

Reply to decision letter reviews: #174

We would like to thank the editor and the reviewers for their useful suggestions and below we provide a detailed response as well as a tally of all the changes that were made in the manuscript. For an easier overview of all the changes made, we also provide a summary of changes.

Please note that the editor's and reviewers' comments are in bold while our answers are underneath in normal script.

A track-changes comparison of the previous submission and the revised submission can be found on: <https://draftable.com/compare/NacKYszQGjnA>

A track-changes manuscript is provided with the file: "PCIRR-RNR-Newman etal 2014 replication & extension main manuscript-TrackChanges.docx"

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
Introduction	<p>R1&R3&R4: We explained the rationale of our exploratory on social norms and added literature on supporting the possible hypotheses.</p> <p>R1: We also elaborated further the addition of options to the political-orientation measure. We consistently use “conservative” and “liberals” in our work. We added the rationale of intuitive true self measure.</p> <p>R2: We elaborated further the rationale behind the authors’ prediction.</p> <p>R4: We changed the word “in synergy” to avoid ambiguity.</p> <p>R1 &R4: We updated the exclusion criteria in the manuscript.</p>
Methods	<p>R1&3: We added an extra question asking respondents to rate their belief about their own intuitive true self. We also updated the tables summarizing all the measures used in this replication and extension.</p> <p>R1: We added one vignette political view manipulation check question in Study 2.</p>
Results	<p>R1&R4: We specified the exploratory test on social norms</p> <p>R1 &R3: We also added comparison between intuitive true self belief and true self evaluations.</p> <p>R2: We updated and added tables based on simulated data.</p>
Discussion	<p>R1&4: We added a future planned discussion regarding the framing of true self rating and the effect of true self across different actions.</p>
Supplementary materials	<p>R1: We updated all the comprehension checks and attention checks.</p> <p>R1&3&4: We added tables about the changes of exploratory on social norms.</p>

Note. Ed = Editor, R1/R2/R3 = Reviewer 1/2/3

Response to Editor: Prof. Chris Chambers

Four reviewers have now kindly evaluated the Stage 1 manuscript, and I am happy to report that they are broadly very positive, which bodes well for the submission achieving in-principle acceptance. The assessments are detailed and constructive, following from careful reading of the manuscript and offering a range of suggestions for improvement. Without providing an exhaustive list, some of the headline issues that I noted include unpacking the theoretical rationale in key areas (including consideration of additional literature), justification of specific predictions (even though, broadly, the reviewers found the hypotheses well motivated), clarification of a range of methodological details, potential strengthening of data quality checks and expansion of manipulation checks, addressing concerns with the MTurk sample selection, resolving 3 discrepancies between methodological details presented in the main text and supplementary information, and considering the validity of ‘true self’ measurements. From an editorial perspective, all of the issues raised seem readily addressable, therefore I am pleased to invite a thorough revision and response.

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit. We appreciate the time and effort you and the reviewers have invested in providing us with constructive positive feedback.

Response to Reviewer #1: Dr. Andrew G. Christy

Overall, my evaluation of this proposal in terms of the five Stage 1 criteria is favorable. In my assessment the proposed research is assessing a scientifically valid question, with an adequate rationale behind the hypotheses being tested, and a basically sound study design and analytic plan that has been explained with sufficient detail to allow other investigators to conduct close-to-exact replications of this work and prevent undisclosed flexibility in the analysis of the resulting data.

For Criteria B – E, I did perceive areas with room for improvement, or at least where the current state of the proposal raised questions that I would encourage the authors to consider as they continue to plan this work. I have outlined these thoughts under the heading of each criterion below. All of these issues are quite minor in scope, and do not seriously compromise the basic soundness of this proposal.

Thank you for the positive supportive opening note, and for your time and effort in helping us improve with detailed constructive feedback.

Rationale for exploratory tests of social norms could be explained more clearly on pg. 13

Thank you, we appreciate the feedback to help us improve.

