Reply to decision letter reviews: #179

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/YeWOVmMiGotH

A track-changes manuscript is provided with the file: PCIRR-RNR-Weiner etal 1988_ replication & extension main manuscript-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	R1: We provided clarifications on key terms in the abstract. R1 & R2: We changed the wording of stigma labels and replaced "pity" with "sympathy" in the Qualtrics survey and in discussion related to the current replication. R2: We added a definition of paraplegia in Qualtrics.
Introduction	R1: We provided more information on the consistency of key findings from the target article. R1: We provided more information about physically-based and mental-behavioral stigmas and made clarifications regarding the theoretical background. R1: We provided clarifications on the target article's experimental design and procedures. R1: We added more references to better support our suggested extensions and exploratory analyses. R1: We updated Table 4 for clarity. R2: We edited the section 'Choice of study for replication' for clarity. R2: We rephrased Hypothesis 8 for clarity. R1 & R2: We provided clarifications on the definition of 'stigma'.

Section	Actions taken in the current manuscript
	R1 & R2: We updated Table 1 and the section 'Choice of study for replication' to clarify the reproducibility issues. R1 & R2: We added exploratory directions. As there is an increased focus on comparing perceptions of physically-based versus mental-behavioral stigmas, we decided to completely follow the original study's method and included the variable "stigma source" in conducting our correlation and regression analyses.
Methods	R1: We changed the simulated dataset to include 800 participants to match our target sample size in the actual data collection. R1: We clarified that there would be equal amounts of participants in each of the conditions. R1: We moved questions on categorizing stigma by source to the end of the survey. R1 & R2: We changed the experiment to a true between-subjects design and updated the data analysis strategy to align with the new design and the exploratory analyses.
Results	R1: We changed all indications of sample sizes to match our updated target sample size of 800 participants in total.
Discussion	R2: We added a 'Limitations' section to discuss the possibility that people may have different perceptions of the term 'stigma'.
Supplementary materials	R1 & R2: We included the Pearson's r from correlations with "stigma source" in the target article in "Effect size calculations of the original study effects".

Note. R1/R2 = Reviewer 1/2

Response to Editor: Prof. Chris Chambers

Two reviewers have now provided rapid and helpful assessments the Stage 1 manuscript. As you will see, both evaluations are generally positive, noting the value of the replication (and extension), as well as the adherence to rigorous open practices. The reviews do, however, note a range of areas requiring careful revision, including strengthening of the study rationale, clarification (and likely correction) of a range of specific methodological details, tightening the link between the critical design elements, ensuring that the analysis plans are statistically valid, and ensuring that the study itself avoids stigmatising language. All of these issues fall within the typical expectations for a Stage 1 RR, and on this basis I am happy to invite a revision.

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

Response to Reviewer #1: Charlotte Pennington

Opening

This is a fantastic Stage 1 Registered Report which aims to (conceptually)* replicate Weiner et al. (1988) to assess whether physically-based stigmas are perceived as less controllable, more stable (irreversible), and associated with more pity, less anger, and more willingness to help compared to mental-behavioral stigmas. It also aims to extend this original study by assessing four additional stigmas that have become prevalent in the last decade and re-assess the original categorizations of stigma sources (origins) by asking participants whether they perceive these as mental-behavioral or physically-based. Overall, I think this is a well-planned study with important implications, and I fully endorse the recognition of replications, and their importance, within psychological science. The proposal meets the PCI RR guidelines for the scientific validity of the research questions, the logic, rationale and plausibility of the proposed hypotheses, the soundness of the proposed methodology and analysis pipeline, the clarity of study procedures and analysis pipeline to reduce undisclosed flexibility, and the consideration of sufficient outcome-neutral conditions (e.g., attention checks, MTurk data quality checks). Although my suggestions for revision below are extensive, they all relate to minor improvements to clarity and detail rather than anything substantial/major. I must say, I am impressed by the quality of this work and am inspired by the implementation of replications within teaching/pedagogy (this represents a students' proposed project and it is of extremely high quality).

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

> **General comments**
> **Before I outline my minor comments below, I just want to provide some general points to reflect upon:**
> **-   The authors suggest this is a direct replication, according to LeBel's guidelines, but in my opinion, it's a conceptual replication due to the addition of four new stigma labels, a change in the Vietnam War label (which I agree with, but my point remains), slight changes in the procedure and measuring participant's own cateogorisations of the stigma origins, and the analytical strategy (e.g., removing "order" because it showed no significant effect in the original, but the inclusion of an additional factor in the original will have changed its power and potentially the subsequence interpretations). This might be worth thinking about with regards to the original authors replying to this replication.**

We appreciate the note, this is a tricky somewhat subjective categorization, yet we feel it is important that this replication and extension project be differentiated from conceptual replications.

According to the criteria set out by LeBel et al. (2018), "close replications" differ from the original in terms of IV stimuli, DV stimuli, procedural details, physical setting, and contextual variables. As our current replication incurs changes only for the above design facets and not the rest, we categorize our study as a "close replication."

The IV stimuli were mostly similar to that of the original with the main change being the replacement of the Vietnam War syndrome label. We indeed added four stigmas as an extension, yet these should have little to no impact on the rest of the replication given the randomized order design. To argue that the effect only replicates if and only if the original items are included with no other changes would be counterproductive for the original authors and/or the community as the implications would be that the effect is very contextual and not at all generalizable, making it much less useful for our understanding of human psyche on stigma.

If our replication would be classified as conceptual, then it would be extremely difficult for any replication to be classified as direct. Any replication, especially of classic articles from decades ago, requires adjustments, if only for the time-based translations of the stimuli. It is also important to differentiate a conceptual replication from direct replication with extensions because direct replications share the same design and materials and are informative for knowledge accumulation and aggregation regarding the specific test of the phenomenon.

