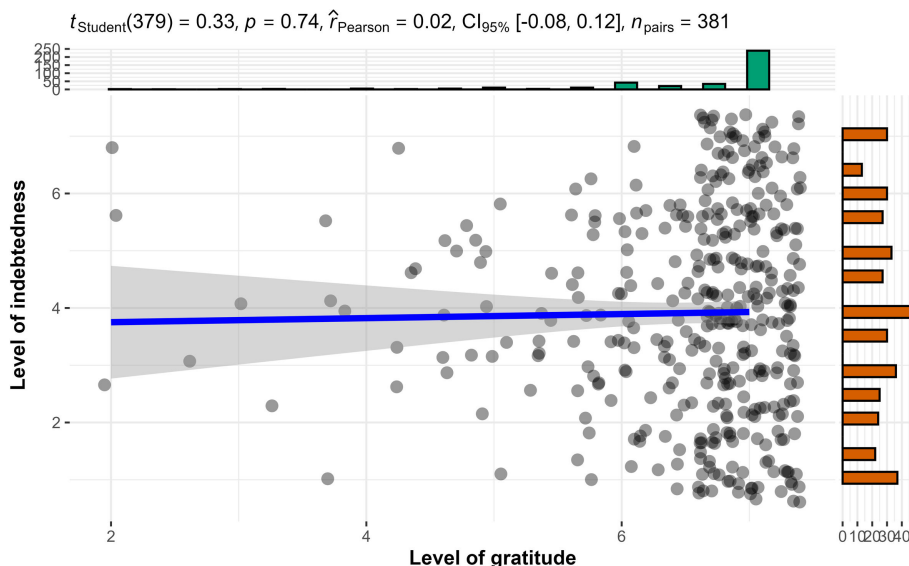


Figure 2. Study 2 benevolent condition (H1c null hypothesis): association between gratitude and indebtedness.

$= 365$, $p < 0.001$; $\eta_p^2 = 0.33$, 90% CI [0.28, 0.37]; figure 3), but not in indebtedness (H3 null hypothesis: $F_{1, 756} = 0.37$, $p = 0.54$; $\eta_p^2 = 0.001$, 90% CI [0.00, 0.01]; figure 4). We, therefore, concluded support for the combination of hypotheses 2 and 3 for higher gratitude when recalling a benevolent favour compared to a selfish favour, after controlling for the magnitude of the favour, but less so for indebtedness.

3.1.1.4. Complementary regression analyses

We conducted regression analyses with a condition variable contrasting benevolent and selfish helper intent and magnitude of favour predicting gratitude and indebtedness. For gratitude, we found support for both intention conditions ($\beta = -0.51$, $t_{756} = -19.11$, $p < 0.001$) and magnitude of favour ($\beta = 0.37$, $t_{756} = 13.84$, $p < 0.001$) as predictors of feelings of gratitude ($R^2 = 0.53$, 95% CI [0.48, 0.58], $F_{2, 756} = 430.49$, $p < 0.001$).

We found support for magnitude of favour ($\beta = 0.38$, $t_{756} = 10.46$, $p < 0.001$), but not for intention conditions ($\beta = 0.02$, $t_{756} = 0.61$, $p = 0.54$), as predictors of feelings of indebtedness ($R^2 = 0.14$, 95% CI [0.10, 0.19], $F_{2, 756} = 60.52$, $p < 0.001$).

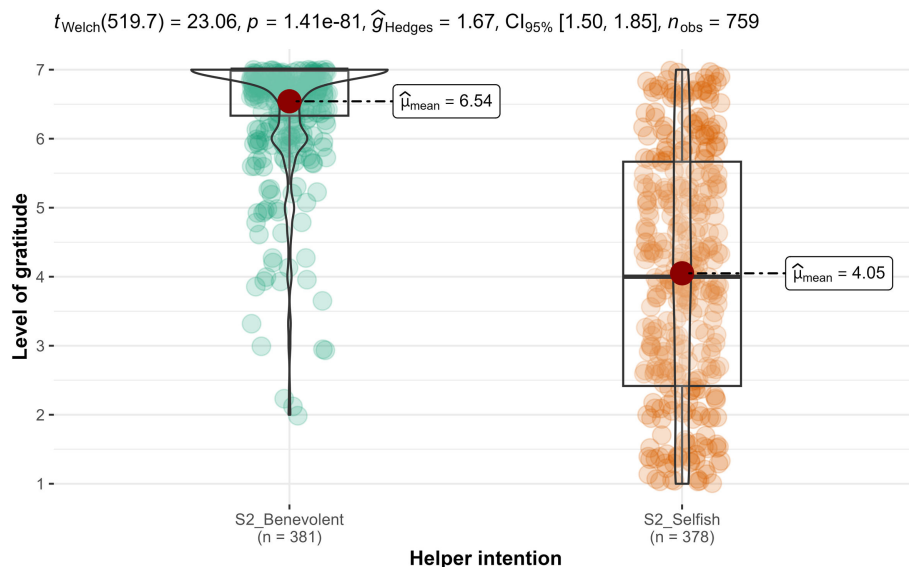


Figure 3. Study 2: helper intention impact on gratitude. Note: scale is from 1 to 7; higher values indicate stronger feelings of gratitude.

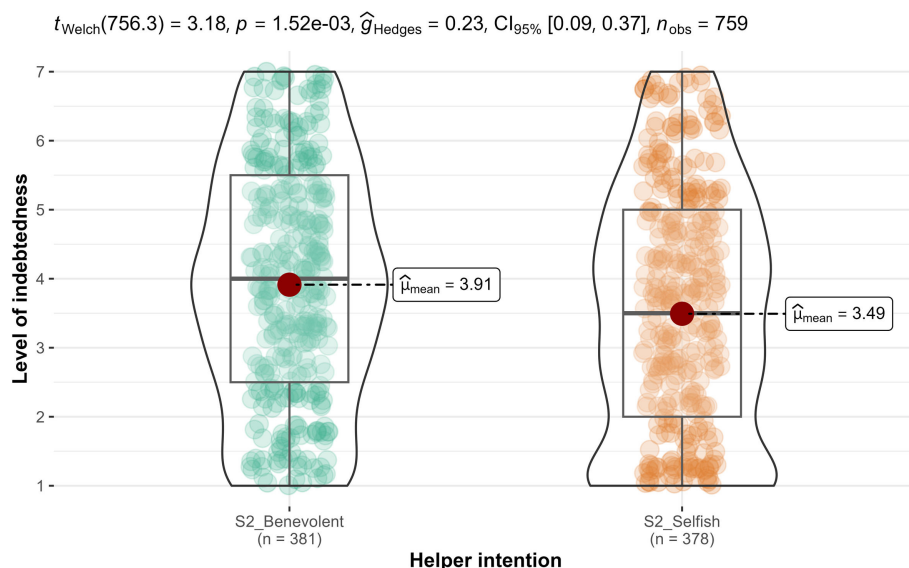


Figure 4. Study 2: helper intention impact on indebtedness. Note: scale is from 1 to 7; higher values indicate stronger feelings of indebtedness.

We concluded support for hypothesis 4 (2r), that helpers' intent (benevolent versus selfish-ulterior) and magnitude of favour predict gratitude, but—as expected—not for hypothesis 5 (3r null hypothesis) that helpers' intent predicts indebtedness.

3.1.1.5. Interaction between intent and emotions (gratitude versus indebtedness): extension analysis of a direct test to core hypothesis

We conducted a mixed ANOVA examining the interaction between intent (benevolent versus selfish-ulterior; between-subject) and emotion type (gratitude versus indebtedness; repeated) and found evidence of main effect for emotion type ($F_{1, 757} = 532.4, p < 0.001, \eta_p^2 = 0.41, 95\% \text{ CI } [0.37, 0.46]$), evidence of main effect for intent ($F_{1, 757} = 216.6, p < 0.001, \eta_p^2 = 0.22, 95\% \text{ CI } [0.18, 0.27]$) and evidence of interaction between intent and emotion type ($F_{1, 757} = 225.0, p < 0.001, \eta_p^2 = 0.23, 95\% \text{ CI } [0.19, 0.27]$). We plotted the findings in figure 5.

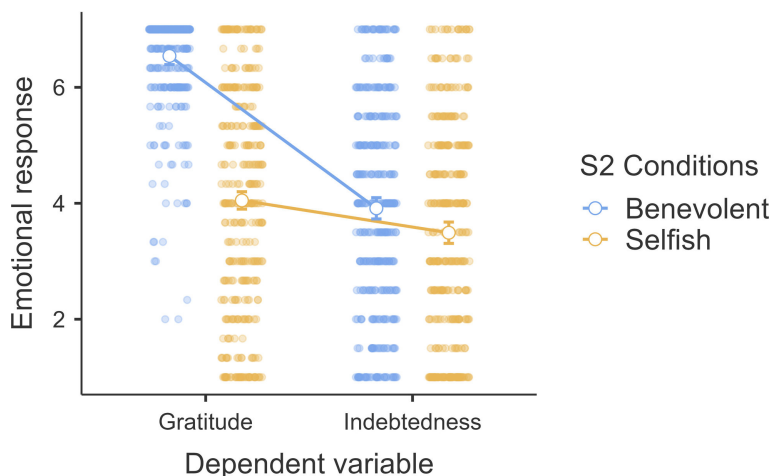


Figure 5. Study 2: interaction between helper intention and emotions. Note: scale is from 1 to 7; higher values indicate stronger feelings of the emotion.

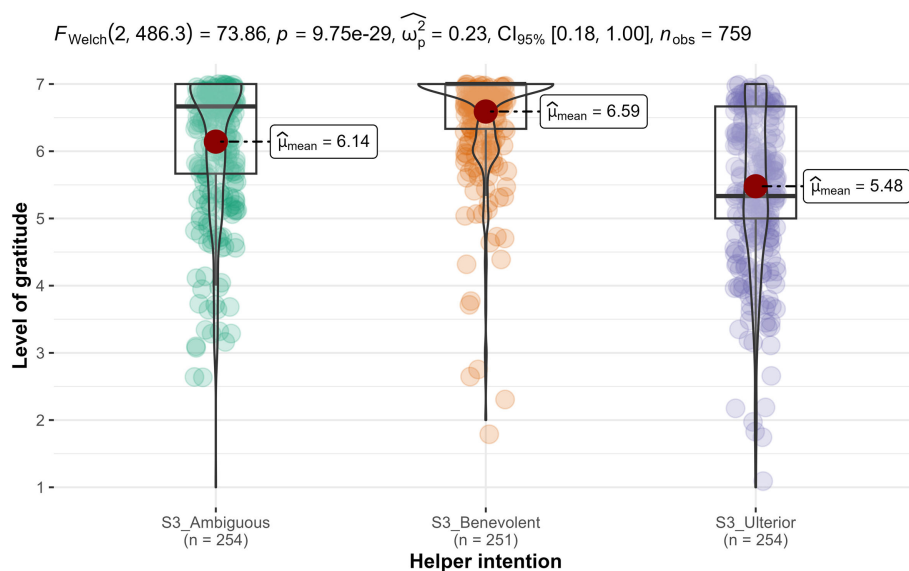


Figure 6. Study 3: helper intention impact on gratitude. Note: scale is from 1 to 7; higher values indicate stronger feelings of gratitude.

