Reply to decision letter reviews: #180

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 the changes.

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

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

A track-changes manuscript is provided with the file: "PCIRR-RNR-Smith-etal-2002-rep-ext-manuscript-v2-G-track-changes.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
General	Ed: We improved the introduction section, added sensitivity analyses, and responded to the concerns over the possible effects of deviations on the observed effect
	R1: We added discussion of the self-behavior view in the introduction, improved the power analyses, and discussed the deviations either in the response letter or in the revised manuscript.
	R2: We revised the abstract, added discussion of the self-behaviour view in the introduction, and discussed the possible effects of deviations on the observed effect.
Introduction	Ed, R1, & R2:We added the discussion of the self-behavior view in the introduction.
	R1: We addressed the issue of previous direct replication.

Section	Actions taken in the current manuscript
Methods	Ed & R1: In addition to the sensitivity analyses for main effects, we also added the analyses for the two-way interactions that are connected to the hypotheses and added the reliability analyses in the analyses plan.
	R1: We discussed the possible effects of attention checks and revised the exclusion criteria. We wish to not perform factor analyses for the scales used as it is beyond the scope of the replication. We also addressed the issue of performing mixed ANOVA instead of multilevel models.
	R2: We offered our rationale to not exclude participants based on the devices used or run exploratory analyses.
Results	
Discussion	R1:We will add the discussion about the possible effects of deviations on the results. We will acknowledge the importance of scale validity research in the future directions section.
	R2: We will add the discussion on how interpersonal motives may interact with public exposure to influence shame and guilt.
Reporting	R1: The first author double-checked the grammar and spelling.
Supplementary materials	R1: We added the protocol for the added sensitivity analyses in the supplementary.

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

Response to Editor: Prof. Chris Chambers

Two expert reviewers have now assessed the Stage 1 manuscript. As you will see, the evaluations are broadly positive, with both reviewers praising the value of the replication and the methodological rigour of the proposal. There are, nevertheless, some conceptual and methodological issues to address in order to satisfy the Stage 1 criteria. Key concerns raised include the accuracy of the framing in the introduction, ensuring that the power analysis targets the appropriate (finest grained) level of the design, consideration of the validity of the attention checks, making clear all deviations from the original study design and ensuring they are scientifically valid, and clarification of the analysis plans.
In revising, please keep in mind that while it is perfectly fine to bring up new literature in the Stage 2 discussion section, it is generally not possible to add new literature to the introduction section after in-principle acceptance. Therefore, in responding to the reviewers' comments, please ensure that the introduction is brought now into a state that will require minimal (if any) changes at Stage 2.

Thank you for the reviews obtained, your feedback, and the invitation to revise and resubmit. We revised the manuscript based on the feedback.

[Disclosure: For the sake of transparency, we note that Reviewer Prof. Roger Giner-Sorolla was a collaborator/coauthor with the corresponding author Gilad Feldman on various large-scale open-science related projects.]

Response to Reviewer #1: Prof. Roger Giner-Sorolla

> **In general there is very good practice here for highlighting original-replication comparisons and for ensuring that the final report follows the registration closely.**

Thank you very much for the positive opening note and the detailed feedback.

> **1. Abstract: "impacted differently by exposure" should be expanded for clarity, e.g. "... by appraisals of potential exposure to disapproving others"**

Thank you for this comment. We elaborated on the findings of Smith et al. (2002) and clarified the meaning of exposure in the revised abstract. The sentence is now as follows:

> "Smith et al. (2002) demonstrated that shame and guilt are distinct and are impacted differently by public exposure, that is, the (potential) exposure to disapproving appraisals of one's misdeeds by others. The impact of public exposure (compared to no exposure) was greater for feelings of shame than for feelings of guilt".

> **2. Authors should follow APA style when citing modern translators and commentators on ancient philosophers, see 7th ed. style manual, p. 325, example 36.**

Thank you.

We revised the references and in-text citations for the two ancient greek works:

> Aristotle. (1941). The basic works of Aristotle. (R. McKeon, Ed.). Random House. (Original work published ca. 350 B.C.E.)

> (Aristotle, ca, 350 B.C.E./1941)

> Plato. (1997). The collected dialogues of plato. (E. Hamilton, H. Cairns, & B. Jowett, Eds.) (Ser. Past masters ser). Intelex Corporation. (Original work published ca. 405 B.C.E.)

> (Plato, ca, 405 B.C.E./1997)

3. The Introduction should position the public exposure/reputation view of shame vs. guilt more clearly against a dominant rival view, the Tangney interpretation of Lewis' Shame and Guilt in Neurosis that shame is distinct from guilt in being subjectively about the whole person rather than a single transgression. This omission is much missed, for example on p. 9 where the GASP is mischaracterized as being mainly about the public/private dimension, when actually it intentionally conflates that dimension with several others derived from the Tangney viewpoint. This addition will only strengthen the case for the importance of this study.