Our team has completed and published extensively using this criteria in dozens of replications in a way consistent with the one used here, and also in PCIRR with manuscripts that have already received in-principle acceptance (example 1 / example 2).

As a sidenote, we have been in touch with the original authors and shared our preprint. They kindly gave us some brief feedback. They did not have any issues with the categorization. Yet their main comment focused on suggesting we further clarify that the literature has moved from discussing sympathy rather than pity, which we noted in the revision in paragraph 2 and 3 under the section "Attribution-Affect-Help Judgment Model":

> "Weiner (1986) proposed that perceived controllability was associated with affective reactions of pity and anger, which were subsequently associated with helping behavior. As the conceptualization of pity has changed since Weiner et al. (1988), and the literature has moved from discussing sympathy rather than pity, we used "sympathy" instead of "pity" in descriptions of our current replication and in our survey materials. Stability is often implicated in pity but not in sympathy, as seen from observers' reactions to individuals with AIDS: pity is often a reaction to the terminal quality of AIDS, whereas sympathy is typically elicited by the perceived onset uncontrollability of the condition (Weiner, 1988, p. 126). Moreover, pity is now characterized as a condescending form of feeling sorry towards people perceived as pathetic or responsible for their misery (Geller, 2006). On the other hand, sympathy is viewed as an expression of concern or sorrow towards distress in an individual's life (Clark, 2010) and is more directly related to help-giving sentiments.

> Thus, when an individual in need is perceived to be in control of their plight, observers tend to feel less sympathetic and angrier toward these individuals and expect them to take responsibility for helping themselves out of the perceived self-inflicted predicament. On the contrary, observers tend to sympathize and want to help those who cannot control the causes leading to their misfortune. We summarized the motivational sequence in Figure 1."

We therefore changed instances of "pity" to "sympathy" in Figure 1, in descriptions or discussion relevant to our current replication and in the survey on Qualtrics.

We also noted these changes in paragraph 7 in the section "Deviations from the original" as follows:

> "Lastly, we made changes to the wording. "Sympathy" was used in our survey to replace "pity" as the former was more relevant to the current context. We also changed the wording of different stigma labels. As "Vietnam War syndrome…""

References:

- Clark, A. (2010). Empathy and Sympathy: Therapeutic Distinctions in Counseling. Journal of Mental Health Counseling, 32(2), 95–101. https://doi.org/10.17744/mehc.32.2.228n116thw397504

- Geller, J. D. (2006). Pity, Suffering, and Psychotherapy. American Journal of Psychotherapy, 60(2), 187–205.https://doi.org/10.1176/appi.psychotherapy.2006.60.2.187

- Weiner, B. (1988). An attributional analysis of changing reactions to persons with AIDS. In R. A. Berk (Ed.), *The social impact of AIDS in the U.S.* (pp.123-132). Cambridge, MA: Abt Books.

    -   The literature around public stigma towards mental and physical health conditions is vast, and there is a lot of work consistent with the theory and hypotheses proposed in the original work by Weiner et al. (1988). For example, reviews suggest that mental health disorders are more heavily stigmatised than physical health conditions, at both the explicit and implicit level. Do you think it's important to include some of this more general work to show how vast and, seemingly, consistent this work is? Or do you want to focus only on the replication? Perhaps this detail could be included in the part where you signify the importance and impact of this original study.

Thank you for the suggestion. Our focus has been on the target article, rather than a comprehensive review of the follow-up literature, yet we are happy to try and include a bit more from the main findings in that literature.

We added the following in paragraph 1 under the section "Choice of study for replication: Weiner et al. (1988)":

> "...A meta-analytic review by Rudolph et al. (2004) summarized 39 helping studies and showed general support for the theorized model and findings. Consistent with the theorized model and findings in Weiner et al. (1988), later studies showed that mental health or psychiatric disabilities are more heavily stigmatized than physical or other health conditions (Angermeyer, 2006; Corrigan et al., 2000; Thornicroft, 2006). To the best of our knowledge, there currently are no published independent direct pre-registered replications of the target article."

References:

- Angermeyer, M. C., & Dietrich, S. (2006). Public beliefs about and attitudes towards people with mental illness: a review of population studies. Acta Psychiatrica Scandinavica, 113(3), 163–179. https://doi.org/10.1111/j.1600-0447.2005.00699.x
- Corrigan, P. W., River, L. P., Lundin, R. K., Wasowski, K. U., Campion, J., Mathisen, J., Goldstein, H., Bergman, M., Gagnon, C., & Kubiak, M. A. (2000). Stigmatizing attributions about mental illness. *Journal of Community Psychology*, 28(1), 91–102.
- Thornicroft, G. (2006). *Shunned: Discrimination against People with Mental Illness* (1st ed.). Oxford University Press.