Given that we planned this as an exploratory test, we decided to follow up on your suggestion and frame this more carefully as competing hypotheses and to try and be more clear about the possible directions.

We updated the “Studies 1 and 2: Perceived social norms” section and elaborated further on the two possible directions, with a link to a similar debate in the experimental-philosophy and social-psychology literatures on the purpose of free-will.

We updated Table 1 with the two competing hypotheses regarding the exploratory extension.

We also designated a subsection in our future discussion section for discussing the exploratory findings and discussing limitations and future directions.

o Rationale given on pg. 14 for adding “independent” and “other” options to political-orientation measure could also be clarified.

Thank you.

We updated the Extensions section with more details in the two subsections “Study 2: Capturing diverse political orientations” and “Study 2: Continuous political orientation measure” with further elaboration of the rationale.

It seems that the words do not were omitted (these additions can represent people who do not identify as Republicans or Democrats).

§ “conservatives or liberals” would be a clearer phrase to use in that sentence also – truer to the content of the item, and clearer as a label for political ideologies.

Thank you for the correction and catching that. The sentence has been updated to:

“who do not self-identify as Conservatives or Liberals”.

Rationale for including the general lay beliefs about good true selves could be clearer (currently no rationale is given, it is simply stated that this measure will be included). If it is being included for purely exploratory purposes (which is what it seems like), that could be made explicit. If the authors intend to test hypotheses related to that variable (even tentative hypotheses), those hypotheses and an accompanying rationale should be stated.]

We appreciate the constructive feedback to improve on this extension.

We now clearly mark the two extensions of lay-beliefs about true selves and social norms as exploratory in the title.

We also updated the extensions section with the “Studies 1 and 2: Intuitive true self belief (exploratory)” subsection elaborating on our rationale and building on the suggestion of reviewer Dr. Caleb J Reynolds that there might be differences between perceptions of true-self of one’s self compared to the evaluations of others’ true self.

We updated the methods subsection “Studies 1 and 2: Intuitive true self beliefs (exploratory)” and the Qualtrics accordingly.

The phrasing of the “surface self” response option implies that everything a person learns (i.e., qualities of the person resulting from “nurture” influences) are by definition not part of that person’s true self.

That doesn’t seem like the most neutral way to measure that concept; while on average a person’s innate qualities might be perceived as more authentic/essential than their acquired qualities, that is not entailed by the bare concepts of the true vs. surface selves. The phrasing of the “true self” response option (“the deepest, most essential aspect of this person’s being”) is more neutral – that phrasing doesn’t directly imply innateness.

Can the “surface self” response option be revised to more closely parallel that language? For example, “This person’s ‘surface self’ (the shallower, non-essential aspects of this person’s being).”

We aimed to conduct a close replication of the study with the materials provided by the original authors. We tried to be very careful to avoid making major adjustments to the framing of the questions, given that we do not know how these changes might impact results.

Therefore, I hope you will understand that we decided to stick with our proposed plan and the original’s material.

Instead, we will address the framing issue of surface self as a limitation of the study in our discussion section together with other limitations indicated by the reviewers.

§ There might be some risk that this phrasing (essential vs. non-essential) is not readily understood by all participants, which could introduce noise/imprecision into the measurements. With that in mind, it may be worth considering more accessible alternative phrasing that preserves the same basic meaning. For example: “The person’s ‘true self’ (who the person really is, deep down)” vs. “The person’s ‘surface self’ (parts of the person that don’t define who they really are).” I’m not convinced this phrasing is perfect, just an example of how these items’ wording might be made more accessible.

