Reply to PCIRR Stage 2 decision letter reviews: Tsang (2006) replication

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response to each item. We also provide a summary table of changes. Please note that the editor's and reviewers' comments are in bold with our reply underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: https://draftable.com/compare/ewZDiUgihZmI (https://osf.io/yxj9m)

A track-changes manuscript is provided with the file: PCIRR-S2-RNR-Tsang2006-replication-extension-main-manuscript-trackchanges.docx (https://osf.io/bwck8)

Summary of changes

Below we provide a table with a summary of the main changes to the manuscript and our response to the editor and reviewers:

Section	Actions taken in the current manuscript
General	R1: We corrected typos and improved clarity and grammar. We standardize the use of past tense.
Methods	R1: We specify the number of subjects who failed to complete the survey and did not begin the survey and therefore were not included in the data analysis.
Results	R1: We revised the wording of "Studies 2 and 3" into "replication of Study 2" and "replication of Study 3." R2: We corrected the typo mistakes of "the rated magnitude of favor".
Discussion	R1: We changed into a more tentative tone for the paragraph suggesting a failure for conceptual replication of Bartlett and DeSteno (2006) R2: We explained the potential influence of order effect on our studies. We extended our explanation to our interpretation for the failed conceptual replication of Bartlett and DeSteno (2006)

Note. R1/R2/R3 = Reviewer 1/2/3

Reply to Editor: Prof. Zhang Chen

> **Thank you for submitting your Stage 2 Registered Report "Revisiting the Effects of Helper Intention on Gratitude and Indebtedness: Replication and extensions Registered Report of Tsang (2006)" for consideration by PCI Registered Reports.**
>
> **Three reviewers who have reviewed the Stage 1 protocol previously have now read your Stage 2 Registered Report. As you will see, all three reviewers are in general positive, and have provided valuable feedback to further improve the manuscript. Based on their comments and my own reading, I would therefore like to invite you to submit a revised version.**

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit.

We addressed all reviewers' feedback below. We also did comprehensive final copyediting to improve clarity and conciseness of the manuscript throughout, which resulted in minor adjustments, all documented in the included track-changes manuscript file.

Reply to Reviewer #1: Prof. Jo-Ann Tsang

> **I appreciate the opportunity to be a part of the review process for this replication. Below are some small suggestions for improvement:**

Thank you for the positive and supportive opening note and the constructive feedback.

> **.1. on p. 19, when listing total participants, was this after exclusions? Please list the number of exclusions, and the reasons for excluding.**

We did not exclude any participants, yet this question helped us realize the need to add more information on an important factor we embed in our online surveys to ensure participant attentiveness.

We added the following to the Procedure subsection of the Method section:

> 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, not sure) - 1) "Are you able to pay close attention to the details provided and carefully answer questions that follow?", 2) "Do you understand the study outline and are willing to participate in a survey with comprehension checks?", and 3) "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 and 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.

A total of 759 passed this stage and proceeded to complete the survey, and were included in our data analysis. We took your query as an opportunity to also include information on those who were filtered at this point, and so we added the following:

> We recruited US American students online through Prolific who answered a Qualtrics survey, with a final sample of 759 participants (Mage = 30.47, SD = 11.02; 297 males, 439 females; 18 other; 5 did not disclose). We note that 907 persons began the survey but 148 did not proceed beyond the consent and verifications (explained in "Procedure"). We did not pre-register any additional exclusion criteria.

> **.2. I noticed in the Methods and Results the authors refer to Studies 2 and 3. This can be a bit confusing, as it is easy to interpret this as the authors running 3 studies. It would be clearer (albeit take more space) to write "replication of Study 2" and "replication of Study 3".**

Thank you. We revised the wording of “Studies 2 and 3” to "replication of Study 2" and "replication of Study 3" throughout the manuscript.

> **.3. On p. 31, under the heading "Replication: Extension Analysis", second line from the heading, I believe the authors meant to write "gratitude but not indebtedness" rather than "gratitude but no indebtedness".**

Thank you for catching that. We appreciate it. We corrected the oversight.

> **.4. I noticed in the Methods in p. 32 (under the "Order Effects" heading), the authors used a mix of past, present, and future tense verbs. It would be clearer if the verb tenses are consistent, such as all past tense. Do check before p. 32, as it may have also occurred before then, but it stood out the most to me on this page.**

We appreciate the feedback. We revised the paragraph and standardized the use of tense to past tense throughout.

> **.5. On p. 52, the authors suggest that their results are counter to previous research linking gratitude to prosociality. I think this is too big of a jump. The authors do not measure prosocial behaviors, or even prosocial behavioral intentions. Additionally, Ma et al.'s 2017 meta-analysis linking gratitude to prosociality suggests there generally is a relationship, and the relationship is not necessarily countered by only one study. If the authors feel strongly about leaving this point in, I suggest making the point more tentatively. This same issue appears again on p. 54 where the authors suggest that De Steno's work was not replicated. Again, the authors did not measure dependent measures that De Steno did, and there is a large literature supporting the relationship between gratitude and prosociality.**

Thank you for the invaluable feedback. We agree.