> **-   As an exploratory analysis, will you assess which conditions may be most-least stigmatised? For example, a new review by Kilian et al. (2021) suggests that alcohol use disorders are one of the most heavily stigmatised health conditions:**
> **-   Kilian, C., Manthey, J., Carr, S., Hanschmidt, F., Rehm, J., Speerforck, S., & Schomerus, G. (2021). Stigmatization of people with alcohol use disorders: An updated systematic review of population studies. Alcoholism, Clinical & Experimental Research, 45, 899-991.**
> **https://doi.org/10.1111/acer.14598**

Thank you for the suggestion. Following Dr. Rathbone's recommendation, we also added an exploratory direction to examine whether perceptions towards a particular category of stigma were more alterable under the influence of the controllability information. Thus, we added two sub-sections in "Exploratory directions": "Comparing the effects of controllability information between physically-based stigmas and mental-behavioral stigmas", followed by "Comparing individual stigmas" as follows:

> "Comparing the effects of controllability information between physically-based and mental-behavioral stigmas
>
> We planned to explore if perceptions towards a particular category of stigma were more alterable under the influence of the controllability information. Weiner et al. (1988) found that information portraying stigmas as controllable had a greater effect on judgments of physically-based stigmas. On the other hand, information portraying stigmas as uncontrollable had a greater effect on judgments of mental-behavioral stigmas. Thus, information inconsistent with the typical belief system most influenced observers' perceptions of stigmas in Weiner et al. (1988). However, the original study did not compare the effects of controllability information on physically-based stigmas versus the effects of uncontrollability information on mental-behavioral stigmas. We aimed to first explore if judgments towards physically-based stigmas were more alterable by controllability information and whether judgments towards mental-behavioral stigmas

were more alterable when uncontrollability information was presented. We then investigated if perceptions of a particular category of stigma were more alterable than the other.

Comparing individual stigmas

We planned to compare which conditions were most-least stigmatized by assessing which stigma was perceived as the most onset-controllable, elicited the most negative affective reactions and the least help-giving intent (and vice versa). We also planned to assess which stigmas were most-least influenced by controllability information. Weiner et al. (1988) did not specifically assess which stigmas were most-least influenced by components within the attribution-affect-help judgment model. However, the authors did note that "attributions for Alzheimer's disease, Vietnam War syndrome and child abuse were subject to relatively minor change." (p. 747) Thus, we aimed to first compare individual stigmas in the control condition, followed by comparing if perceptions on a particular stigma were more alterable than the other."

We changed the details of our data analysis plan accordingly. We added the description that our plan to run the control condition and the two information conditions together in the same ANOVA model allowed us to compare the effects of controllability information between the two types of stigmas in the first paragraph of "Exploratory analyses", under "Data analysis strategy":

"The target article conducted separate analyses for the control neutral condition and the contrast between the two information conditions, and as exploratory supplementary analyses we ran the three conditions together in the same ANOVA model. Such allowed us to compare the participants' perceptions of controllability across the control and experimental conditions. We could also further compare the effects of the manipulation between physically-based and mental behavioral stigmas across conditions."

We then addressed the new exploratory direction of comparing individual stigmas in paragraph 2 under "Exploratory analyses", under "Data analysis strategy":

"We further aimed to evaluate which conditions were the most-least stigmatized, as informed by the extremities in the ratings and the stability of ratings across conditions. In addition, we aimed to explore…"

As there is an increased focus on comparing perceptions of physically-based versus mental-behavioral stigmas in this revision, we decided to completely follow the original study's method and included the variable "stigma source" in conducting our correlation and regression analyses (see Table 2 in Weiner et al., 1988, p. 741 for the findings from correlation analyses based on Experiment 1 and 2; text on p. 741 and p.743 for the findings from the multiple regression analyses in Experiment 1 and Experiment 2, respectively). Specifically, we tested for the

hypothesis that "Physical stigmas, perceived onset uncontrollability, positive affective reactions, and help-giving tendencies are positively correlated" as per Weiner et al. (1988) (p. 741).

We first updated Table 2 and Table 3 by reporting Pearson's *r* from correlations with stigma source in the rows reporting the original findings in testing Hypothesis 3. We also reported the original standardized regression (β) coefficients between stigma source and help in the rows reporting the original findings in testing Hypothesis 4 in Table 2 and 3. In the supplementary, we added the relevant effect size and confidence interval calculations in the subsection "Effect size calculations of the original study effects" under "Analysis of the original article" (see (9) and (21))

We updated Hypothesis 6 and Hypothesis 7 in Table 5 by adding "Physical stigmas":

> "H6: Physical stigmas, perceived onset controllability, positive affective reactions, and help-giving tendencies are positively correlated.

> H7: Among the positive correlations between physical stigmas, perceived onset uncontrollability, positive affective reactions, and help-giving tendencies, the correlation between positive affective reactions and help-giving is the strongest."
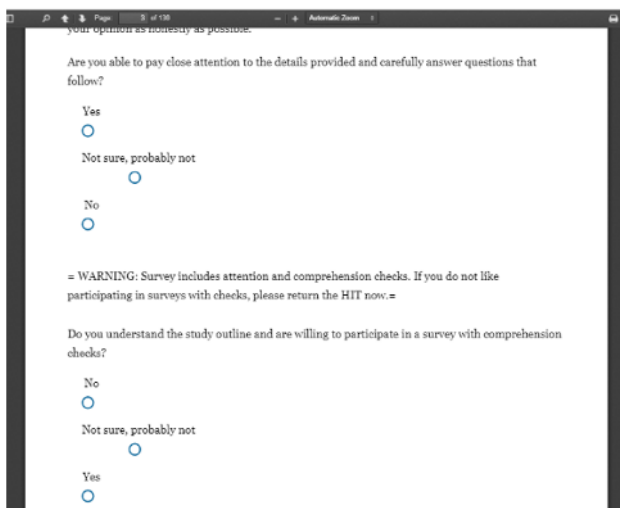
We updated Table 12 in "Results" by adding the variable "Physical stigmas" in the correlation and regression analyses.

To note, unfortunately, I could not initially access the original article by Weiner et al. (1988) because it was behind a paywall that my institution did not have access to. The Recommender – Chris Chambers – kindly sent this to me, but I think it's worth considering that many other readers might also not be able to access this, so it's important that the information presented in the current manuscript does not rely on having access to the Weiner et al. article. I note ambiguities in phrasing within my recommended revisions below.