We understand the concern. Same as our response above, we decided to stick with the original materials, and instead will be addressing this in our discussion section as a limitation.

o On a similar note – the wording of the true-self rating items in the Study 2 replication is also rather convoluted and potentially confusing. In particular, the word “oneself” in those items is strange – just saying “them” would be clearer (e.g., “...there was always something deep within them calling this person...”). But the overall phrasing and structure of these items is odd and seems likely to confuse at least some participants. Alternative phrasing like: “Now that they have [CHANGE], this person is now...” with a response scale indicating degrees of closeness/farness from being their true self (e.g., 1 = much farther from being their true self, 4 = no closer to or farther from being their true self, 7 = much closer to being their true self)

§ The same considerations apply as above – if the authors judge that the wording should remain relatively unchanged in the interest of fidelity to the original work, that is reasonable. But in that case I would still advocate for changing “oneself” in these items to “them”!

Thank you, great suggestion. This was our own adaptation, and so we are happy to change and improve.

To strike the balance of fidelity and fluency in the sentence, we changed

“there was always something deep within oneself calling this person..”

to:

“there was always something deep within them calling them..”.

o In the consenting process in the Qualtrics survey, it states that the study will include attention/comprehension checks. However, it doesn't seem that any checks are actually included.

While there are quite a few safeguards in place to filter out bots and non-serious participants, it does seem like including some simple instructional attention checks embedded throughout the study, and/or simple comprehension checks (e.g., a single item accompanying each vignette asking participants to identify the nature of the change the target person underwent) could add value when it comes to ensuring data quality. The latter type of items (comprehension checks) would probably be sufficient, since they effectively double as attention checks. Including these checks would likely result in a more exclusions. This could be offset by setting a relatively high bar for exclusion (i.e., participants would only be excluded if they displayed a persistent pattern of inattentiveness/non-comprehension across multiple check items). Increasing the planned sample size further could also offset a larger number of anticipated exclusions, though I'm not sure that's necessary.

We included the attention and comprehension checks within the consent form before embarking on the study.

We summarized the comprehension checks and manipulation checks in the supplementary materials in the section "Attention and comprehension checks". We also introduced manipulation check extensions that were not included in the original study.

We believe that those combined with the many other checks that we embedded are more than sufficient to ensure high data quality. We have used similar checks and controls in many other replications successfully.

Also, to try and increase attention to key details in the manipulations and items, we highlighted the direction of change in each vignette and enlarged font size for the description of agent's changes.

Given that we already multiplied the sample size by 2.5 and added extra for exclusions we calculated that we are well-powered to detect very weak effects, we felt like adding additional participants was not necessary.

In the supplemental materials, it is stated that recruitment on MTurk will be limited to workers who have completed between 5,000 and 100,000 HITs. That strikes me as odd; only highly experienced MTurk users would be eligible for participation in this study, and more casual users would be excluded. Allowing less experienced MTurk users to participate would seem desirable for a number of reasons (i.e., greater proportion of naïve participants, somewhat more representative of people in general).]

We are interested in the representativeness, but of minimally qualified and experienced participants. We set this to 5000 given that 5000 tasks ensures that these MTurkers have completed enough tasks to be familiar with the platform and are serious about their work in this platform (when combined with approval rates). The average task/survey on MTurk is 5-10 minutes which means that an MTurker who completed 5000 tasks has a minimum of ~100-150 hours on the platform, which in a work week of 30-50 hours means being on the platform for 2-3 weeks, enough for the workers to become familiar with the platform and understand the workflow and has participated in enough tasks to overcome technical issues.

Addressing the issue of naivety on MTurk does not seem crucial here, and aiming to address this can introduce a host of participant quality issues that we would rather avoid. Our experience from the many replications we completed using these settings has resulted in very high level responding, and we would like to keep it that way.

The exclusion criteria described in the manuscript do not match exactly with those outlined in the supplemental material

Thank you for pointing this out. We removed the section from the supplementary and replaced the section in the main manuscript with:

“We focus on our analyses on the full sample of all participants who completed the study. However, if we fail to find support for the hypotheses we will rerun our analyses with exclusions, and report a comparison of the findings pre and post (see “Pre-exclusions versus post-exclusions” section in the supplementary materials). We will consider these exploratory. Our planned exclusions in case of failure at this stage are: 1) Participants indicating a low proficiency of English (self-report < 5, on a 1-7 scale), 2) participants who self-report not being serious about filling in the survey (self-report < 4, on a 1-5 scale).”