First, we removed the paragraph referring to prosocial behaviors. Second, we amended all references to prosocial behavior, in order to make it clear that we are looking at reciprocity, which is different from prosocial behavior. Third, we clarify that our reciprocity extension was meant to look at inclination to reciprocate, rather than actual behavior or intent.

We therefore changed the references to our extension to “inclination to reciprocate” throughout, also with adjustments of the introduction. We believe this captures what we did and what we found much better.

Reply to Reviewer #2: Prof. Cong Peng

> **Overall, I believe the authors did an excellent job in implementing the study. They adhered closely to the stage 1 registration to conduct the study, test the hypotheses, and report the results. Very glad to see that this important pioneering work by Tsang (2006), which I cited a lot in my own manuscripts, has been successfully replicated. Congratulations to the authors, and thank you for all the efforts in working this out.**
>
> **Here are some of my concerns and suggestions to improve the manuscript.**

Thank you for the positive and supportive opening note and the constructive feedback.

> **.1. For Study 3, H7, the authors proposed a different hypothesis on gratitude regarding the ambiguous condition (benevolent>ambiguous>selfish) compared to Tsang (benevolent>selfish>ambiguous). This hypothesis is not well justified, nor are the results (and its comparison with Tsang) adequately discussed. Additionally, I would expect the authors to report the t-test comparing the benevolent vs. selfish conditions on gratitude, which is crucial but currently missing. Specifically, the effect size seems much larger in this study (6.59 vs. 5.48) compared to the original one (6.76 vs. 6.49). These findings are important and could be discussed.**

We appreciate the feedback, and the opportunity to clarify and improve.

In Table 1 that includes the H7c hypothesis we previously included the following information:

> Gratitude is higher in the ambiguous condition compared to the selfish-ulterior condition. $p = .07$, $d = 0.49$ [-0.04, 1.01]

This indicated that this hypothesis was not supported, and therefore a null result.

As in the other hypotheses stated in Table 1, we reframed some of the hypotheses from null hypotheses to hypotheses that can be tested with null hypothesis significance testing. This was also the case for H7c. It seems logical to us that an ambiguous control condition would be somewhere between the two clear manipulations, which is why we framed it this way.

To make it clearer that this was the intent with hypothesis H7c, we now write it in the following way:

> Gratitude is [not] higher in the ambiguous condition compared to the selfish-ulterior condition. [Reframed from the target article's null result]

Examining our results, indeed we:

> "... found support for feelings of gratitude in the ambiguous condition (n = 254; M = 6.14, SD = 1.09) as weaker than in the benevolent condition (n = 251; M = 6.59, SD = 0.79; Md = -0.45; t (756) = -4.76, p < .001; d = -0.42, 95% CI [-0.60, -0.25]; H7b), but stronger than in selfish-ulterior condition (n = 254; M = 5.48, SD = 1.24; Md = 0.66; t (756) = 7.09, p < .001; d = 0.63, 95% CI [0.45, 0.81]; H7c)"

If we compare the standardized effect size between the original and the replication using the LeBel et al. (2019) method then the original effect (0.49) is within the confidence intervals of our replication ([0.45 - 0.81]) and therefore deemed "consistent". If we apply the better criteria proposed by Lakens, then the Test for Heterogeneity shows Q(df = 1) = 0.26, p-val = 0.61, also with no indication for heterogeneity in effects. We added that to our Rmarkdown code. Looking at means alone can be a bit misleading, because of the different samples and therefore error of the effect.

We initially assumed that benevolent > ambiguous and ambiguous > selfish-ulterior would lead to the conclusion that benevolent > selfish-ulterior, and so we did not originally include that information. Yet, we agree that it is valuable to also include the statistical reporting of that comparison and therefore added the following at the end of the above paragraph:

> [...] ; between benevolent and selfish-ulterior: t (756) = 11.83, p < .001; d = 1.05, 95% CI [0.87, 1.24]

Thank you for these notes and suggestions, very valuable feedback.

> **.2. The authors did several two-way ANOVAs to examine the interaction effects. What these results mean and its implications are somewhat unclear and unexplained.**

Thank you. The aim with the two-way ANOVAs was elaborated under the section "Data analysis strategy" during the Stage 1 submission. It was meant as a more accurate stringent direct test of the comparison between gratitude and indebtedness. We revised it slightly to help make it even clearer:

> 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 2-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.

We also summarized the interpretation in Table 8 as a complementary to the hypotheses stated in Table 1. These ANOVAs serve as a conceptual replication for testing Hypotheses 2 and 3 simultaneously that "Impact of intent on gratitude (benevolent > selfish) is stronger than on indebtedness.".

> **.3. On page 40 par.2, under manipulation check, it states, "However, we found no support for rated helper intention in the benevolent condition as different from that in the ambiguous condition". This reporting appears incorrect, as this paragraph is about the rated magnitude of favor.**

Thank you very much for pointing that out. This was indeed an oversight.