OSF materials, code, and data
- The OSF Page includes simulated power analyses, the Qualtrics survey, and the simulated dataset and planned data analysis; everything is very clearly labelled and in relevant folders. I looked through the documents and they are clear and would aid an independent replication and reproducibility. I document any minor issues found below.
- In the screenshot below from the Qualtrics materials, should the second question now follow the same format as the first, with YES, NOT SURE, and NO, or is this an attention check in itself? (i.e., a participant might quickly click 'NO' in the second question thinking it is 'YES' because of the structure of the first question.

Yes, this is intentional, and serves partly as an attention check before embarking on the study. Indicating anything other than "yes" is taken as no consent and the participant is then asked to return the HIT. The randomized order of the items ensures attentiveness.

> **- In the power analysis, I noticed that alpha was set at .10. Is this correct? Should it be a = .05 or a = .01?**

It is possible that we have misunderstood the feedback, yet we believe that this comment may reflect a minor misunderstanding, since in our power analyses we set our alpha to 0.05: "pwr.anova.test(k = 4, f = 0.2698435, sig.level = 0.05, power = 0.95)"

If you are referring to calculating the effect size where we indicate a = 0.10, then this is only used for the calculation of the confidence intervals for the eta-squared.

In that we followed some of the best practices, and elaborate on this in our effect size guide (https://mgto.org/effectsizeguide) citing Lakens (2014). In the guide we write:

> "Normally, we calculate 95% confidence intervals (i.e., 95% of such intervals are expected to contain the true parameter value if we conduct an infinite number of identical studies). Nonetheless, for some effect sizes (e.g., eta-squared, partial eta-squared, R-squared), we calculate 90% confidence intervals. This is because η^2 is squared and always positive, and F-tests are one-sided. Reporting 95% CI for eta squared may result in situations in which the CI includes zero but the p-value falls below .05, whereas reporting 90% CI prevents such a problem. For further information regarding this issue, read Daniel Lakens blog on confidence intervals and Steiger (2004)."

This is a collaborative guide, so we would appreciate further feedback from you or the community on that if you believe this could be improved.

Reference:

- Lakens, D. (2014, June 07). Calculating confidence intervals for Cohen's d and eta-squared using SPSS, R, and Stata. Retrieved June 05, 2020, from http://daniellakens.blogspot.com/2014/06/calculating-confidence-intervals-for.html

  Title
  - Would this be clearer as "A replication and extension of Weiner et al. (1988)" (compared to "Replication and extensions of Weiner et al. (1988)".

Thank you for the suggestion. We hope that you do not mind us keeping the original title as to indicate clearly that we are including two extensions.

We would gladly revisit this if given clear editorial guidelines to do so.

Abstract
-   The findings sub-section mention "stigma source" and "emotional reactions" but these are fairly ambiguous given the prior information given – could the same terms be used as is currently presented in the first introductory sentences? I am confused by what "stigma source" is when reading the Abstract alone; could this be 'origins' instead (and you could provide the e.g., of mental-behavioural and physical-based again here).

Thank you for the suggestion. We changed the sentence as follows in the abstract:

"Our replication [failed to find/found] support for the original findings on the associations between stigma origin (i.e., mental-behavioral or physically-based), stability, emotional reactions (i.e., sympathy, liking, and anger) and willingness to help [summary effect sizes + CIs will be added here]."

Introduction
-   Please check APA or BPS guidelines for terminology, but I think the term "obese people" can be seen as stigmatising, and it is best to use the term "individuals with obesity".

Thank you for bringing this up. We changed all mentions of "obese people" to "individuals/people with obesity".

We noted changes in stigma labels in paragraph 7, under the section "Deviations from the original":

"Lastly, we made changes to the wording. "Sympathy" was used in our survey to replace "pity" as the former was more relevant to the current context. We also changed the wording of different stigma labels. As "Vietnam War syndrome" seemed to be outdated, we changed the stigma label to "Post-traumatic stress disorder (PTSD)," which we thought was more generalizable and fitting for current times. In addition, we used the terms "individuals/people with obesity" and "individuals/people with AIDS" to not perpetuate stigma in research.

-   "Stigmas are negative social reactions to perceived deviations from the norm (Jones et al., 1984)." There is a distinction between public and self-stigma; could you clarify here that this is a definition of Public stigma rather than capturing self-stigma.

Thank you for highlighting this. We updated the sentence in paragraph 2 of the section "Background":

"Public stigmas are negative social reactions to perceived deviations from the norm (Jones et al., 1984) and are connotative of negative outcomes, or undesirable, debilitating conditions (Goffman, 1963; Weiner et al., 1988)."

References:

- Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity.* New York: Simon & Schuster
- Jones, E. E., Farina, A., Hastorf, A. H., Markus, H., Miller, D. T., & Scott, R. A. (1984). Social stigma. San Francisco: Freeman

> **-   More information could be given about the Weiner et al. (1988) study on Page 8 of the Introduction, such as examples of "stigmas with somatic origins" and "mental-behavioral origins". I was surprised to not see more detail here given that this is the study under replication.**

Thank you for the feedback.

We added a third paragraph under "Background":

"In Weiner et al. (1988), the list of physically-based stigmas included attributes typically thought to be "determined" by genetic and environmental forces beyond personal control, such as blindness and Alzheimer's disease. Their list of mental-behavioral stigmas included conditions typically associated with deviant acts that arise from an individual's intentions, such as obesity, which is commonly thought to be caused by laziness (Jones & Davis, 1965; Maselli & Altrocchi, 1969; DeJong, 1980)."