3.1.2. Study 3

3.1.2.1. Manipulation check: helper intention

We conducted independent samples *t*-tests (Welch's; two-tailed) and found that participants rated benefactor as being less selfish in the benevolent condition ($n = 251$; $M = 1.71$, s.d. = 1.17) compared to the selfish-ultior condition ($n = 254$; $M = 4.81$, s.d. = 1.36; $M_d = 3.10$; $t_{493} = 28$, $p < 0.001$; $d = 2.49$, 95% CI [2.26, 2.72]). We also found that participants rated helper as being more selfish in the selfish-ultior condition compared to the ambiguous condition ($n = 254$; $M = 3.02$, s.d. = 1.59; $M_d = 1.79$; $t_{494} = 14$, $p < 0.001$; $d = 1.24$, 95% CI [1.05, 1.43]). Participants rated helpers as being less selfish in the benevolent condition compared to the ambiguous condition ($t_{464} = -11$, $p < 0.001$; $d = -0.98$, 95% CI [-1.16, -0.79]).

3.1.2.2. Covariate: magnitude of favour

We conducted independent samples *t*-tests (Welch's; two-tailed) and found that the rated magnitude of favour in the selfish-ultior condition ($n = 254$; $M = 5.67$, s.d. = 1.32) was lower than in the benevolent condition ($n = 251$; $M = 6.27$, s.d. = 0.99; $M_d = -0.60$; $t_{469} = -5.8$, $p < 0.001$; $d = -0.52$, 95% CI [-0.69, -0.34]).

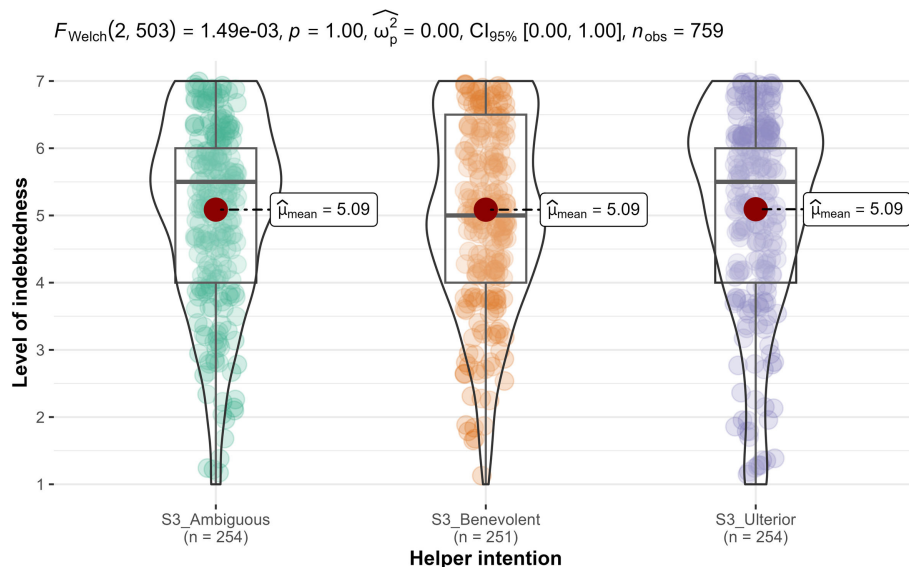


Figure 7. Study 3: helper intention impact on indebtedness. Note: scale is from 1 to 7; higher values indicate stronger feelings of indebtedness.

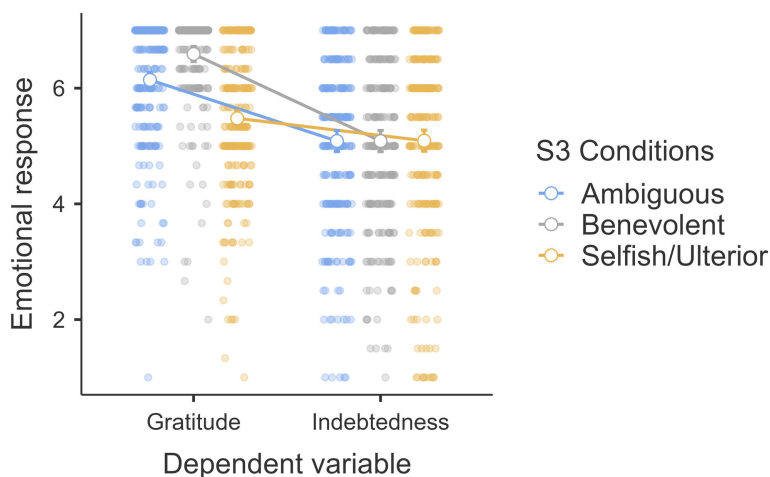


Figure 8. Study 3: interaction between helper intention and emotions (H7b/c + H8 b/c). Note: scale is from 1 to 7; higher values indicate stronger emotional response.

and in the ambiguous condition ($n = 254$; $M = 6.16$, s.d. = 1.09; $M_d = -0.49$; $t_{489} = -4.5$, $p < 0.001$; $d = -0.40$, 95% CI $[-0.58, -0.22]$). However, we found no support for the magnitude of favour in the benevolent condition as different from that in the ambiguous condition ($t_{499} = 1.2$, $p = 0.20$; $d = 0.11$, 95% CI $[-0.07, 0.28]$).

3.1.2.3. Associations between gratitude and indebtedness

We conducted Pearson's correlation tests and found support for a positive association between gratitude and indebtedness in the benevolent condition ($r_{249} = 0.25$, 95% CI $[0.13, 0.36]$, $p < 0.001$), the selfish-ultior condition ($r_{252} = 0.17$, 95% CI $[0.05, 0.29]$, $p < 0.01$) and the ambiguous condition ($r_{252} = 0.23$, 95% CI $[0.11, 0.34]$, $p < 0.001$). We, therefore, concluded support for hypothesis 6, that gratitude is positively associated with indebtedness in ambiguous conditions.

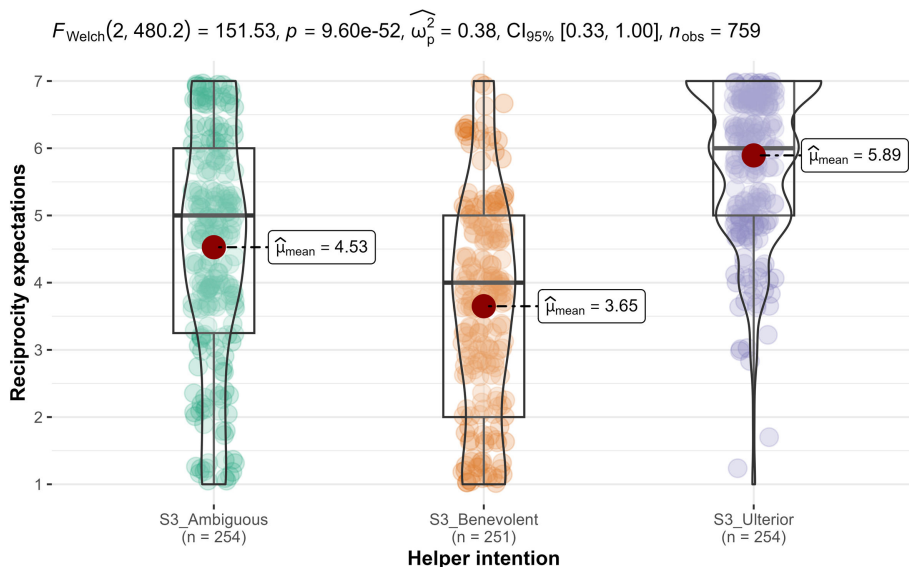


Figure 9. Study 3: helper intention impact on reciprocity expectations. Note: scale is from 1 to 7; higher values indicate higher reciprocity expectations.

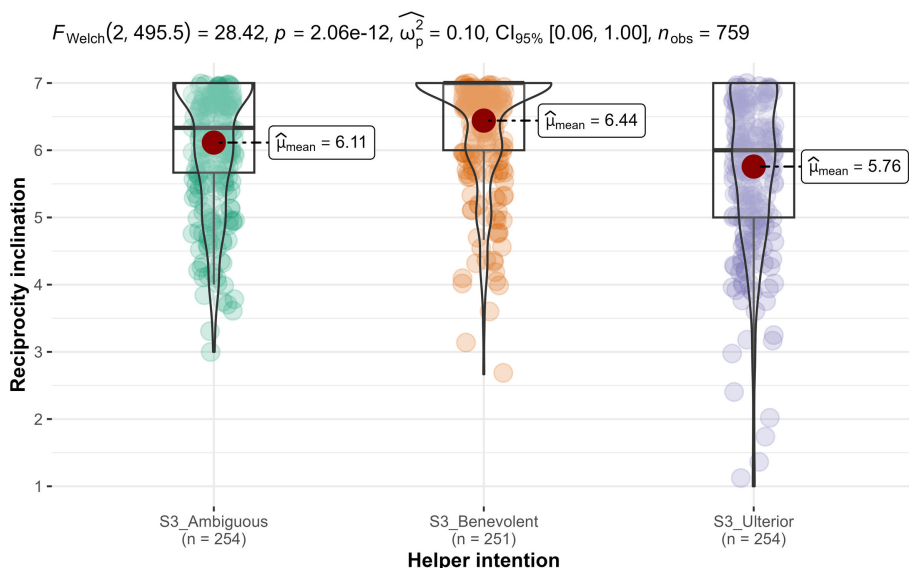


Figure 10. Study 3: helper intention impact on reciprocity inclination. Note: scale is from 1 to 7; higher values indicate higher reciprocity inclination.

3.1.2.4. Core hypothesis: impact of intent (benevolent > ambiguous > selfish) on gratitude is stronger than on indebtedness

We conducted one-way ANOVAs and found support for helper intention's impact on gratitude ($F_{2, 756} = 70.9, p < 0.001; \eta_p^2 = 0.16, 90\% \text{ CI } [0.12, 0.20]$; figure 6), but not on indebtedness ($F_{2, 756} = 0, p = 1.0; \eta_p^2 = 0.00$; figure 7). Therefore, we concluded support for hypothesis 7a, that gratitude is different across the three conditions (benevolent, selfish-ultior and ambiguous), and—as expected—no support for hypothesis 8a (reframed from null hypothesis) that indebtedness is different across the three conditions.