First, we made it clear that this result was for magnitude of favor, and then added a heading to clearly indicate that magnitude of favor is a covariate, not a manipulation check.

We corrected the relevant paragraph accordingly (added text is bolded):

> **<u>Covariate: Magnitude of favor</u>**
>
> We conducted independent samples t-tests (Welch; two-tailed) and found that the rated in the selfish-ulterior condition ($n = 254$; $M = 5.67$, $SD = 1.32$) was lower than the magnitude in the benevolent condition ($n = 251$; $M = 6.27$, $SD = 0.99$; $Md = -0.60$; $t(469) = -5.8$, $p < .001$; $d = -0.52$, 95% CI [-0.69, -0.34]) and in the ambiguous condition ($n = 254$; $M = 6.16$, $SD = 1.09$; $Md = -0.49$; $t(489) = -4.5$, $p < .001$; $d = -0.40$, 95% CI [-0.58, -0.22]). However, we found no support for **<u>the magnitude of favor</u>** in the benevolent condition as different from that in the ambiguous condition ($t(499) = 1.2$, $p = .20$; $d = 0.11$, 95% CI [-0.07, 0.28]).

> **.4. I did not see the report related to the order effect. I understand that the authors registered to do this only if the data failed to support the hypothesis. But as the authors indicated, "one disadvantage is that answers to one scenario may bias participants". Therefore, I believe it's still important for the readers to understand whether, or to what extent, combining two studies into one may have generated bias.**

Thank you.

To address this request we added the following to the manuscript:

> We 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.

We amended our Rmarkdown to have the default to be run on the full sample, yet with a parameter that can be passed to request the analyses to be conducted only when the studies are presented first. We also included a R file script to run the Rmarkdown in both modes. All those are now included in our OSF folder. As we wrote above, we find the results highly consistent.

Further, we added a new paragraph under the section titled "Implications, Limitations, and Directions for Future Research."

> 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.

> **.5. The results for the extension part are a bit confusing to me. The authors concluded that gratitude and indebtedness are both not associated with reciprocity tendency, but no statistical analysis was presented to directly support this statement. Moreover, based on this conclusion, the authors discussed that this is a conceptual replication of Bartlett & DeSteno (2006) and Peng et al., (2020), which I disagree with. In these two articles, the IV is Help vs. No Help, and the DV is actual helping behavior rather than reciprocity tendency, which are quite different from the current experiment setting.**

Thank you.

In our revision, we revamped and elaborated further the extension results as well as the discussion, to help make things a bit clearer.

Gratitude and indebtedness were associated with reciprocity inclinations (previously "tendency"). We added the following:

> Finally, we added exploratory correlations on 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], $p < .001$), and a weaker positive association with indebtedness ($r(757) = 0.29$ [0.22, 0.35], $p < .001$; z for differences between correlations = 5.71, $p < .001$).

The association we did not find support for an association between expectations for reciprocity and reciprocity inclination:

> We then conducted correlation tests (Pearson's Correlation) and found no support for a link between reciprocity inclination and perceived reciprocity expectation ($r(757)= -0.06$, 95% CI [-0.13, 0.01], $p = .08$). We failed to find support for Hypothesis 13 that perceived reciprocity expectations is correlated with reciprocity inclination.

We changed our previous references to “prosocial behavior” and the connection to Bartlett and DeSteno (2006) and Peng et al., (2020). See also our reply to Prof. Jo-Ann Tsang (reviewer 1) comment number #5. We instead now focus on concluding the following:

> Our findings support Watkins et al.’s (2006) argument 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. (2006), it would be associated with decreased indebtedness and increased gratitude. Our extensions help link between Tsang (2006) and Watkins et al. (2006) into a more comprehensive theory that higher benevolent intent is correlated with lower expectations and therefore higher gratitude than indebtedness.

Best of luck with your future research!

Thank you for all the invaluable feedback.

Reply to Reviewer #3: Dr./Prof. Sarahanne Miranda Field

> **It was a pleasure to see this article in its final form, after having reviewed this as a stage 1 protocol. I thought the protocol was solid, and the completed study reinforces that. It has been carefully and thoroughly conducted (which of course is partly due to the cooperation of the original authors), and written up with plenty of detail which allow the reader to evaluate its quality.**
>
> **[...]**
>
> **All-in-all, congratulations to the authors on a very well-done study! I have no qualms in seeing this article in print!**

Thank you for the positive and supportive opening note and the constructive feedback.

> **As I believe I mentioned in my original review, I think it could be enhanced by the addition of Bayesian statistics, but that isn't a deal-breaker, especially as the p-values do not really come into the ambiguous territory (i.e., most of them are quite small and would be likely to easily recalculate to pro-alternative BFs).**

We appreciate the suggestion, yet opted to keep things as is with only reporting of NHST and effect sizes with confidence intervals. As you pointed out, most of our findings are in support of the target's hypotheses, and Bayesian reporting requires nuance, some of it debatable and subjective, such as setting an informative agreed upon prior, which we would rather not go into at this Stage 2.