References:

- DeJong, W. (1980). The Stigma of Obesity: The Consequences of Naive Assumptions Concerning the Causes of Physical Deviance. Journal of Health and Social Behavior, 21(1), 75. https://doi.org/10.2307/2136696
- Jones, E. E., & Davis, K. E. (1965). From Acts To Dispositions The Attribution Process In Person Perception. Advances in Experimental Social Psychology, 219–266. https://doi.org/10.1016/s0065-2601(08)60107-0
- Maselli, M. D., & Altrocchi, J. (1969). Attribution of intent. Psychological Bulletin, 71(6), 445–454. https://doi.org/10.1037/h0027348

> **-   Page 9, remove the term "suggested" from "followed by our suggested extensions" if this article receives IPA.**

Thank you. We removed the word "suggested" in paragraph 5 under the section 'Background'.

> **-   Page 9, what is "attributional search"?**

We understand the term "attributional search" may have been unclear, and we made the following amendment to paragraph 1 under the section "Attribution-Affect-Help Judgment Model":

> "Attribution theory posits that people attribute causes to observed events or behavior, and the perceived causality of an outcome induces various affective reactions and behaviors (Weiner, 1986)."

> **-   Please correct the references on Page 12: "importance of reproducibility and replicability in psychological science (Open Science Framework & Lakens, 2012)". Is this Open Science Collaboration, 2012, 2015? And another reference to Lakens?**

Thank you for pointing this out. The correct reference is "Open Science Collaboration, 2012" in paragraph 4 of "Choice of study for replication: Weiner et al. (1988)". We deleted the reference to Lakens.

> **Table 1**
> **-   Can you further explain the issue of 'randomization'– as it currently reads, I cannot see what the specific issue is (and the table is meant to provide an overview of these issues).**

The original authors did not fully randomize the display of the items. Of all possible combinations of the display of items, they fixed those into two specific orders and then tested order effects by contrasting these specific two. Choosing two specific sets with a determined order has implications for understanding and controlling the impact of stigma order display. This was understandable in 1988 with the administration of surveys using paper and pencil, but nowadays we can do a full randomization of all stigma items.

We therefore changed to a fully randomized order.

We tried to improve on explaining this by changing the description of 'Randomization' in Table 1 as follows:

> The original article chose two fixed stigma display orders out of all possible stigma display combinations, thereby contrasting specific stigmas and not fully addressing order. [...] We randomized the display order of the stigmas.

> **Can you do the same for "Comparison between the control and experimental condition in Experiment 2". In simple terms, it would be good**

> **to see the issues briefly described in this table so it is clear (just like with the "scale of meansurement" and "nature of stigma" inclusions).**

To try and better clarify the issue we amended the 'Comparison between the control and experimental condition in Experiment 2' in Table 1:

> "The design in Experiment 2 manipulated controllability information in three conditions: 1) neutral control condition with no information on stigma onset controllability, 2) stigma controllability information and 3) stigma uncontrollability information to examine causal relationships between perceived controllability and affective reactions, along with help-giving behavior.
>
> However, the researchers only contrasted the two information conditions (i.e., conditions #2 and #3) in testing associations between variables within the attribution-affect-help judgment model.
>
> We discuss this in more detail in the methods section."

> **Replication target's expeimental design, hypotheses and findings**
> **-   Page 13, first paragraph: What do you mean by "adding one generalized hypothesis for causality" – can you expand on this? Did you also test this generalized hypoithesis? Can you also expand on what is meant by "instrumentality" or "intervention techniques" in this paragraph; this assumes prior knowledge of the reader/the terms are quite ambiguous.**

We appreciate the opportunity to clarify, and we understand that the term "generalized hypothesis" can be confusing. We deleted the word "generalized" in "Replication target's experimental design, hypotheses, and findings."

In the original study, the purpose of Experiment 2 is to establish whether perceived uncontrollability causes positive affective reactions (i.e., liking, pity) and a higher likelihood of help-giving judgments by adding the manipulation of controllability information. As we are replicating Experiment 2 in our current study, we are testing the hypothesis that perceived controllability causes different affective reactions and help-giving judgments.

In sum, we made the following changes to "Replication target's experimental design, hypotheses, and findings":

> "Experiment 2 tested for the same core hypotheses as Experiment 1, adding one hypothesis to test for the presumed causal relations between perceived controllability and the control-related variables (i.e., affective reactions and help-giving judgment). Our replications focused on the core hypotheses relating to the base theory. We did not seek

to test associations with the instrumentality of intervention techniques, i.e., the usefulness of different methods to improve life satisfaction of stigmatized individuals."

To remove duplications regarding the instrumentality of interventions in paragraph 1 under "Experiment 1: Baseline", we made the following changes:

"In Experiment 1, Weiner et al. (1988) recruited 59 American college students and asked them to rate ten stigmas on the following dimensions: the perceived controllability of stigma onset (indexed by responsibility and blame), affective reactions of liking, pity, and anger towards individuals with a particular stigma, the perceived stability of the stigmas (indexed by changeability) and the perceived instrumentality of five intervention techniques (technical job training, professional job training, welfare, medical treatment, and psychotherapy)."

Experiment 1: Baseline
-   Does Weiner et al. explain how the different stigmas were cateogorised? For example, I would say that obesity is a physical health stigma, but Weiner et al. describe this as a mental behavioral stigma. Further information to clarify how these stigmas were cateogorised into the two groups would be helpful. Reading on, I see that your extension covers this by assessing "Participants' categorization of stigmas". Perhaps you can use my example here within this paragraph on Page 17 to show how their categorizations might not be have been fully accurate (and there is literature that supports obesity as a physical health condition).