Following the ANOVAs, we conducted *post hoc* contrasts analyses for hypotheses 7b and 7c and found support for feelings of gratitude in the ambiguous condition ($n = 254; M = 6.14, \text{ s.d.} = 1.09$) as weaker than in the benevolent condition ($n = 251; M = 6.59, \text{ s.d.} = 0.79; M_d = -0.45; t_{756} = -4.76, p < 0.001; d = -0.42, 95\% \text{ CI } [-0.60, -0.25]$; H7b), but stronger than in the selfish-ultior condition (H7c: $n = 254; M$

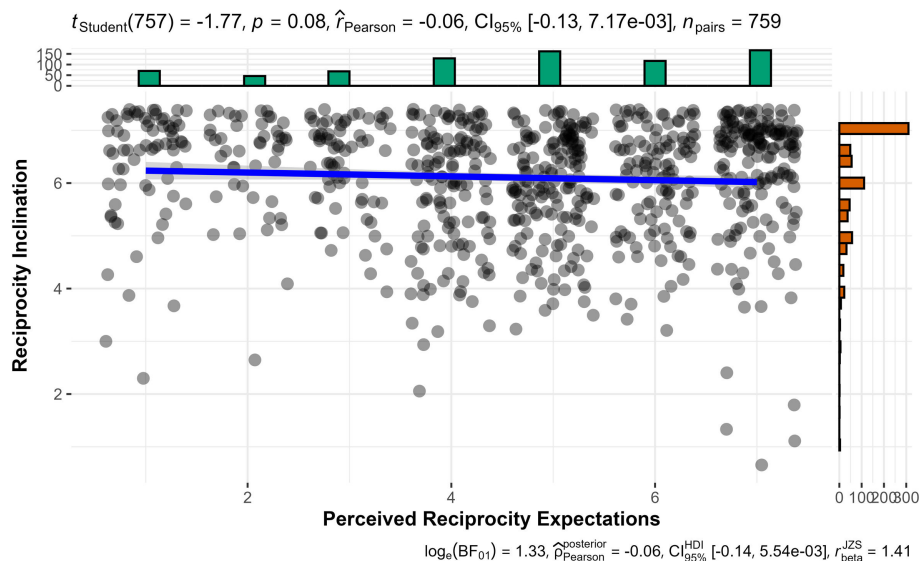


Figure 11. Study 3: the association between perceived expectations for reciprocity and reciprocity inclination.

= 5.48, s.d. = 1.24; $M_d = 0.66$; $t_{756} = 7.09$, $p < 0.001$; $d = 0.63$, 95% CI [0.45, 0.81]; between benevolent and selfish-ulterior: $t_{756} = 11.83$, $p < 0.001$; $d = 1.05$, 95% CI [0.87, 1.24]).

Also, we conducted *post hoc* contrasts analyses for hypotheses 8b and 8c (reframed from null hypotheses) and—as expected—found no support for differences in indebtedness comparing the benevolent condition ($M = 5.09$, s.d. = 1.42) to both the ambiguous condition ($M = 5.09$, s.d. = 1.58; $M_d = 0$; $t_{756} = 0.01$, $p = 0.9948$; $d = 0.00$, 95% CI [-0.17, 0.18]; H8b) and the selfish-ulterior condition ($M = 5.09$, s.d. = 1.43; $M_d = 0$; $t_{756} = -0.05$, $p = 0.969$; $d = 0.00$, 95% CI [-0.18, 0.16]; H8c).

3.1.2.5. Interaction between intent and emotions (gratitude versus indebtedness): extension analysis of a direct test to core hypothesis

We conducted a mixed ANOVA examining the interaction between intent (benevolent versus selfish-ulterior versus ambiguous; between-subject) and emotion type (gratitude versus indebtedness; repeated) and found support for a main effect for emotion type ($F_{1, 756} = 275.57$, $p < 0.001$, $\eta_p^2 = 0.27$, 95% CI [0.22, 0.31]), a main effect for intent ($F_{2, 756} = 19.72$, $p < 0.001$, $\eta_p^2 = 0.05$, 95% CI [0.03, 0.08]) and an interaction between intent and emotion type ($F_{2, 756} = 30.12$, $p < 0.001$, $\eta_p^2 = 0.07$, 95% CI [0.05, 0.10]; H7b/c and H8b/c combined; figure 8).

3.1.2.6. Associations between helper intentions, gratitude and indebtedness

We conducted correlation tests (Pearson's correlation) in the ambiguous condition and found support for more selfish intention as being negatively associated with gratitude ($r_{252} = -0.50$, 95% CI [-0.59, -0.40], $p < 0.001$), yet not associated with indebtedness ($r_{252} = -0.01$, 95% CI [-0.14, 0.11], $p = 0.80$).

3.2. Extensions: perceived expectations for reciprocity and reciprocity inclination

We added perceived expectations for reciprocity and reciprocity inclination as two extension dependent variables in study 3 and evaluated how they differ across conditions and are associated with gratitude and indebtedness.

First, we conducted independent samples *t*-tests (Welch's; two-tailed) and found support for higher perceived expectations for reciprocity in the selfish-ulterior condition ($n = 254$; $M = 5.89$, s.d. = 1.18) than in the benevolent condition ($n = 251$; $M = 3.65$, s.d. = 1.79; $M_d = -2.24$; $t_{433} = -17$, $p < 0.001$; $d = -1.51$, 95% CI [-1.71, -1.31]; figure 9). We found support for hypothesis 11a over 11b.

Then, we conducted independent samples *t*-tests (Welch's; two-tailed) and found support for higher reciprocity inclination in benevolent condition ($n = 251$; $M = 6.44$, s.d. = 0.84) than those in the selfish-ulterior condition ($n = 254$; $M = 5.89$, s.d. = 1.18; $M_d = 0.55$; $t_{433} = 17$, $p < 0.001$; $d = 0.44$, 95% CI [0.28, 0.60]; figure 10).

rior condition ($n = 254$; $M = 5.76$, s.d. = 1.2; $M_d = 0.68$; $t_{455} = 7.4$, $p < 0.001$; $d = 0.66$, 95% CI [0.48, 0.84]; figure 10). We found support for hypothesis 12b over 12a.

We then conducted correlation tests (Pearson's correlation) and found no support for a link between reciprocity inclination and perceived reciprocity expectations ($r_{757} = -0.06$, 95% CI [-0.13, 0.01], $p = 0.08$; figure 11). We failed to find support for hypothesis 13 that perceived reciprocity expectations is correlated with reciprocity inclination.

We also conducted correlation tests (Pearson's correlation) and found support for a negative correlation between perceived reciprocity expectations and gratitude ($r_{757} = -0.28$, 95% CI [-0.35, -0.22], $p < 0.001$) and a positive correlation between perceived reciprocity expectations and indebtedness ($r_{757} = 0.17$, 95% CI [0.10, 0.24], $p < 0.001$). We successfully found support for hypotheses 14 and 15 that gratitude and indebtedness are associated with perceived expectation for reciprocity.

Finally, we added exploratory correlations for the associations with reciprocity inclination. We found support for reciprocity inclination having a positive association with gratitude ($r_{757} = 0.52$ [0.46, 0.57], $p < 0.001$) and a weaker positive association with indebtedness ($r_{757} = 0.29$ [0.22, 0.35], $p < 0.001$; z for differences between correlations = 5.71, $p < 0.001$).

4. Discussion

In our replication and extensions registered report for the effects of helper intention on gratitude and indebtedness demonstrated by Tsang [1], results were mostly consistent with the findings reported in the target article (see table 8 for a summary).

4.1. Replication

Overall, we found that benevolent (versus selfish) intentions were more strongly associated with gratitude than with indebtedness. Comparison of hypotheses 2 and 3 in study 2 ($H2: \eta_p^2 = 0.33 > H3: \eta_p^2 = 0.001$) as well as hypotheses 7b/c and 8b/c in study 3 ($H7b/c: d = 0.42/0.63 > H8b/c: d = 0$) both revealed that impact of helper intent on gratitude is much stronger than on indebtedness. These results are consistent with the conceptualization by Tsang [1] of gratitude and indebtedness as distinct emotions and with different emotional levels depending on perceived benefactors' intentions.

Our replication reveals slightly weaker effect sizes for some of the effects compared to those reported in the original article. For example, the association between gratitude and indebtedness is weaker across conditions in study 2 (hypothesis 1a: original: $r = 0.57$; replication: $r = 0.34$) and in the ambiguous condition in study 3 (hypothesis 6: original: $r = 0.42$; replication: $r = 0.23$). However, all is still very consistent with the core hypotheses in the target article about the influence of helper intent on gratitude and indebtedness suggested by Tsang [1]. Overall, seven out of nine of the hypotheses in study 3 were consistent and shared remarkably similar effect sizes with the original article, indicating the robustness and replicability of the conceptualization by Tsang [1] on gratitude and indebtedness after more than one and a half decades.

Some may question the value of replication studies, asserting that highly cited studies are inherently reliable and replicable. We believe these reactions often stem from a hindsight bias (also known as the 'knew-it-all-along phenomenon') that many, including researchers, harbour towards replication studies. We previously demonstrated an ironic display of hindsight bias over the replicability of a classic experiment on hindsight bias (study 3; [46]). To try and address hindsight bias over the replicability of our target article, we conducted a prediction poll on Twitter/X on 26 March 2023 [47] and found that 24 out of 30 (80%) of the researchers in the community predicted an unsuccessful replication of studies 2 and 3 by Tsang [1], which is one of the lowest predictions among other targets included in the same Twitter/X poll. These predictions stand in strong contrast to the very successful replication we reported here and further highlights the importance of testing intuitions and the possible misperceptions that some may hold towards the importance of comprehensive independent registered reports of direct replication.

4.2. Extensions: perceived expectations for reciprocity and reciprocity inclination

We ran extensions examining the relationship between gratitude, indebtedness, perceived reciprocity expectations and reciprocity inclination. Our findings showed that: (i) perceived expectation for

reciprocity was lower in benevolent helping intent than in selfish-ulterior helping intent; (ii) reciprocity inclination was stronger in receiving a benevolent helping than a selfish-ulterior helping; (iii) expectation for reciprocity was negatively associated with gratitude and positively associated with indebtedness; and (iv) there was no indication that expectations for reciprocity is associated with reciprocity inclination. Overall, we found empirical support for the hypothesis that gratitude and indebtedness are correlated to expectation for reciprocity, yet failed to find support for the association between expectation for reciprocity and reciprocity inclination.

Our findings support the argument by Watkins *et al.* [6] that expectation for reciprocity would be associated with higher indebtedness but lower gratitude. Provided that benevolent helping intent is associated with lower expectations for reciprocity, then according to the experimental paradigm of Watkins *et al.* [6], it would be associated with decreased indebtedness and increased gratitude. Our extensions help link between Tsang [1] and Watkins *et al.* [6] into a more comprehensive theory that higher benevolent intent is correlated with lower expectations and therefore higher gratitude than indebtedness.