Thank you for this valuable constructive feedback.

We added a subsection titled "The Self-behavior Distinction" in the introduction and briefly introduced this view. We also corrected the interpretations of the GASP development by clarifying that the development of the scale took into consideration both views.

The revised section is as follows:

"Note that the public vs. private distinction is not the only mainstream thought regarding the differentiation between shame and guilt. Another influential explanation posits that the object of negative evaluation (self vs. behaviour) can differentiate shame and guilt (Tangney, Stuewig, & Mashek, 2007). While shame focuses on the negative evaluation of the global self (e.g., being a bad person), guilt is associated with the negative appraisal of the specific behaviour (e.g., having done a bad thing). Previous research has found that when describing shame-inducing situations, participants expressed more concerns about negative evaluations of the self, compared with guilt-inducing situations. While the opposite holds for concerns about the effect on others (Tangney, 1994).

However, as expressed by Tangney and colleagues (2007), the two schools of thought need not be mutually exclusive. The salience of public exposure could shift the individual's attention to the self and thus induce a greater feeling of shame. While in private, the individual may pay more attention to the effects of their behaviours on others. Other research has also pointed out that both accounts receive empirical support and a new scale measuring shame and guilt-proneness (Guilt and Shame Proneness scale, GASP) has been developed taking into consideration arguments from both sides (Cohen, Wolf, Panter, &Insko, 2011). GASP measures guilt-proneness using negative behaviour evaluations towards private transgressions and shame proneness using negative self-evaluations towards public transgressions.

In view of the debate over the two schools of thought, it is thus more important to ensure the empirical foundations of the theorizing are reliable and replicable, which we hope to contribute by replicating one of the classic findings: Smith et al.(2002)."

4. An important part of a case to replicate a specific article is that it has not been directly replicated already, and/or that conceptual replications/extensions are few - can the authors speak to this issue?

We agree that the value of a replication project also depends on whether or not there have been replications already. To the best of our knowledge, there are no direct replications of Smith et al. (2002). We added this information in the revised manuscript:

> "However, to the best of our knowledge, despite its impact there seem to be no published direct replications of Smith et al. (2002)."

5. Power analysis should make clearer that main effects and not the interaction + simple effects (because the original did not find it significant) were the main hypotheses of the study. If testing the interaction is important then the study should be powered with regard to simple effects tests interpreting the interaction

This is an important comment.

In this replication study, we aim to replicate the main findings hypothesized by the original article: the effect of public exposure on shame and moral belief on guilt. We do not plan to test whether public exposure and moral belief interact to influence the two emotions. However, to test the proposition that public exposure impacts shame more than guilt, we also planned the mixed 3 (Public Exposure: private vs. implicit public vs. explicit public) × 2 (Moral belief: Low vs. High) × 2 (Emotion: shame vs. guilt) ANOVA with Emotion being the within-subject factor, following Smith et al. (2002). In the revised manuscript, we also added the sensitivity analyses for this test, specifically, the 3 (Public Exposure: private vs. implicit public vs. explicit public) × 2 (Emotion: shame vs. guilt) interaction and the 2 (Moral belief: Low vs. High) × 2 (Emotion: shame vs. guilt) interaction.

The added analysis is as follows:

> "We also ran sensitivity analysis for the two-way interactions (Public Exposure × Emotion and Moral belief × Emotion) mixed 3 (Public Exposure: private vs. implicit public vs. explicit public) × 2 (Moral belief: Low vs. High) × 2 (Emotion: shame vs. guilt) ANOVA using Morepower (Campbell & Thompson, 2012). Results showed that a sample of 1260 would allow the detection of $f = 0.10$ with a power of 95% and alpha of 5%."

6. Was an attempt made to contact the original authors for the missing disobeying scenarios?

The corresponding author reached out to the original authors but did not receive a response. However, we do not wish to add this information to the manuscript as there could be many reasons why the authors did not respond and we would rather keep things positive.

7. pp. 17-18 - if I understand correctly participants can only proceed after correctly answering or re-answering attention check questions. This is a deviation from the original that will likely increase effect size and should be questioned. It might be more defensible to not force this kind of learning and look at responses only with (better test of idea) or also without (better replication of original) a correct check response.

Thank you for the comment.

We are hoping that the employment of comprehension checks this way will lead to higher quality data and better use of the sample's data, given that it is likely that participants will pay closer attention to the manipulated details and will therefore ensure both understanding and attentiveness.

We consider this a major strength of the current design, and highly needed given the chosen target sample. This signals seriousness on the experimenters' part and the need for close attentiveness on the participants' part. We much prefer this to excluding participants post hoc. We have employed this design in many of the other replications our team has completed with this target sample and have come to consider these as important.