This is a good point. Weiner et al. (1988) did not explain how the different stigmas were categorized, and this is one of the reasons why we felt it was important to assess that participants and others would categorize stigmas in a similar way. We used this specific example to make our argument regarding the extension even stronger in the sub-section "Participants' categorizations of stigmas" under the section "Extensions":

"In the original study, the hypotheses revolved around classifying stigmas by source, yet such categorizations were solely based on the researchers' classifications. Participants might categorize the stigmas differently. For example, some may perceive obesity to be caused by a lack of will. Others may be aware that obesity is a complicated medical condition caused by the interplay of genetic, environmental, and metabolic factors (Upadhyay et al., 2018). Thus, we thought it imperative to assess participants' physical versus mental-behavioral origin classifications to ensure these match with the researchers' assumptions."

References:

- Upadhyay, J., Farr, O., Perakakis, N., Ghaly, W., & Mantzoros, C. (2018). Obesity as a Disease. Medical Clinics of North America, 102(1), 13–33. https://doi.org/10.1016/j.mcna.2017.08.004

> **Four new current stigmas**
> **-   "Since the original article's publication, the coverage and prevalence of different physiological and psychological conditions have changed considerably." - can you support this statement with a reference?**

Great suggestion, thank you. We added several references to support the statement as follows to paragraph 2 of the sub-section "Four new current stigmas" under "Extensions":

> "Since the original article's publication, the coverage and prevalence of different physiological and psychological conditions have changed considerably. To illustrate, there was an increase in newspaper coverage of Alzheimer's disease between 1988 and 1997 (Adelman & Verbrugge, 2000) and a decrease in the death rates from heart disease between 2000 to 2012 (Mozaffarian et al., 2015). In terms of psychological disorders, the volume of news coverage of mental illness appears to trend downward from 1995 to 2014 (McGinty et al., 2016). At the same time, there have been increasingly more public figures disclosing their stories of mental disorders. People have also been more open to admitting their psychotherapy experiences (Hinshaw, 2009, p. ix)."

References:

- Adelman, R. C., & Verbrugge, L. M. (2000). Death Makes News: The Social Impact of Disease on Newspaper Coverage. Journal of Health and Social Behavior, 41(3), 347. https://doi.org/10.2307/2676325
- McGinty, E. E., Kennedy-Hendricks, A., Choksy, S., & Barry, C. L. (2016). Trends In News Media Coverage Of Mental Illness In The United States: 1995–2014. Health Affairs, 35(6), 1121–1129. https://doi.org/10.1377/hlthaff.2016.0011
- Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., de Ferranti, S., Després, J. P., Fullerton, H. J., Howard, V. J., Huffman, M. D., Judd, S. E., Kissela, B. M., Lackland, D. T., Lichtman, J. H., Lisabeth, L. D., Liu, S., Mackey, R. H., Matchar, D. B., . . . Turner, M. B. (2015). Heart Disease and Stroke Statistics—2015 Update. Circulation, 131(4). https://doi.org/10.1161/cir.0000000000000152

> **Rethinking the "child abuse' stigma item**
> **-   "The described stigma regarding "child abuse" differed from other stigmas in that individuals suffering from the other stigmas were commonly described as the victims, yet the stigma "child abuse" was about the person who perpetrated the abuse rather than its victim".**

> **Can you provide the specific statement from Weiner et al. which shows that this was about the perpetrator and not the victim; this will solidify your point.**

We made the following changes to the sub-section "Rethinking the 'child abuse' stigma item" under the section "Exploratory directions":

> "The described stigma regarding "child abuse" differed from other stigmas in that individuals suffering from the other stigmas were commonly described as the victims. However, "child abuse" was about the person who also perpetrated abuse rather than only its victim. Such is indicated by the controllability information description in Weiner et al. (1988): "5. Child abuse - Had been an abused child, experiencing severe stress and near a nervous breakdown; intentionally abused own child" (p.742). Weiner et al. (1988) also mentioned that "child abuse cannot be transformed from a sin into a sickness" (p. 747). Thus, child abuse seems to be a puzzling choice given the factors of focus (such as sympathy and help)."

We made the following changes to the description of the methodological issue "Nature of stigma" in Table 1:

> "We identified one stigma in the original list - "child abuse" - described differently from the others. Individuals suffering from the other nine stigmas in the original article were commonly described as victims. However, "child abuse" was about the person who also perpetrated abuse rather than only its victim.
>
> We discuss this in more detail in the methods section. "

> **Table 5 (updated to Table 4 in Version 2 of the manuscript)**
> **In my opinion, Table 5 needs the most work to aid reader understanding.**
> **-   Why are there no relevant hypotheses proposed in Table 1 for the first row of Alzheimer's disease, blindness, cancer, paraplegia, child abuse, and drug abuse. It would be best to specifically state these here for completeness.**

There were baseline-relevant replication hypotheses in another table.

To link the two and be more clear to readers who might be wondering the same thing, we added "All original hypotheses: H1 - H6 (see Table 3)" in the first row of Table 4 for clarity.

> **-   The sentences following the first of the "stigmas in the original study that will not replicate as well in the current replication" are confusing and could be reworded to aid clarity;**
> **here you state that H1 was "stigmas having a mental-behavioral origin are perceived as more onset-controllable than stigmas having a somatic genesis", and then go onto state "heart disease, which is categorized as a physically-based stigma, will be perceived as more onset-controllable in our current replication. US Americans may perceive heart disease to be a mental-behavioral stigma nowadays (Waters et al., 2014). Such current perceptions on heart disease will contradict with original findings, which indicated that heart disease received low ratings on perceived controllability".**
> **Paradoxically, then, such a proposed finding (heart disease will have higher ratings on perceived controllability) would actually support the original hypothesis that stigmas having a mental-behavioral origin are perceived as more onset-controllable because perceptions of it have changed from being physically-based to mental-behavioral based. In simple terms, the contradiction is with the categorisation of this condition (from physical to mental-behavioral), rather than contradicting the prediction ("stigmas having a mental-behavioral original are perceived as more onset controllable").**

Thank you for highlighting this, and we agree that our initial phrasing can be improved. We deleted the sub-headings "Stigmas in the original study that will replicate well in the current replication" and "Stigmas in the original study that will not replicate as well in the current replication" in Table 4. We clarified the differences between conditions under the "Condition" column.