4.3. Implications, limitations, and directions for future research

We concluded a successful replication, yet noted several limitations in the way we implemented the replication. First, we did not take into account the inflation from 2006 at which the study was first held, to 2023 at which our replication was held. The stimuli used in our \$200 in 2006 is fairly equal to \$300 in 2023. This was a dilemma that we acknowledged when we conducted our replication, but we decided not to change the stimuli. We considered the target's claims and theory, but we did not find in the article or in any follow-up literature that indicated it as an important factor, and thus, we decided to conduct a direct replication for our current study without changing the value. Our successful replication with similar effect sizes with the target further supports the conclusion that this was not a critical factor.

Second, we have not performed a systematic review and meta-analysis of the literature pointing to the findings in the literature that built up on the target article. The scope for this direct replication with extensions was rather narrow and purely focused on the empirical effort to reproduce and replicate the original findings, and thus we decided to keep our literature review concise, mostly to explain how the target article was embedded in the broader literature. We believe our successful replication can serve as a cornerstone for future systematic reviews on research of gratitude and indebtedness and provide empirical evidence for a more comprehensive meta-analysis since the studies by Tsang [1] were first conducted.

Third, we added the expectation for reciprocity as an extra dependent variable on top of the original manipulation check of helper intentions, given that the scenario also explicitly mentioned expectations in the benevolent (versus selfish) condition. This could be a duplication. We decided to keep what we replicated and what we extended discrete for our direct replication, but we see room for regarding the expectation for reciprocity as a manipulation check, combining it with the original dependent variable of perceived helpers' motivations.

Fourth, we deviated from the target article's design by having participants take part in both the replication of study 2 and the replication of study 3 in random order. This may potentially introduce order effects, where responses to one scenario might influence responses to subsequent scenarios. However, by randomizing the sequence of two studies for each participant, we minimized potential biases. Our exploratory analyses indicated that this had little to no impact on the findings, with results consistent when comparing the full sample (as pre-registered) to studies presented first (exploratory). Given the strong alignment between the effects observed in the original research and our replication results, and the exploratory analyses, we believe order had little to no impact on our findings.

Our replication and extension takes the first step in combining the key relevant experimental paradigm of Tsang [1] and Watkins *et al.* [6] about the influence of helper intent and expectation for reciprocity on gratitude and indebtedness, with a conceptual replication of Bartlett & DeSteno [2] and Peng *et al.* [16] about the relationship between gratitude, indebtedness and reciprocity. The strong association between helper intentions and expectations for reciprocity in our findings may serve as the empirical foundation to develop a comprehensive framework for explaining the influence of helper intentions on gratitude and indebtedness. By contrast, the absence of an association between expectations for reciprocity and reciprocity inclination in our results may be seen as lending support for the findings by Peng *et al.* [16] over that of Bartlett & DeSteno [2].

Following a successful replication of Tsang [1] and our extension of studying reciprocity, we recommend more regular replications in the field, to state theoretical factors and predictions that might impact the effects and future replications and to examine moderators like the amount of money or degree of favour involved. Together with the empirical support for the impact of helper intentions on gratitude and indebtedness found in this replication, we believe it would be ideal to conduct a comprehensive systematic review, two decades after the research by Tsang [1] was first conducted.

Lastly, although we found an association between reciprocity expectations and gratitude and indebtedness in these studies, the causal effect between these variables remains unclear. Future research could consider trying to manipulate perceived expectations for reciprocity and reciprocity inclination, to further test the causal relationship between helper intent, reciprocity, gratitude and indebtedness.

Ethics. The project was part of a large mass replications and extensions project, which received ethics approval from the University of Hong Kong (no. EA210265).

Data accessibility. We provided all materials, data, and code on [48].

Supplementary material is available online [49].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. C.F.C.: conceptualization, formal analysis, investigation, methodology, resources, software, validation, visualization, writing—original draft, writing—review and editing; H.C.L.: formal analysis, methodology, writing—original draft; F.Y.L.: formal analysis, methodology, writing—original draft; W.I.: formal analysis, methodology, writing—original draft; C.F.S.L.: formal analysis, methodology, writing—original draft; K.Y.Y.T.: supervision; G.F.: conceptualization, data curation, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. The project was supported by the University of Hong Kong Teaching Development Grant awarded to G.F.

References

1. Tsang JA. 2006 The effects of helper intention on gratitude and indebtedness. *Motiv. Emot.* **30**, 198–204. (doi:10.1007/s11031-006-9031-z)
2. Bartlett MY, DeSteno D. 2006 Gratitude and prosocial behavior: helping when it costs you. *Psychol. Sci.* **17**, 319–325. (doi:10.1111/j.1467-9280.2006.01705.x)
3. Gray SA, Emmons RA, Morrison A. 2001 Distinguishing gratitude from indebtedness in affect and action tendencies. In *Poster presented at the annual meeting of the American Psychological Association*. San Francisco, CA.
4. Peng C, Nelissen RMA, Zeelenberg M. 2018 Reconsidering the roles of gratitude and indebtedness in social exchange. *Cogn. Emot.* **32**, 760–772. (doi:10.1080/02699931.2017.1353484)
5. Greenberg MS. 1980 A theory of indebtedness. In *Social exchange: advances in theory and research* (eds KJ Gergen, MS Greenberg, RH Willis), pp. 3–26. New York, NY: Plenum Press.
6. Watkins P, Scheer J, Ovnick M, Kolts R. 2006 The debt of gratitude: dissociating gratitude and indebtedness. *Cogn. Emot.* **20**, 217–241. (doi:10.1080/02699930500172291)
7. Ortony A, Clore GL, Collins A. 1988 *The cognitive structure of emotions*. Cambridge, UK: Cambridge University Press.
8. McCullough ME, Tsang JA, Emmons RA. 2004 Gratitude in intermediate affective terrain: links of grateful moods to individual differences and daily emotional experience. *J. Personal. Soc. Psychol.* **86**, 295–309. (doi:10.1037/0022-3514.86.2.295)
9. Bar-Tal D, Bar-Zohar Y, Greenberg MS, Hermon M. 1977 Reciprocity behavior in the relationship between donor and recipient and between harm-doer and victim. *Sociometry* **40**, 293. (doi:10.2307/3033537)
10. Graham S. 1988 Children's developing understanding of the motivational role of affect: an attributional analysis. *Cogn. Dev.* **3**, 71–88. (doi:10.1016/0885-2014(88)90031-7)
11. Lane J, Anderson NH. 1976 Integration of intention and outcome in moral judgment. *Mem. Cogn.* **4**, 1–5. (doi:10.3758/bf03213247)
12. Tesser A, Gatewood R, Driver M. 1968 Some determinants of gratitude. *J. Personal. Soc. Psychol.* **9**, 233–236. (doi:10.1037/h0025905)
13. McCullough ME. 2002 Savoring life, past and present: explaining what hope and gratitude share in common. *Psychol. Inq.* **13**, 302–304. <https://psycnet.apa.org/record/2003-01827-007>
14. Gouldner AW. 1960 The norm of reciprocity: a preliminary statement. *Am. Sociol. Rev.* **25**, 161. (doi:10.2307/2092623)
15. Mathews MA, Green JD. 2010 Looking at me, appreciating you: self-focused attention distinguishes between gratitude and indebtedness. *Cogn. Emot.* **24**, 710–718. (doi:10.1080/02699930802650796)
16. Peng C, Malafosse C, Nelissen RMA, Zeelenberg M. 2020 Gratitude, indebtedness, and reciprocity: an extended replication of Bartlett & DeSteno (2006). *Soc. Influ.* **15**, 4–16. (doi:10.1080/15534510.2019.1710248)

17. Algoe SB, Gable SL, Maisel NC. 2010 It's the little things: everyday gratitude as a booster shot for romantic relationships. *Pers. Relationships* **17**, 217–233. (doi:10.1111/j.1475-6811.2010.01273.x)
18. Frijda NH. 1988 The laws of emotion. *Am. Psychol.* **43**, 349–358. (doi:10.1037/0003-066x.43.5.349)
19. Fredrickson BL. 2004 Gratitude, like other positive emotions, broadens and builds. In *The psychology of gratitude* (eds RA Emmons, ME McCullough), pp. 145–166. Oxford, UK: Oxford University Press. (doi:10.1093/acprof:oso/9780195150100.003.0008)
20. Emmons RA, Crumpler CA. 2000 Gratitude as a human strength: appraising the evidence. *J. Soc. Clin. Psychol.* **19**, 56–69. (doi:10.1521/jscp.2000.19.1.56)
21. Algoe SB, Fredrickson BL, Gable SL. 2013 The social functions of the emotion of gratitude via expression. *Emotion* **13**, 605–609. (doi:10.1037/a0032701)
22. Bartlett MY, Condon P, Cruz J, Baumann J, Desteno D. 2012 Gratitude: prompting behaviours that build relationships. *Cogn. Emot.* **26**, 2–13. (doi:10.1080/02699931.2011.561297)
23. Kubacka KE, Finkenauer C, Rusbult CE, Keijsers L. 2011 Maintaining close relationships: gratitude as a motivator and a detector of maintenance behavior. *Personal. Soc. Psychol. Bull.* **37**, 1362–1375. (doi:10.1177/0146167211412196)
24. Williams LA, Bartlett MY. 2015 Warm thanks: gratitude expression facilitates social affiliation in new relationships via perceived warmth. *Emotion* **15**, 1–5. (doi:10.1037/emo0000017)
25. Goyal N, Adams MM, Wice M, Sullivan S, Miller JG. 2022 Gratitude endures while indebtedness persuades: investigating the unique influences of gratitude and indebtedness in helping. *Cogn. Emot.* **36**, 1361–1373. (doi:10.1080/02699931.2022.2118237)
26. Naito T, Sakata Y. 2010 Gratitude, indebtedness, and regret on receiving a friend's favor in Japan. *Psychologia* **53**, 179–194. (doi:10.2117/psysoc.2010.179)
27. Ames DR, Flynn FJ, Weber EU. 2004 It's the thought that counts: on perceiving how helpers decide to lend a hand. *Personal. Soc. Psychol. Bull.* **30**, 461–474. (doi:10.1177/0146167203261890)
28. Welsh DT, Baer MD, Kim JK, Thiel C, Smith IH. 2022 Desirable or deceitful? How social exchange dynamics shape responses to pro-coworker unethical behavior. *Pers. Psychol.* **75**, 619–644. (doi:10.1111/peps.12483)
29. Nosek BA *et al.* 2022 Replicability, robustness, and reproducibility in psychological science. *Annu. Rev. Psychol.* **73**, 719–748. (doi:10.1146/annurev-psych-020821-114157)
30. Zwaan RA, Etz A, Lucas RE, Donnellan MB. 2018 Making replication mainstream. *Behav. Brain Sci.* **41**, e120. (doi:10.1017/s0140525x17001972)
31. Komter AE. 2004 Gratitude and gift exchange. In *The psychology of gratitude* (eds RA Emmons, ME McCullough), pp. 194–212. Oxford, UK: Oxford University Press. (doi:10.1093/acprof:oso/9780195150100.003.0010)
32. Chen Z. 2024 Grateful or indebted? revisiting the role of helper intention in gratitude and indebtedness. *Peer Community Registered Report* (doi:10.24072/pci.r.100373)
33. Chen Z. 2025 Grateful or indebted? revisiting the role of helper intention in gratitude and indebtedness. *Peer Community Registered Report* (doi:10.24072/pci.r.100373)
34. Feldman G. 2023 *Registered report stage 1 manuscript template*. See <https://doi.org/10.17605/OSF.IO/YQXTP>.
35. R Core Team. 2022 *R: a language and environment for statistical computing (version 4.1)*. R packages retrieved from CRAN snapshot 2023-04-07. See <https://cran.r-project.org>.
36. Jané M *et al.* 2024 *Guide to effect sizes and confidence intervals*. (doi:10.17605/OSF.IO/D8C4G)
37. Faul F, Erdfelder E, Lang AG, Buchner A. 2007 G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191. (doi:10.3758/bf03193146)
38. Simonsohn U. 2015 Small telescopes: detectability and the evaluation of replication results. *Psychol. Sci.* **26**, 559–569. <https://doi.org/10.1177/0956797614567341>
39. Chen EY, Chee MX, Feldman G. 2023 Revisiting the differential centrality of experiential and material purchases to the self: replication and extension of Carter and Gilovich (2012). *Collabra* **9**, 57785. (doi:10.1525/collabra.57785)
40. Yeung SK, Feldman G. 2022 Revisiting the temporal pattern of regret in action versus inaction: replication of Gilovich and Medvec (1994) with extensions examining responsibility. *Collabra Psychol.* **8**, 37122. (doi:10.1525/collabra.37122)
41. LeBel EP, Vanpaemel W, Cheung I, Campbell L. 2019 A brief guide to evaluate replications. *MP.* **3**, 1–9. (doi:10.15626/MP.2018.843)
42. LeBel EP, McCarthy RJ, Earp BD, Elson M, Vanpaemel W. 2018 A unified framework to quantify the credibility of scientific findings. *Adv. Methods Pract. Psychol. Sci.* **1**, 389–402. (doi:10.1177/2515245918787489)
43. Nosek BA, Lakens D. 2014 Registered reports: a method to increase the credibility of published results. *Soc. Psychol. (Göttingen, Germany)* **45**, 137–141. <https://doi.org/10.1027/1864-9335/a000192>
44. Patil I. 2021 Visualizations with statistical details: the 'ggstatsplot' approach. *J. Open Source Softw.* **6**, 3167. (doi:10.21105/joss.03167)
45. The jamovi Project. 2023 *jamovi. (version 2.6.17)*. See <https://www.jamovi.org>.
46. Chen J, Kwan LC, Ma LY, Choi HY, Lo YC, Au SY, Tsang CH, Cheng BL, Feldman G. 2021 Retrospective and prospective hindsight bias: replications and extensions of Fischhoff (1975) and Slovic and Fischhoff (1977). *J. Exp. Soc. Psychol.* **96**, 104154. (doi:10.1016/j.jesp.2021.104154)
47. Feldman G. 2023 *'Will it replicate? Tsang, 2006 | studies 2 and 3' poll on Twitter (Mar 26, 2023)*. See <https://x.com/giladfeldman/status/1639889119013752832>.
48. Chan CF, Ching LH, Lau FY, Ip W, Lui S, Tam KYY, Feldman G. 2025 Revisiting the effects of helper intention on gratitude and indebtedness: replication and extensions registered report of Tsang (2006) (doi:10.17605/OSF.IO/GHFY4)