Manipulation checks were added to the Study 1 replication, but not in the Study 2 replication. It would seem possible to add a similar morality rating there, allowing the authors to test whether the “pro-liberal” and “pro-conservative” changes were indeed perceived as changes for the better by liberal and conservative participants, respectively.]

Great suggestion, thank you for that. We agree.

We added the following measure:

“Do you perceive this person’s change as liberal or conservative?”.
(-100 = Pro conservative; 0 = Neither; 100 = Pro liberal.)

Response to Reviewer #2: Dr. Cillian McHugh

Thank you for the opportunity to review this work. I am enthusiastic about the proposed project. I think there is considerable value in replicating previous work. The authors here also include an extension element that could potentially further our understanding of the mechanisms at play in this paradigm. I applaud the authors on their efforts. I think this is a very strong proposal and I look forward to seeing what the results will be.

Thank you for the supportive and encouraging opening note. We appreciate that very much.

I have some comments for possible improvements (particularly in relation to theory and framing). I have provided my comments under the recommended evaluation criteria below.

However, I think there is room for a more in depth discussion of the theoretical rationale for the hypotheses. This is especially true for H3 (the extension). The authors provide a descriptive rationale for why one might expect the proposed relationship, but this rationale does not appear to be grounded in existing literature or theoretical approaches.

Thank you for highlighting this, we made some changes to that section. Please see our reply to a similar comment made above.

I would also like to see an expanded discussion of the theory informing H1 and H2, why did the original authors arrive at these predictions? What is their relevance in contemporary theory?

Please note however, that I'm not requesting an exhaustive review of the literature, but I would like to see a bit more depth of discussion than is currently provided.

We appreciate the feedback that would try to do better to clarify and improve the introduction. Given that this is a replication, and that we are based on an existing paper with theory and hypotheses, we tried to minimize repeating what is in the original, and focus on the empirical aspect of replication. We do recognize, however, that we need to provide more context to readers.

To address this feedback, we briefly elaborated on the authors' prediction under the subsection of "Hypotheses and findings in target article" in the manuscript.

In addition to unpacking the rationale that supported the development of the original hypotheses, there is some interesting work on moral character that I think is relevant to the current study that could enhance the framing of the introduction. In particular these offer an alternative perspective and potentially offering competing hypotheses. I have provided a very brief summary of some of the key points in my own work (see McHugh et al., 2022, pp 136-137) but I encourage the authors to study the original papers and either integrate these potential alternative predictions into their manuscript, or to clearly differentiate between these lines of research. The papers I am thinking of are Klein and O'Brien (2016), Siegel et al., (2018) Klein and O'Brien (2016) demonstrate that when judging transformations of moral character, people appear to perceive negative transformations much quicker than positive transformations. Participants were presented with vignettes describing changes in patterns of behavior and asked to indicate how many consecutive instances of the new behavior would need to occur to convince them that the actor's moral character had transformed. Participants relied on fewer instances to judge a negative transformation than a positive transformation.

Similarly, Siegel et al. (2018) report that beliefs about bad agents are more volatile than beliefs about good agents, noting that, "bad people often behave morally, but good people rarely behave immorally" (p. 750). Both these examples potentially undermine the true self narrative. I suspect the difference in design (repeated vs single exposure) means these may not be directly comparable to the current work. However, I think there is broader theoretical overlap that I think could strengthen the current work, if integrated (e.g., offering a framework for interpreting results that run counter to H1).