For the specific condition in which the contradiction is with the stigma categorization and not the hypothesis, we concluded it as "3. The categorization of the stigma(s) in the current replication

will be different from that in the original" within the "Condition" column. Under "Relevant hypotheses and Descriptions", we made the following changes:

> "Americans may perceive heart disease to be a mental-behavioral stigma nowadays (Waters et al., 2014). Thus, heart disease, categorized as a physically-based stigma (following the original's categorization) will be perceived as more onset-controllable in the current replication.
> Such current perceptions of heart disease will contradict the original conceptualization of heart disease and the original findings, which indicated that heart disease received low ratings on perceived controllability. The conceptualization of heart disease as a mental-behavioral stigma, leading to higher ratings on perceived controllability, confirms the original hypothesis, but does not match the original's categorization of heart disease."

> **-   Condition 2 is presented under "Stigmas in the original study that will not replicate as well in the current replication". However, it seems that, actually, you will support the original findings ("Original hypotheses will be disconfirmed again in the current replication.").**

Thank you for pointing this out. For the specific condition in which we predict current findings will support original findings and disconfirm hypotheses, we concluded it as follows in Table 4, under the "Condition" column:

> "2. The findings on the stigma(s) in the current replication will be similar to the original findings. However, original hypotheses on the stigma(s) will be disconfirmed again in the current replication."

> **Method: Pre-registration and open-science**
> **-   "We pre-registered the experiment on the Open Science Framework (OSF) and data collection was launched later that week. Pre-registrations, power analyses, and all materials used in these experiments are available in the supplementary materials. We provided all materials, data, code, and pre-registration on the OSF: https://osf.io/gwcbt/. We provided additional openscience details and disclosures in the supplementary materials under "Open Science disclosures" sub-section".**
> **Can I clarify that this is written in past tense to aid the completion of the Stage 2 manuscript, but that data collection has not yet begun? Also, note that a space is needed between open and science.**

Yes. We confirm that this section is written in past tense to simulate what the manuscript would look like after the completion of the Stage 2 manuscript, and that data collection has not yet begun. We added a space between open and science.

We also posted clarifications at the beginning of the 'Method' section that:

"For the purpose of the simulation, we wrote things in past tense, but no pre-registration or data collection took place yet.".

Power analysis/sample size:
-   "To demonstrate what the results would look like after data collection, we simulated a dataset of 1000 participants using Qualtrics, which we will later update with the real data and our sample of ~800".
Why did you base this on a sample of 1000 participants not your plan of 800 participants?

There was no specific reason for this, and that simulated sample has no special meaning. The purpose of the data simulation was to help show what the final report will look like after data collection.

Thank you for pointing this out, and we understand the difference in sample size can be confusing. We re-simulated the dataset based on 800 participants. We changed the sentence in paragraph 2 in the section "Power analysis" as follows:

"To demonstrate what the results would look like after data collection, we simulated a dataset of 800 participants using Qualtrics, which we will later update with the real data."

> **- "A sensitivity analysis indicated that a sample of 800 would allow the detection of f = 0.14 (groups = 3, df = 1) and d = 0.29 (independent samples with 266 participants in each condition; both 95% power, alpha = 5%, one-tail)". Sorry to be pedantic, but 266 * 3 = 798 and not 800.**

We changed the sentence in the paragraph 2 in the section "Power analysis" as follows:

> "A sensitivity analysis indicated that a sample of 798 (rounded up to 800) would allow the detection of $f = 0.14$ (groups = 3, df = 1) and $d = 0.29$ (independent samples with 266 participants in each condition, both 95% power, alpha = 5%, one-tail), effects much weaker than any of the effects reported in the target article."

> **- Do you plan for equal amounts of participants in each of the three conditions? Some text is confusing because it implies two conditions, when there are three, e.g. "In the simulation the no-information condition and the experimental condition each comprised 500 participants as samples, but in the actual data collection the three conditions will have a fairly even split".**

We appreciate the opportunity to clarify. We plan for a fairly even split of participants in each of the three conditions in the actual data collection process, i.e., 1) the neutral control condition with no information on stigma onset controllability, 2) stigma controllability information and 3) stigma uncontrollability information.

We rectified all text that implied there were only two conditions.

We also updated indications of sample size in the 'Results' section to demonstrate that there is a fairly even split of the 800 participants among the three conditions. Based on the simulated data, there were 267 participants in the control (i.e., no info) condition, 266 participants in the controllability info condition, and 267 participants in the uncontrollability info condition, $N = 267$.

> **Table 7 (updated to Table 6 in Version 2 of the manuscript)**
> **- Table 7 suggests that the gender was not reported in Weismann et al. (1988), but perhaps you could state what they did report: "149 male and female UCLA students and 171 male and female University of Manitoba students". I understand that they did not split them, but it would be good to have this information.**

Thank you for the suggestion. We included the original study's reported descriptive statistics on participants' gender in Table 6.