49. Chan CF, Lim HC, Lau FY, Ip W, Lui CFS, Tam KYY *et al.* 2025. Supplementary Material Fro: Revisiting the Effects of Helper Intentions on Gratitude and Indebtedness: Replication and Extensions Registered Report of Tsang (2006). FigShare (doi:[10.6084/m9.figshare.c.7790589](https://doi.org/10.6084/m9.figshare.c.7790589))

Revisiting the Effects of Helper Intention on Gratitude and Indebtedness: Replication and extensions Registered Report of Tsang (2006) **Supplementary**

Analysis of the original article	2
Original article methods	2
Type of study	2
Experimental design	2
Manipulation checks	5
Original article results	6
Sample size before and after exclusions	6
Included sample description	6
Study 2	6
Study 3	6
Findings in the original article	6
Study 2	7
Study 3	7
Effect size calculations of the original study effects	9
Power analysis of original study effect to assess required sample for replication	10
Sensitivity analyses	12
Handling outliers: Strategy	12
Materials and scales used in the replication + extension experiment	13
Experiment	13
<i>Study 2</i>	13
<i>Study 3</i>	14
Funnelling section	15
Comprehension Check	16
Scales used in the experiments	16
Comparisons and deviations	18
Original versus replication	18
Pre-registration plan versus final report	19
Additional analyses and results	20
Additional information about the study	20
Replication evaluation	21
Replication closeness	21
Replication versus the original	22
References	23

Analysis of the original article

Original article methods

Type of study

True randomized experimental studies are conducted in the original article.

Experimental design

Study 2

99 undergraduates were recruited to participate in this study, and they were run singly in laboratory cubicles. Participants were first asked to remember and write about a situation that happened to them in the past year. Then, participants were randomly assigned into 2 conditions, namely the Benevolent condition ($N=45$) and Selfish Condition ($N=47$). They are asked to recall a situation in which someone has done something good for benevolent and selfish reasons. After that, they are asked to take a minute to think back and re-experience the thoughts and emotions that they were feeling during the situation, and then instructed to write about the details of the situation on a lined paper.

After that, participants were then asked questions about their current emotions in reaction to the situation they were asked to recall. Participants responded on a 1-7 Likert-type scale (1= Would feel very little of this emotion, 7 = Would feel a lot of this emotion) A gratitude scale was constructed from the emotion adjectives *grateful*, *thankful*, and *appreciative*, $\alpha = .96$, and an indebtedness scale was constructed from the emotion adjectives *indebted* and *obligated*, $\alpha = .71$. Participants were also asked to rate the helper's motivation in the situation (1 = Very concerned about me, 7 = Motivated mostly by selfish reasons) and the magnitude of the favor ("How big of a favor do you think the other person did for you?" 1 = A very small favor, 7 = A very big favor). In a nutshell, the experimental design is summarized in **Table 1** below.

This was a between-subject experimental design, in which the benefactors' intention is the only manipulated independent variable in the study. Participants were randomly assigned into 2 different groups(i.e., conditions), and all the dependent variables were measured. The magnitude of the favor was added as a covariate.

Table S1*Experimental Design for Study 2*

Independent variables (IV)		
IV: Motives of helper in recalling conditions		
<u>Condition 1</u> : Benevolent conditions -		
Asked to recall a situation that someone has done something good for a benevolent reason		
<u>Condition 2</u> : Selfish conditions –		
Asked to recall a situation that someone has done something good for a selfish reason		
Dependent variables (DV)		
DV1:	DV2:	DV3:
Gratitude	Indebtedness	Perceived helpers' motivations
Covariate (C)		
C:		
Magnitude of the favor		

Data collected were analyzed with three kinds of statistical tests, including correlation tests, ANCOVAs, and regression analyses. First, three correlation tests were conducted to study the correlation between gratitude and indebtedness in different helper intention conditions (i.e., Selfish, Benevolent and across conditions). Then, two ANCOVAs, with the perceived magnitude of favor added as a covariate, were conducted to investigate the effects of perceived helper conditions on gratitude and indebtedness respectively. Lastly, two regression analyses were conducted with conditions and magnitude of favor predicting gratitude and indebtedness.

Designs for the statistical tests in this study are summarized in the following Tables 2-4:

Table**S2***Correlation tests in Study 2*

Tests	Conditions	Variables	
		1	2
1	Selfish	Gratitude	Indebtedness
2	Benevolent	Gratitude	Indebtedness
3	Selfish + Benevolent	Gratitude	Indebtedness

Table**S3***ANCOVAs in Study 2*

Tests	Variables		
	IV	DV	Covariate
1	Motives of helper	Gratitude	Magnitude of favor
2	Motives of helper	Indebtedness	Magnitude of favor

Table**S4***Regression analyses in Study 2*

Tests	Variables		
	IV1	IV2	DV
1	Motives of helper	Magnitude of favor	Gratitude
2	Motives of helper	Magnitude of favor	Indebtedness

Study 3

86 undergraduates were recruited to participate in this experiment, and they were run in groups of 2-30 individuals. Participants are randomly assigned to read one of the three scenarios, including the Benevolent Motives (N = 28), Ulterior Motives (N = 29), and Ambiguous Motives (N=29) scenarios. The scenario is as follows:

“It’s the beginning of the semester, and you’re standing in line at the bookstore to buy all the books for your classes. You are waiting in line with a friend, and the both of you joke about how long the line is taking. After a long wait, the cashier rings you up, and you find out that the total cost for your books is \$400, which is much more expensive than what you expected. You only have \$200 in your checking account. As you are standing there wondering what to do, your friend offers to pay the extra \$200 for you: “Don’t worry about it. I’ve been in that situation before and it’s a real bummer! Let me pay for it and you won’t have to stress about getting your books in time for the first exam or anything.” Although your friend isn’t rich, you know that your friend can afford the \$200.”

However, the last sentence of the paragraph would be different according to the assigned benefactors’ intention.

- **Benevolent Motives:** “You can tell that your friend is really concerned about you and wants to help you out, so you say yes.”
- **Ulterior Motives:** “You know that your friend is really doing you this favor in order to borrow your car next weekend, but you really need those textbooks, so you say yes.”
- **Ambiguous Motives:** “You really need those textbooks, so you say yes. The next weekend that same friend asks you if they can borrow your car to run some errands.”

The experimental design is mostly the same as Study 1 in the original article, with “Ambiguous Motives” added as an extra condition apart from Benevolent and Selfish conditions. Details regarding the experimental design are shown in Table 5. Measures of gratitude, $\alpha = .85$, indebtedness, $\alpha = .64$, and perceived benefactor intention are identical in Study 2.

This is a between-subject experimental design with 3 conditions, in which the benefactors’ motives provided in the paragraph are the only manipulated independent variable in the study. Participants are randomly assigned into different groups. All the dependent variables are measured for each subject.

Table

Experimental Design for Study 3

S5

Independent variables (IV)			
IV: Motives of helper in the provided in the paragraph			
<u>Condition 1</u> : Benevolent conditions			
<u>Condition 2</u> : Selfish conditions			
<u>Condition 3</u> : Ambiguous conditions			
Dependent variables (DV)			
DV1:	DV2:	DV3:	DV4:
Gratitude	Indebtedness	Perceived helpers’ motivations	Magnitude of the favour

Manipulation checks

Study 2

The author manipulates the recalling situations of perceived helpers’ motives. The perceived helpers’ motives between the Benevolent and Selfish conditions are tested with independent-sample t-tests. Participants in the Benevolent condition are expected to perceive their helper’s motives as being less selfish than participants in the Selfish condition. Also, following the original article, the magnitude of the favor between the Benevolent condition and the Selfish condition would also be checked.