Klein, N., & O'Brien, E. (2016). The Tipping Point of Moral Change: When Do Good and Bad Acts Make Good and Bad Actors? *Social Cognition*, 34(2), 149–166. <https://doi.org/10.1521/soco.2016.34.2.149>

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2022). Moral Judgment as Categorization (MJAC). *Perspectives on Psychological Science*, 17(1), 131–152. <https://doi.org/10.1177/1745691621990636>

Siegel, J. Z., Mathys, C., Rutledge, R. B., & Crockett, M. J. (2018). Beliefs about bad people are volatile. *Nature Human Behaviour*, 2(10), 750–756. <https://doi.org/10.1038/s41562-018-0425-1>]

Thank you for the suggestion and the citations. We appreciate those, and have spent considerable time diving into the suggested literature to try and understand how we might integrate this into our introduction.

Apologies, but after giving this much thought, we hope you will understand our decision to not elaborate on this point and this literature in this manuscript. We feel that this goes above and beyond the scope of a replication, and has the potential of both confusing readers and distracting from the main point of simply repeating what the original findings have found. Given that we are not conducting an extensive review of the literature, we also worry about citing specific studies from what seems like a very broad moral judgment literature that is not directly related to the target article and true-self attributions.

What we can say is that we will keep those citations in mind when constructing our discussion section following Stage 2 to reflect on the findings, especially if those will go counter to the original's.

We hope that this is reasonable, but are of course willing to revise this decision, given clear editorial directives.

Re-check the results for Study 2. Table 12 is empty. I also couldn't see any mention of an interaction discussed when reporting the 2 (political view: liberal and conservative; between) x 2 (item types: liberal and conservative; within) mixed-model ANOVA (the original study reports such an interaction, p. 207)]

Thank you for raising this. The result section was meant as a simulation and an illustration for what the results section might look like after data collection. The random noise data is meaningless at this point, so we did not fill in all the tables.

That said, we do appreciate the opportunity to improve and make changes to the results section to make those more comprehensive, clear, and with better illustrations of the results. We did not make the tables labels clear enough, and in this revision we changed those labels and other aspects that would make it clear that we are reporting this interaction in what is now Table 9.

Response to Reviewer #3: Dr. Caleb J. Reynolds

This Stage 1 submission is clear, precise, and documents essential elements of an RR throughout. I commend the authors on their careful and thorough preparation, and I recommend the submission with no reservations.

Thank you for the positive and supportive opening note. We appreciate that very much.

The hypothesis about social norms is clear and seems reasonable, but it not terribly well justified in the text beyond the assertion that, if true, it would suggest an interesting paradox (p. 13). Given that this is the only hypothesis not from the target article, another sentence or two of justification for this prediction would be helpful (indeed, wouldn't the implication of a paradox suggest one should hypothesize the opposite?).

Thank you for the feedback to do better here.

Please see our detailed response to a similar point made by Reviewer 1 Dr. Andrew G. Christy. In our revision we elaborated on this direction further, as you suggested. Please see “Studies 1 and 2: Perceived social norms (exploratory)” subsection in the introduction.

The measure of intuitive true self belief (p. 26) seems odd. In true self ratings in both studies, true self is the particular target's true self, that is, true self is localized. Yet for this general measure, true self becomes this disembodied universal that exists apart from individual people. I know the literature sometimes talks about true self in a way that projects this, but I'm skeptical that this is how lay people think about the concept of true self. Assuming this is an ad hoc measure, I would suggest the authors consider revising the items. Perhaps something calling out individuals' true selves might help? e.g., “Most [all?] people's true self is morally good”.

This is a great point and a good idea. This does require clarification and it points to potential differences in perceived true-self in self and others.

We changed the framing of this question, and split it into two. We now make the distinction between one's own true-self and the average person's true-self. Please see revised “Studies 1 and 2: Intuitive true self belief (exploratory)” subsection in the introduction.

Response to Reviewer #4: Dr. Sergio Barbosa

I thank the authors for the opportunity to review this paper and for an interesting and rigorous project that will surely meaningfully contribute to literature. I hope I get to see this published manuscript soon and that my comments might make this process a bit less difficult.