The increased attentiveness may have an effect on the observed effect size, though we would hope that the participants of the original's and many studies conducted are attentive, even without these checks. This was noted as a deviation from the original.

8. Is the outlier exclusion a deviation from the original?

The original article did not report the criteria for outlier exclusion.

The classification of outliers and exclusion criteria can be a deviation. In this revision, we decided to remove the outlier analyses in order to be consistent with the original article, and given feedback advising us to remove these analyses in our other PCIRR submissions.

However, the majority of the exclusion criteria are necessary to make sure the sample is comparable to the original sample (e.g., English proficiency, seriousness etc.). After consideration, we decided to remove the exclusion criteria of guessing the hypothesis since this was not mentioned in the original paper. (see supplementary materials *Exclusion criteria*

section). In the revised manuscript, we added the statement about the possible deviation in the main manuscript.

> "Note that the original article did not report any criteria or operations for exclusions. Therefore, there is a possibility that the exclusion process will constitute a deviation from the original study. However, most of the exclusion criteria we employed were to make the sample more comparable to the original sample and ensure data quality, which makes it unlikely to be a consequential deviation."

> **9. I agree it is important to test the validity of the measures as face-wise several of the shame-related measures seem conceptually and empirically shaky (e.g. I usually find anger at self loads on guilt and not shame). Rather than simply testing correlations it might be better to test omnibus reliability and also discriminant validity (e.g., factor analysis of all items) of all the shame and guilt measures, both direct and indirect, together. This would let us assess, improving on the original, whether each "related" measure is correctly classified or not.**

The validity of the measures is of importance. However, this goes far beyond the scope of the current replication and is not core to our aims. This deserves a full-on investigation on its own and there have already been other follow-up studies that tackle this direction (e.g., the development of the GASP scale by Cohen et al., 2011). We would rather not include the factor analyses yet we will acknowledge that there is also much other research that can be done with the data of the proposed study in the discussion section. We will, of course, make all our data/code available for anyone who would like to follow up and look at these and other future directions.

> **10. Also, there is a discrepancy where the results test individual scale reliability but the analysis plan does not.**

Thank you for this comment. We added the tests of reliability in the analyses plan: confirmatory analyses section.

> "Therefore, after performing internal consistency analyses (i.e., Cronbach's Alpha) following the original study, we first ran and reported the Pearson correlations between explicit shame and shame-related reactions as well as the correlations between explicit guilt and guilt-related reactions to establish the validities of the measures."

> **11. If transforming data into long format, multilevel (hierarchical) analysis needs to be followed with a participant as a random factor in order to properly account for nonindependence of observations. However, treating shame/guilt as a within-participants factor while keeping wide format data**

would be simpler. It is not clear from the writing which analysis will be used

Thank you for paying attention to such detail. The reason we transform data to long format is that the afex package requires long format data to conduct ANOVA in R (in comparison to SPSS, which uses the wide format). We have specified the random factor (ie. participant ID) and accounted for the nonindependence of data in the analyses code.

The specific code is as follows:

Three-way ANOVA

Explicit_emotion<- afex::aov_car(Emotion_score ~ Publicity*Moral_belief*Emotion + Error(subject/Emotion), data=dat_long)

Explicit_emotion

Although multilevel modeling (MLM) makes fewer assumptions than mixed ANOVA and it could be advantageous to employ such technique in current practice, we choose mixed ANOVA for the following reasons:

1. The original article did repeated-measures ANOVA with emotion (Shame vs. Guilt) as a within-subject factor and since we are trying to replicate the study, it is important that we employ the same analyses approach.
2. One of the great advantages of MLM is that it can deal with missing data instead of list-wise deletion. In our case, participants will be asked to complete all the measures and incomplete cases will be removed from analyses (see supplementary materials *Exclusion criteria* section). Therefore, the advantage of handling missing data does not warrant its usage here.

   12. In general there are a few grammar and word errors noted that suggest further proofreading is needed.

Thank you for pointing out this issue.

The first author double-checked the spelling and grammar in this revision.

Response to Reviewer #2: Dr Uriel Haran

This report describes a replication plan for Study 1 of Smith et al.'s 2002 paper about the effect of public exposure of wrongdoing on ratings of guilt and shame. The paper has been pretty influential, with over 600 citations on Google Scholar, and has been published in the Journal of Personality

and Social Psychology a long time ago. Therefore, other than determine whether the study is worthy of replication (which it definitely is), there is not much to say about the merits of that research, the relevance of the original research question, the rigor of the original study design etc. As long as the authors follow the research protocol employed in the original study, they are not obliged to independently satisfy the evaluation criteria of that work.