Study 3

The author manipulates the helpers’ motives across situations. The perceived helpers’ motives across the 3 conditions (Benevolent, Ambiguous, and Utterior condition) are tested with independent-sample t-tests. Participants are expected to rate the helper as less selfish in the Benevolent Motives condition compared to the Utterior Motives condition.

Original article results

Sample size before and after exclusions

Study 2

97 participants before exclusions and 92 participants after exclusions are reported. 5 participants (3 from selfish conditions and 2 from benevolent conditions) were excluded from data analyses as they didn't follow instructions about the type of situation to recall. Eventually, 45 participants are assigned to Benevolent condition while 47 participants are assigned to Selfish Condition.

The degree of freedom in ANCOVA ($df1=1$; $df2=89$) and regression analyses ($df1=1$; $df2=89$) were consistent with the number of participants and conditions. However, it is worth noticing that the degree of freedom for the correlational tests for the benevolent condition ($df=42$) and selfish condition ($df=46$) is not consistent with the number of participants reported. The degrees of freedom for a correlation in $n-2$. Thus, 1 more participant and 1 less participant are analyzed in the Benevolent condition and Selfish Condition respectively.

Study 3

86 participants were recruited to join the study. No known participants were excluded from the study. Participants are randomly assigned into Benevolent ($N=28$), Ulterior ($N=29$) and Ambiguous scenario ($N=28$).

In this study, the degree of freedom 2 for ANOVAs ($df2=83$) was consistent with the number of participants ($N=86$). However, the degree of freedom for the correlational tests was not reported.

Included sample description

Study 2

- Age (unreported)
- Gender (male = 16, female = 76)
- Location (run singly in laboratory cubicles at Baylor University)
- Sample type (undergraduates at Baylor University, questionnaire packet was given)

Study 3

- Age (unreported)
- Gender (male = 13, female = 49, unreported = 24)
- Location (run in groups of 20-30 individuals at Baylor University)
- Sample type (undergraduates at Baylor University, questionnaire packet was given.)

Findings in the original article

The following finding is directly extracted from the original article. Please refer to the original article for more details. Please note that all the following p-values for t-tests are two-tailed according to the author. A few comments would be given for some of the suspicious results. **Table 6** shows the mean levels of gratitude and indebtedness by a condition in three studies as shown in the original article.

Study 2

Part 1: Correlational tests

- Across conditions, participants expressed a high amount of gratitude ($M = 5.01$, $SD = 2.04$), and lower levels of indebtedness ($M = 2.78$, $SD = 1.50$).
- Gratitude and indebtedness were significantly correlated between conditions $r(90) = .57$, $p < .001$, and within the Selfish, $r(46) = .61$, $p < .001$, but not the Benevolent condition, $r(42) = .20$, $p = .20$.
- ***Pearson's r being 20 is statistically unrealistic so it is suspected to be a typo error in the original article. The R could possibly be 0.2 for the benevolent condition.

Part 2: ANCOVAs

- Participants reported significantly more grateful emotions after writing about a favor given unselfishly compared to one given selfishly after controlling for favor magnitude, $F(1,89) = 21.61$, $p < .001$.
- There was no significant effect of intention on feelings of indebtedness after controlling for favor magnitude, $F(1,89) = 1.21$, $p > .20$.

Part 3: Regression analyses

- Selfish/unselfish condition [$\beta = .32$, $t(89) = 4.65$, $p < .01$] and magnitude of the favor [$\beta = .62$, $t(89) = 9.02$, $p < .01$] both significantly and uniquely predicted feelings of gratitude, $R^2 = .73$, $F(2,89) = 114.40$, $p < .001$.
- Selfish/unselfish condition [$\beta = .13$, $t(89) = 1.10$, $p > .20$] did not significantly predict indebtedness, although the magnitude of the favor [$\beta = .42$, $t(89) = 3.69$, $p < .001$] did, $R^2 = .26$, $F(2, 89) = 15.22$, $p < .001$.

Study 3

Part 1: Correlational tests

- Across the three conditions (Benevolent, Ambiguous and Ulterior), participants expressed both a high amount of gratitude ($M = 6.45$, $SD = .71$) and indebtedness ($M = 5.30$, $SD = 1.35$).
- Gratitude and indebtedness were significantly correlated only in the Ambiguous Motives condition ($r = .42$, $p < .05$);
- Correlations between Gratitude and indebtedness in the Benevolent Motives ($r = -.13$, $p > .20$) and Ulterior Motives condition ($r = .26$, $p = .17$) were not significant.

Part 2: ANOVAs and independent-sample t-tests

- ANOVAs suggested that helper intention conditions had a significant effect on gratitude [$F(2,83) = 6.72$, $p < .01$], but not indebtedness [$F(2,83) = .18$, $p > .20$].
- Participants felt significantly more grateful in the Benevolent Motives condition compared to the Ambiguous Motives condition [$t(41.63) = 2.07$, $p < .05$], and marginally less grateful in the Ulterior Motives condition compared to the Ambiguous Motives condition [$t(56) = -1.84$, $p = .07$].
- In contrast, levels of indebtedness in the Benevolent Motives condition were not significantly different from the Ambiguous Motives [$t(55) = .47$, $p > .20$] or the Ulterior Motives condition [$t(56) = -.10$, $p > .20$].
- Looking at the Ambiguous Motives condition alone, ratings of selfish intentions were significantly negatively correlated with gratitude, $r(27) = -.40$, $p < .05$, but uncorrelated with indebtedness, $r(27) = .00$, $p > .20$.

Table S6

Mean levels of gratitude and indebtedness by conditions in Study 2 and 3 from the original article, PDF file p. 3

Study	Intention condition		
	benevolent	Selfish	Ambiguous
Study 2			
Gratitude	6.48 (.63)	3.67 (1.96)	
Indebtedness	3.36 (1.42)	2.24 (1.38)	
Study 3			
Gratitude	6.76 (.31)	6.49 (.62)	6.11 (.92)
Indebtedness	5.43 (1.35)	5.26 (1.37)	5.22 (1.38)

Note. Values enclosed in parentheses represent standard deviations

Effect size calculations of the original study effects

The effect size and confidence intervals of the original study are summarized in Table 1 in the main manuscript, while the calculations for the manipulation checks in the original study effects are summarized in **Table S7**, followed by the code blocks for R (Version 1.4.1717) computations and their output correspondingly (Figures 1- 13). Packages of "MOTE", "MBESS" and "esc" are adopted in the calculation in R.

For Rmarkdown code for calculating the target's effects and the power analysis, please see files in the OSF folder: `Effect_size_and_power_analysis_for_original_studies*.RMD/html`

Table S7

Summary of Effect Size and Confidence Interval of the manipulation checks.

Manipulation checks	Factor	<i>p</i>	Effect size	95%CI
1	The perceived helpers' motives in Benevolent condition being less selfish compared to Selfish condition	<.001	<i>d</i> = 2.576	[2.023, 3.129]
2	The rated magnitude of favor being larger in Benevolent condition being less selfish compared to Selfish condition.	<.001	<i>d</i> = 1.277	[0.901, 1.653]
3	The perceived helpers' motives in Ulterior Motives being more selfish compared to Benevolent Motives.	<.01	<i>d</i> = 1.714	[1.107, 2.321]
4	The perceived helpers' motives in Ulterior Motives being more selfish compared to Ambiguous Motives.	<.01	<i>d</i> = 1.205	[0.646, 1.765]
5	The rated magnitude of favor being smaller in Ulterior Motives compared to Benevolent Motives.	<.01	<i>d</i> = 0.765	[0.342, 1.189]
6	The rated magnitude of favor being smaller in Ulterior Motives compared to Ambiguous Motives.	<.05	<i>d</i> = 0.428	[0.014, 0.841]

Note. All calculations are corrected to 3 decimal places if possible.

Effect = Cohen's *d*. 95%CI = 95% Confidence Interval.

*Checks with significant results, i.e., $p < 0.05$

Power analysis of original study effect to assess required sample for replication

For Rmarkdown code for calculating the target's effects and the power analysis, please see files in the OSF folder: `Effect_size_and_power_analysis_for_original_studies[version].RMD/html`

The *Priori Power Analysis* in the Original Study is summarized in Table S8. We used R studio and G-power to conduct the following analyses. The required R code blocks, the R packages for R computations as well as their output are shown in the Figure 23 to 29. R package "pwr" is adopted in the analysis.

The required sample size for 0.95 power and 0.05 alpha is **264**. This is calculated by rounding up the highest minimum sample size required for testing the hypothesis, i.e., factor 7b. Since the calculated sample size is the number of each group for an independent sample t-test, the calculated number of each condition needs to be times 3. The calculation is structured in the following way:

88 samples for each group x 3 groups = 264.

Power Analyses were not conducted for the hypotheses which had insignificant results in the target article (i.e., $p > .05$). These factors are shown with N.A. in the "minimum required N". Otherwise, abnormally high required sample size would be calculated from the Power Analyses. Also, these hypotheses are not hypotheses of main interest in this research.

Table S8
Summary of Priori Power Analysis in the Original Study

Factor	Description	Statistical Test Conducted in the Original Study	Minimum Required N
1a	Gratitude being associated with indebtedness across conditions (Selfish and Benevolent).	Pearson correlation	34
1b	Gratitude being associated with indebtedness in Selfish condition.	Pearson correlation	29
1c	Gratitude being associated with indebtedness in Benevolent condition.	Pearson correlation	N.A.
2	More gratitude emotions being experienced when recalling a selfish favor compared to an unselfish favor after controlling magnitude of favor	ANCOVA	36*
3	More Indebtedness emotions being experienced when recalling a selfish favor compared to an unselfish favor after controlling magnitude of favor	ANCOVA	N.A.
4	Helper's motives (selfless/ ulterior) and magnitude of favor being able to predict feelings of gratitude	Regression Analysis	N.A.

5	Helper's motives (selfless/ ulterior) and magnitude of favor being able to predict feelings of indebtedness	Regression Analysis	<i>N.A.</i>
6	Gratitude being associated with indebtedness in Ambiguous condition.	Pearson correlation	67
7a	Level of gratitude being different across three conditions (Benevolent, Ulterior and Ambiguous).	ANOVA	99 ($n_1=33$, $n_2=33$, $n_3=33$)
7b	Level of gratitude being higher in Benevolent conditions compared to Ambiguous conditions.	Independent samples t-test (two-tailed)	176** ($n_1=88$, $n_2=88$)
7c	Level of gratitude being higher in Ambiguous conditions compared to Ulterior conditions.	Independent samples t-test (two-tailed)	<i>N.A.</i>
8a	Level of indebtedness being different between three conditions (Benevolent, Ulterior and Ambiguous).	ANOVA	<i>N.A.</i>
8b	Level of indebtedness being higher in Benevolent conditions compared to Ulterior conditions.	Independent samples t-test (two-tailed)	<i>N.A.</i>
8c	Level of indebtedness being higher in Benevolent conditions compared to Ambiguous conditions.	Independent samples t-test (two-tailed)	<i>N.A.</i>
9	Rating of selfish intentions being associated with gratitude in Ambiguous conditions.	Pearson correlation	75
10	Rating of selfish intentions being no associated with indebtedness in Ambiguous conditions.	Pearson correlation	<i>N.A.</i>

Note. α (two-sided) = .05, power = .95. N = sample size; n_m = sample size of condition m.