Thank you for the positive and encouraging opening note.

However, i do feel that seriousness and language proficiency items deserve to be explicitly written in-text rather than left in supplementary material. As is it reads as quite confusing and the reader is left wondering how do authors assess these.

We appreciate the feedback.

We now include the seriousness and language proficiency item in the manuscript. The measures are detailed under subsection “Procedure” in the “Method” section and the use in potential exclusions elaborated under subsection “Exclusion Criteria” in the “Method” section.

However, I am not convinced by hypothesis regarding social norms having a positive link to true self. I would expect quite the opposite effect as people might view conformity to a social (as opposed to moral) norm as reflecting the desire to please others or to avoid the social costs of counter-normative actions rather than the expression of personal conviction (i.e. arguably part of the true self). Of course, authors need not agree with me but I believe further justification of this hypothesis is certainly beneficial.]

Thank you, good point. A similar point was made by Reviewer 1 Dr. Andrew G. Christy.

In our revision we elaborated on this direction further, and as you suggested we now clearly mark this direction as “exploratory” and frame this as competing hypotheses.

Please see “Studies 1 and 2: Perceived social norms (exploratory)” subsection in the introduction.

On a similar note, I think interpretation of this, possible, result as reflecting that participants believing that the true self is in synergy with others is not the only plausible explanation to this pattern of results, admitting they are actually observed after data collection. One might also interpret the correlation between social norms and true self as reflecting the true self being influenced by what people believe others believe is moral without necessarily being "in synergy". I am interpreting "in synergy" as some sort of coordination in real life. Perhaps this is simply an issue with this particular expression but I believe it leaves some ambiguity that authors might want to avoid.

Thank you for the feedback.

We agree, and have revised the section discussing this extension. We also removed reference to “synergy” and reframed this about alignment with perceived social norms. Given the correlational design we would not be able to draw any insights about causality, but we hope that this exploratory direction can help spur future research about the role of social norms.

However, if authors find it interesting and doable with available resources I do suggest comparing scenarios in which the agent changes their behavior and where they change their opinions/ thoughts. I am sure authors are familiar with literature on the effect of intentions on moral judgment or on harmless wrongs relevant to this topic. I believe a possible extension without more data being collected is actually observing whether true self link to moral or social norms is identical when considering actions or thoughts.

In this study we focus our efforts on the replication of the classic article, which we hope will encourage future research directions. This is an interesting suggestion, but one that goes far beyond the scope of our current investigation.

I believe they have done a great job at it and have nothing to suggest other than discouraging the use of G*Power as its interpretability and usefulness is hotly debated. I suggest a full simulation approach if at all possible. Even as a user of the pwr package I am not sure it allows for power and sensitivity analysis of interactions which, I believe, compels us to consider simulation approach.

Thank you for the comment. We are not sure what debate you were referring to regarding GPower, yet we appreciate the suggestion to try and do even better and to supplement this with a simulation.

First, we note that the mixed design is not needed, given that the between factor is block (order) which the original authors never reported stats for and concluded no support for effects.

Therefore, although we implemented this in our design, we do not feel that it is important to be powered for that or include that in the power analysis. Therefore, for the core hypotheses this is a simple repeated design, which can be handled by pwr.

Moreover, especially given the within-design, we see no reason for concern. Our sensitivity analyses using G*power indicates that we are extremely well powered, above and beyond what is needed for much weaker effects than those reported in the original.

To address this concern, we added a simulation analyses of the mixed design using ANOVA/Super_power shiny app. We note that since descriptives and effects are not reported for the between factor, we just set it at no effect, and we are not aiming to detect that effect.

In the supplementary we now report both the G*Power sensitivity analysis, subsections “Supplementary power analysis” and “Sensitivity analysis”, right after the effect size calculations and power analysis.