The intended replication follows the original protocol pretty closely, including the experimental design, adjusted sample size and analytical approach. The authors had to write the stimuli of one of the conditions, which was not provided in the original article. The new text seems valid: it is close enough in length and in format to the texts in the other conditions. The authors are also adding a couple of attention check items, which is a legitimate addition, and standard in such studies in 2022.

Thank you, for the positive encouraging opening note.

1. One difference that is noteworthy and should be discussed in the replication manuscript is the physical settings in which participants will complete the study. The original study was conducted in the lab with about 30 people in the room per session. The replication will be conducted online, with participants completing the study on their own electronic devices, presumably in private.

Thank you for this comment. We acknowledged this deviation in the original manuscript in Table 4 Classification of the replication, based on LeBel et al. (2018):

Physical settings   Different   Replication was conducted online individually; The original experiment was carried out in person in groups of 30.

There is the possibility that this deviation will have a significant effect on the observed effect size. More precisely, the private nature of the current setting could make it harder for participants in the public exposure conditions to imagine the situation, shrinking the observed effect size.

However, the same can be said about the original study. The relatively public nature of the physical setting could also heighten the emotional response of participants in the private condition. Therefore, it is reasonable to assume that this deviation will not impact the observed effect greatly.

We will discuss how the deviations could affect the results in the Stage 2 report including this one.

> **2. I recommend also coding the type of device on which participants do the study, and if possible prevent people from completing it on their mobile phones, as research in information systems finds systematic differences in user attention between tasks performed on a PC or a tablet and tasks performed on mobile phones.**

The type of device may or may not have some kind of impact. However, recent research has shown that almost none of the Mturk participants used a mobile device in their study (Eyal et al., 2021). Taking into consideration that we have employed plenty of techniques, designed a fair compensation scheme, and provide participants with clear and respectful instructions, we believe that the device used is not a major concern for the data quality or attentiveness.

> **3. The abstract states that guilt and shame are similar in that they both are associated with negative evaluations of oneself, but this is not entirely accurate. Guilt, unlike shame, is associated with negative evaluations of one's behavior, separately from one's view of one's qualities and characteristics (i.e. "I did a bad thing" rather than "I am a bad person"). This externalization of the emotion-eliciting wrongdoing is what distinguishes guilt from shame. See Tangney & Dearing, 2002.**

Thank you for this valuable feedback.

1. We changed the expression to the following in the revised manuscript: "Shame and guilt are unpleasant self-conscious emotions associated with negative evaluations *of oneself or one's behavior*." (Page Line)
2. The self/behavior distinction is a major school of thought on the research of shame and guilt. The lack of discussion about this perspective has also been raised by reviewer Prof. Giner-Sorolla. In the revised manuscript, we added the subsection discussing this approach and its relevance to the current study. The section is as follows:

> "Note that the public vs. private distinction is not the only mainstream thought regarding the differentiation between shame and guilt. Another influential explanation posits that the object of negative evaluation (self vs. behaviour) can differentiate shame and guilt (Tangney, Stuewig, & Mashek, 2007). While shame focuses on the negative evaluation of the global self (e.g., being a bad person), guilt is associated with the negative appraisal of the specific behaviour (e.g., have done a bad thing). Previous research has found that when describing shame-inducing situations, participants expressed more concerns about negative evaluations of the self, compared with guilt-inducing situations. While the opposite holds for concerns about the effect on others (Tangney, 1994).

> **4. There is more modern research in psychology about guilt vs. shame in general (e.g., Cohen et al., 2011; Tangney et al., 2007) and particularly**

about the public-private dimension. Some of these works are cited later in the report (p. 9 Choice of replication), but it should be discussed earlier.

Thank you for the advice on re-organizing the introduction section.

We made changes to address this in our revision. Specifically, we added the subsection discussing the self-behavior distinction view and how these two major explanations are supported and how they influenced subsequent research, that is, the development of GASP.

5. One wrinkle in the public-private distinction between guilt and shame is that guilt enhances interpersonal motives such as the desire to be loved and accepted by others (Baumeister et al., 1994). These motives might moderate the effect of public exposure of one's behavior on one's feelings of guilt and shame. Again, this does not affect your replication research but might be a point worth addressing in your discussion.

Thank you for this comment. If we understand you correctly, what you are suggesting is that given that guilt theoretically is associated with enhanced affiliative motive, public exposure could also have an impact on guilt by making the objects of affiliation (i.e., others) more salient. That is, we could observe increased levels of both shame and guilt in the public exposure conditions compared with the private condition.

This is indeed a valid hypothesis. However, in the original study, Smith et al.(2002) did not report support for an effect of public exposure on guilt. We, therefore, are not sure regarding the possibility of observing an effect of public exposure on guilt. In the case we do find such an effect, we will discuss this possible explanation in the Stage 2 report and possible directions on how to disentangle these mechanisms in the Future Directions section.