Factors with *N.A.* are not of main interest in this study

* The power analysis for factor 2 was specifically calculated with G power since the R packages used don't support the power analysis for ANCOVA

**Factor 7b owns the highest minimum sample size required for the testing the hypothesis

Sensitivity analyses

With $N = 750$, 95% power, $\alpha = 5\%$, one-tail.

Two conditions independent samples t-test, $n = 375$:

t tests - Means: Difference between two independent means (two groups)

Analysis: Sensitivity: Compute required effect size

Input:	Tail(s)	=	One	
	α err prob=		0.05	
	Power ($1-\beta$ err prob)	=	0.95	
	Sample size group 1	=	375	
	Sample size group 2	=	375	
Output:	Noncentrality parameter δ	=	3.2926870	
	Critical t	=	1.6468933	
	Df	=	748	
	Effect size d	=	0.2404639	

Three conditions between-subject one-way factor ANOVA, $N = 750$:

F tests - ANOVA: Fixed effects, omnibus, one-way

Analysis: Sensitivity: Compute required effect size

Input:	α err prob=		0.05	
	Power ($1-\beta$ err prob)	=	0.95	
	Total sample size	=	750	
	Number of groups	=	3	
Output:	Noncentrality parameter λ	=	15.5053645	
	Critical F	=	3.0077784	
	Numerator df	=	2	
	Denominator df	=	747	
	Effect size f	=	0.1437839	

Two conditions out of three conditions, independent samples t-test, $n = 250$:

t tests - Means: Difference between two independent means (two groups)

Analysis: Sensitivity: Compute required effect size

Input:	Tail(s)	=	One	
	α err prob=		0.05	
	Power ($1-\beta$ err prob)	=	0.95	
	Sample size group 1	=	250	
	Sample size group 2	=	250	
Output:	Noncentrality parameter δ	=	3.2941867	
	Critical t	=	1.6479191	
	Df	=	498	
	Effect size d	=	0.2946410	

Handling outliers: Strategy

We included all data, we did not classify outliers.

Materials and scales used in the replication + extension experiment

We are grateful to Professor Tsang, the author of the original article, for providing us with the target article's materials.

For the most up to date survey details, see Qualtrics export files: RRR_Tsang_2006_Studies_2-3.QSF/DOCX/PDF in the OSF folder.

Experiment

The order of the following studies are randomized. The following instructions are the direct replicates from the original materials in hopes of accurate replication of the original studies.

Study 2

Instructions:

“Task of recall and re-experience

For this study, you will be asked to remember a situation that happened to you in the past year. Try and remember this past situation as clearly as possible:

Please try and recall a past experience where you felt that:

- 1. Someone else had caused, and was controlling, what was happening in the situation,*
- 2. The other person was doing something good for you for (X)* reasons, and*
- 3. The positive consequences of this other person's actions were important to you.*

Take a minute to think back and re-experience the thoughts and emotions that you were feeling during this experience. In particular, think about the things that this person did that made you think he/she was acting unselfishly.”

The (X)* seen by subjects would be different according to their randomly allocated conditions. For benevolent conditions, subjects would see the word “*unselfish*”. whereas for selfish conditions, subjects would see the word “*selfish*”. Subjects also need to answer a few comprehension check questions before proceeding to the next page. Please see section “Comprehension check” for more details.

After answering the comprehension check questions and moving to the next page, subjects were asked to describe the past situation they were recalling briefly. Questions include:

- 1. Please briefly describe this past situation: What happened? (1-3 sentences)*
- 2. What actions did this person do to benefit you? (1-2 sentences)*
- 3. How large were the positive consequences that came from this person's actions? (1-2 sentences)*

After that, subjects are required to answer the following questions in randomized order:

- 1. Please rate the other person's motivations on the following scale.
1 = very concerned about me, 7 = motivated by selfish reasons*
- 2. Please rate your understanding of the other person's expectations of you to reciprocate:
1 = No expectations to reciprocate, 7 = Very high expectations to reciprocate*
- 3. People may feel different emotions in response to various situations. Please choose the number by each adjective to indicate the extent to which you feel each of the following emotions right now in reaction to thinking about the past situation:*
 - a. Grateful*

- b. *Thankful*
- c. *Appreciative*
- d. *Indebted*
- e. *Obligated*
- f. *Calm*
- g. *Pleased*
- h. *Resentful*
- i. *Upset*
- j. *Annoyed*

** the sequence of emotions shown is randomized.*

4. *What do you think were the other person's specific motivations for doing something good for you?*
5. *How costly (in terms of money, time, effort, etc.) do you think this situation was for the person who did something good for you?*
6. *How big of a favor do you think the other person did for you?*
7. *How costly (in terms of money, time, effort, etc.) do you think this situation was for the person who did something good for you?*
1 = Cost very little, 7 = Cost very much

Study 3

Instructions:

"Situational questions"

The following page contains a short scenario for you to read. In this scenario, a friend does you a favor. Pay particular attention to the motivations that the friend might have in doing you a favor (did your friend have good intentions, or was your friend being selfish?). After reading the scenario, you will be asked questions about what you think the friend's motivations were."

Then subjects would proceed to the next page and read the following paragraph.

"Try to imagine yourself in the following situation. Imagine what you would think and feel if the following events happened to you:

'It's the beginning of the semester, and you're standing in line at the bookstore to buy all the books for your classes. You are waiting in line with a friend, and both of you joke about how long the line is taking. After a long wait, the cashier rings you up, and you find out that the total cost for your books is \$400, which is much more expensive than what you expected. You only have \$200 in your checking account. As you are standing there wondering what to do, your friend offers to pay the extra \$200 for you: "Don't worry about it. I've been in that situation before and it's a real bummer! Let me pay for it and you won't have to stress about getting your books in time for the first exam or anything." Although your friend isn't rich, you know that your friend can afford the \$200.'"

The last sentence of the paragraph would be different according to the assigned conditions, i.e., benefactor intention.

- Benevolent Motives: *"You can tell that your friend is really concerned about you and wants to help you out, so you say yes."*

- Ulterior Motives: “You know that your friend is really doing you this favor in order to borrow your car next weekend, but you really need those textbooks, so you say yes.”
- Ambiguous Motives: “You really need those textbooks, so you say yes. The next weekend that same friend asks you if they can borrow your car to run some errands.”

After reading the paragraphs, subjects could move to the next page and answer the questions that followed. The order of questions shown to the subjects is in random order.

1. What do you think were the friend's motivations for offering to pay for the textbooks? (1-2 sentences)
2. Please rate the other person's motivations on the following scale: 1 = very concerned about me, 7 = motivated by selfish reasons
3. People may feel different emotions in various situations. Please circle the number by each adjective to indicate the extent to which you would feel each of these emotions in the scenario you just read:
 - a. Grateful
 - b. Thankful
 - c. Appreciative
 - d. Indebted
 - e. Obligated
 - f. Calm
 - g. Pleased
 - h. Resentful
 - i. Upset
 - j. Annoyed
4. How much of a favor do you think the friend did by giving money for the textbooks?
1 = A very small favor, 7 = A very big favor
5. Have you ever had a friend loan you money for textbooks?
Yes or No
- 6.
7. To what extent do you have the urge to do the following ways?
 - a. I would feel like helping my friend in return
 - b. I would feel like giving my friend a gift in return
 - c. I would feel like doing something for my friend in return

Funnelling section

Three funnelling questions:

- What do you think the purpose of the last part was?
- Have you ever seen the materials used in this study or similar before? If yes - please indicate where
- Did you spot any errors? Anything missing or wrong? Something we should pay attention to in the next runs? (Briefly, up to one sentence, write "none" if not relevant)

Finally, participants were asked to fill in demographics and were debriefed. No filler items were included.

Comprehension Check

Study 2

We added two comprehension-check questions for Study 2. The two questions were designed based on the instructions. Participants were not allowed to proceed to the next page unless they answered the questions correctly. The answers were presented in randomized order.

1. What type of helping behavior are you asked to recall?
 - a. Unselfish (benevolent) help
 - b. Selfish help
 - c. Any kind of help

Ans: a. Help someone else gave to me in Benevolent Condition

b. Selfish help in Selfish Condition

2. Whose helping behavior are you asked to recall?
 - a. Help I gave to someone else
 - b. Help someone else gave to me
 - c. Any kind of help

Ans: b. Help someone else gave to me

Study 3

We added three comprehension-check questions for Study 3. One was directly extracted from the original study, with additional two new questions. Participants were not allowed to proceed to the next page unless they answered the questions correctly. The answers were presented in randomized order.

1. How much money did the friend offer to give to help pay for the textbooks?

Ans: 200

2. What was the favor offered in the scenario?
 - a. A birthday gift
 - b. Helping you with your homework
 - c. Paying for the textbooks for you

Ans: C. Paying for the textbooks for you

3. According to the text: Why is your friend offering to help you?
 - a. I know without doubt it is because my friend wanted to borrow my car
 - b. It is not clear about the two being related, but the weekend after helping me this friend asked to borrow my car
 - c. My friend is really concerned about me

Ans: a. for Selfish Condition

b for Ambiguous condition

c for Benevolent Condition

Scales used in the experiments

A 1-7 Likert scale is used to measure gratitude, indebtedness, perceived helper's motivations, the magnitude of favor, and action tendencies. For the first four scales, they are directly extracted from the original study. For the fifth and sixth scales, they were designed for the extension study.

1. **Gratitude**

“To what extent would you feel the following emotions in this scenario?”

Emotion adjectives include *grateful*, *thankful*, and *appreciative*.

(1 = Would feel very little of this emotion 7 = Would feel a lot of this emotion)

(source: McCullough et al., 2004)

2. **Indebtedness**

“To what extent would you feel the following emotions in this scenario?”

Emotion adjectives include *indebted* and *obligated*.

(1 = Would feel very little of this emotion, 7 = Would feel a lot of this emotion)

(source: Greenberg, 1980)

3. **Perceived helpers' motivations**

“Please rate the helper's intention in the scenario.”

(1 = Very concerned about me, 7 = Motivated mostly by selfish reasons)

4. **Magnitude of the favour**

“How big of a favor do you think the other person did for you?”

(1 = A very small favor, 7 = A very big favor)

5. **Perceived expectations for reciprocity (Extension)**

“Please rate your understanding of the other person's expectations of you to reciprocate”

(1 = No expectations to reciprocate, 7 = Very high expectations to reciprocate)

6. **Reciprocity tendency (Extension)**

Items include

“I would feel like helping my friend in return”, “I would feel like giving my friend a gift in return”, and “I would feel like doing something for my friend in return.”

(1 = Slight urge, 7 = Very strong urge)

(source: Watkins et al., 2006)

Comparisons and deviations

Original versus replication

	Original	Replication	Reason for change
Study design	Mode: Participants completed the studies with pen and paper in the laboratories.	Participants completed the studies with an online questionnaire.	to reach more and a wider variety of participants; to conduct the studies with lower cost and higher efficiency.
Sample characteristics	Sample size: Study 2: 92; Study 3: 86 Geographic origin: Undergraduates studying at Baylor University	$N = 759$ [US American]	To increase the generalizability of results by including a more and larger variety of participants
Procedure	Items on gratitude, indebtedness was not randomized''	Items on gratitude, indebtedness were randomized	To reduce the order effect
	Study 2 and 3 are conducted separately	Study 2 and 3 are conducted with the same participants in one setting; The order of Study 2 and 3 are random to the participants	to reduce the order effect; to avoid the influence of decline to particular studies; to find potential consistency within participants' answers (whether an answer is predictive of another answer); to allow subsequent separate analysis on participants who took Study 2 / 3 as their first study.
	1 comprehension check question in Study 3	2 comprehension check questions in Study 3	To make sure the participants read and understand the instructions and scenarios in the materials
Conditions	No Change	No Change	NA

Pre-registration plan versus final report

Table S9

Deviation documentation: Pre-registration plan versus final report

Components in your preregistration (e.g., stopping rule, analyses, hypotheses, exclusion rules)	Location of 1) preregistered decision/plan and 2) rationale for decision/plan [Location / link]	Were there deviations? What type? [no / minor / major]*	If yes - describe details of deviation(s) [brief description / location / link]	Rationale for deviation [brief description / location / link]	How might the results be different if you had/had not deviated [brief description / location / link]	Date/time of decision for deviation + stage	Any additional notes
Study design	/	/	/	/	/	/	/
Measured variables	/	/	/	/	/	/	/
Exclusion criteria	/	/	/	/	/	/	/
IV	/	/	/	/	/	/	/
DV	“Method” section - Table 3 and 4 of main manuscript; Qualtrics Study 2: DV 2,3,4 Study 3: DV 1,2,3,4, Covariate	Minor	Wordings of the questions stick to the original studies materials provided by the author.	Questions deviate from the original studies.	It would no longer be direct replication.	After receiving Stage 1 in-principle acceptance, right before data collection.	/
Data analysis	/	/	/	/	/	/	/

Note. Minor - Change probably did not affect results or interpretations; Major - Change likely affected results or interpretations.

Additional analyses and results

Additional information about the study

1. Duration of Study Sessions: 21 days
2. No. of sessions: 1
3. Time of Day: Participants were allowed to do the test at any time of their convenience.
4. Data collection dates: From 3rd March 2024 - 24th March 2024
5. Participant Recruitment: Participants were recruited using Prolific.
6. No. of participants dropped out from the study: 148

This study was conducted on Prolific with American participants. We imposed the following settings in recruiting our participants:

1. Participants were paid \$1.8 as a fixed participation reward. This amount was determined by multiplying the expected completion time (in mins.) with the minimal federal wage in the U.S. (i.e., \$0.125 per minute).
2. The expected completion time for each session was set at **10** minutes in advance.
3. We limited all workers' Task Approval Rate to be between 90% and 100%.
4. We limited each worker's number of Task approved to be between 50 and 10000.
5. We restricted workers' location to be in the U.S.
6. We restricted answering from PC and or tablet (not mobile).

Replication evaluation

Replication closeness

Lebel, McCarthy, Earp, Elson, and Vanpaemel (2018):

Target similarity		Highly similar				Highly dissimilar	
Category	Direct replication			Conceptual replication			
Design facet	Exact replication	Very close replication	Close replication	Far replication	Very far replication		
Effect/hypothesis	Same/similar	Same/similar	Same/similar	Same/similar	Same/similar		
IV construct	Same/similar	Same/similar	Same/similar	Same/similar	Different		
DV construct	Same/similar	Same/similar	Same/similar	Same/similar	Different		
IV operationalization	Same/similar	Same/similar	Same/similar	Different			
DV operationalization	Same/similar	Same/similar	Same/similar	Different			
Population (e.g. age)	Same/similar	Same/similar	Same/similar	Different			
IV stimuli	Same/similar	Same/similar	Different				
DV stimuli	Same/similar	Same/similar	Different				
Procedural details	Same/similar	Different					
Physical setting	Same/similar	Different					
Contextual variables	Different						

Figure S1. Criteria for evaluation of replications by LeBel et al. (2018).

A classification of relative methodological similarity of a replication study to an original study. “Same” (“different”) indicates the design facet in question is the same (different) compared to an original study. IV = independent variable. DV = dependent variable.

“Everything controllable” indicates design facets over which a researcher has control.

Procedural details involve minor experimental particulars (e.g., task instruction wording, font, font size, etc.).

"Similar" category was added to the Lebel et al. (2018) typology to refer to minor deviations or extensions aimed to adjust the study to the target sample that are not expected to have major implications on replication success. See Olsson-Collentine, van Assen, and Wicherts (2020) on meta analysis showing minor to no expected impact due to variations in sample population or setting.

Replication versus the original

LeBel, Vanpaemel, Cheung, and Campbell (2019) criteria:

A Signal Detected in Original Study

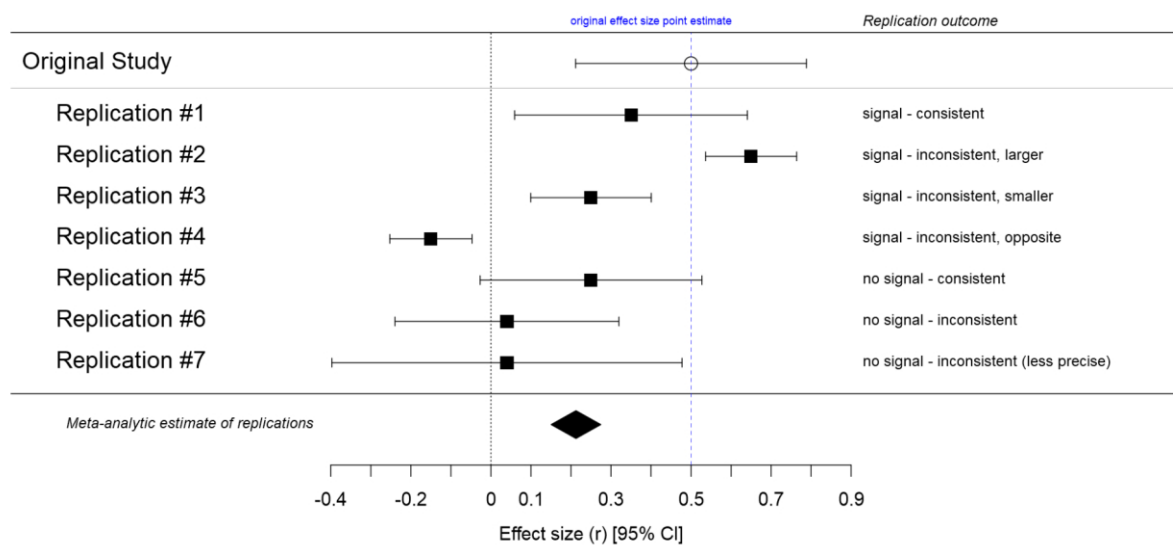


Figure S2. Interpretation criteria for evaluation of replications outcomes by LeBel et al. (2019), if the original study detected a signal. A simplified replication taxonomy for comparing replication effects confidence intervals to target article original effect sizes.

B Signal Not Detected in Original Study

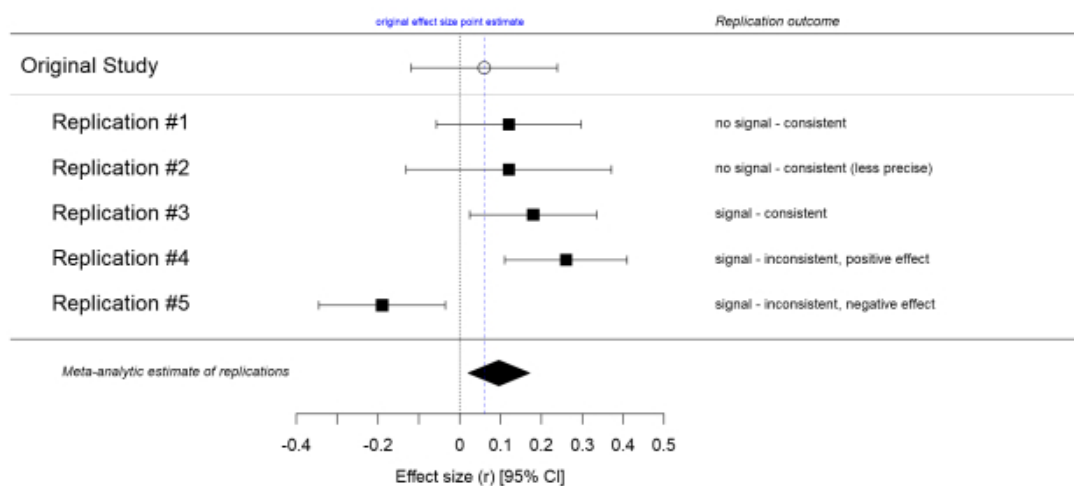


Figure S3. Interpretation criteria for evaluation of replications outcomes by (LeBel et al., 2019), if the original study failed to detect a signal.

References

- Buchanan E, Gillenwaters A, Scofield J, Valentine K (2019). `_MOTE: Measure of the Effect: Package to assist in effect size calculations and their confidence intervals_`. R package version 1.0.2, <URL: <http://github.com/doomlab/MOTE>> .
- Dimitris Rizopoulos (2006). `ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses`, *Journal of Statistical Software*, 17 (5), 1-25. URL <http://www.jstatsoft.org/v17/i05/>
- Hadley Wickham and Evan Miller (2021). `haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files`. R package version 2.4.3. <https://CRAN.R-project.org/package=haven>
- Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). `dplyr: A Grammar of Data Manipulation`. R package version 1.0.7. <https://CRAN.R-project.org/package=dplyr>
- H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- John Fox and Sanford Weisberg (2019). *An {R} Companion to Applied Regression*, Third Edition. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Ken Kelley (2021). `MBESS: The MBESS R Package`. R package version 4.8.1. <https://CRAN.R-project.org/package=MBESS>
- LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, 1, 389-402.
- LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate replications. *Meta-Psychology*, 3.
- Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. *Journal of Open Source Software*, 6(61), 3167, doi:10.21105/joss.03167
- Revelle, W. (2021) `psych: Procedures for Personality and Psychological Research`, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Version = 2.1.9,.
- Tsang, J. (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>