Design and procedure
-   You ask for specific feedback on the following: "We followed the target's decision to try and balance the controllability/uncontrollability conditions by having the two information conditions evenly mixing controllability and uncontrollability stigma information.
[Sidenote: After contemplating this design for long, we admit to struggling to understand this decision and having strong reservations regarding their design, we believe this should have been randomized, yet decided to try and replicate their design as is. More information is provided in the "deviations" section below. It is possible that we misunderstood or overlooked something important here, and so we would appreciate reviewers feedback on this point]."
Feedback: I think they evenly mixed controllability and uncontrollability so that more of one (e.g., controllability) didn't bias the responses (i.e., participants would have viewed a greater proportion of one condition, which might have biased their responses/confounded stigma ratings). I have spent the past hour reading through, reflecting, and evaluating their Methods on page 727 and I believe you have got this replication method/procedure correct (however, I note that their Method section of Experiment 2 is very complex and contradictory in parts).

Thank you for the comments and the support.

Given Dr. Joanne Rathbone's suggestion, we decided to change the experimental design to make it truly a between-subject design. Participants would either 1) receive no information, 2) receive information that indicates conditions were onset controllable, or 3) conditions were onset uncontrollable.

Please refer to our response to Dr. Rathbone's comment on 'Design and procedure' below for more details.

-   Sorry if I've missed it, but whereabouts in the Procedure do you measure participant's own categorizations of mental-behavioral or physical? I can see that this is asked at the start of the Qualtrics questionnaire (by looking at the questionnaire), but I don't seem to be able to find it within the written procedure?
Moreover, do you think that by presenting this first, this might weaken the manipulation of mental-behavioral vs. physical stigmas? (i.e. people's pre-existing beliefs influence the stigmas, and this being salient nullifies the experimental manipulation?).

Thank you for pointing this out, and we agree that asking participants' categorizations of stigma origin first might impact the manipulation.

We therefore grouped the questions on categorizing the fourteen stigmas by source together and presented them at the end of the survey, after participants had rated the stigmas on the other dimensions in Qualtrics.

Following changes in the display sequence of the survey items, we updated the description of the procedure in paragraph 3 under the section 'Design and procedure' as follows:

> "To ensure attentiveness, we first introduced participants with an attention check asking them to indicate which stigma they were currently rating. Participants then responded to eight questions per stigma: 1) perceived stability (changeability), 2) perceived onset controllability for a stigma: responsibility and blame, 3) three questions on affective reactions towards perceived stigma onset controllability: sympathy, anger, liking, and 4) two questions on help-giving tendencies: charitable donation and assistance. Questions on categorizing the fourteen stigmas by source were grouped and presented after participants had responded to the above eight questions per stigma. The presenting order of the fourteen questions were randomized. Such arrangement was to ensure the robustness of the manipulation in verifying if perceived sources of stigmas influence perceptions of stigmas, as hypothesized in Weiner et al. (1988). If the manipulation was presented before the above eight questions, it might reinforce participants' pre-existing beliefs that influence their perceptions of stigmas, and such being salient would nullify the experimental manipulation. Except for perceived stigma origin, all dependent variables were on a 9-point scale anchored at the extremes (e.g., 0 - *Not responsible at all;* 8 - *Entirely responsible*)."

> **Data Analysis Strategy:**
> -   **This is clear and I agree with all analytical decisions here – the addition of a 3-way ANOVA is a good extension.**

Thank you. Unless we misunderstood something here, we believe what you meant to write was "3-conditions ANOVA"?

In light of the changes requested by Dr. Rathbone to a fully between-design we updated our data analysis accordingly. Please see our reply on that point below.

> **Results**
> -   **Where are the simulated results for H1 and H2? Am I missing something here? For example, H1: Stigmas having a mental-behavioral origin are perceived as more onset-controllable than stigmas having a somatic genesis.**

We appreciate the opportunity to clarify. We would like to point out that H1 and H2 are the extension hypotheses according to Table 5. H1a - g state that the ten physically-based stigmas in Weiner et al. (1988) and diabetes and stroke (the newly added stigmas) are

perceived as physically based stigmas, whereas H2a - g state that the ten mental-behavioral stigmas in Weiner et al. (1988) and major depressive disorder and anxiety disorder (the newly added stigmas) are perceived as mental-behavioral stigmas. The results for H1a - g and H2a - g will be presented in Table 9.

 The replication hypotheses, H3 - H5a - e, compare physically based and mental-behavioral stigmas in terms of perceived stability, onset-controllability, and affective reactions and help-giving judgments elicited from observers. The H1 you were inquiring about is actually H4 according to Table 5: Compared to physically based stigmas, mental-behavioral stigmas are perceived to be more onset-controllable. H5a - e are hypotheses comparing observers' feelings of sympathy, anger, liking, and tendencies to provide personal assistance and charitable donations between the two stigma sources. The results for H4 and H5a - e will be presented in Table 11.

The simulated results for H1 and H2 in Table 9 (the extension hypotheses) and for H4 and H5a - e (the replication hypotheses) in Table 11 are not presented fully as we meant the tables as a preview of what the results might look like after the data collection and the analyses we ran. Given that all data is random noise, we did not see the point of completing all the tables and only completed several rows for illustration purposes.

Response to Reviewer #2: Dr. Joanne Rathbone

Opening
The report proposes a replication of Experiment 2 from Weiner et al. (1988). It was nice to see that the report goes beyond simply replicating the original study, using the opportunity to address limitations of the original study and extend upon it. The report clearly outlines what aspects will be direct replications, where the planned replication study deviates from the original and why, what analyses are exploratory, and how the replication study will also build upon the original study.

Thank you for the opening note.

However, I did pick up on some inconsistencies between the aims, hypotheses, and planned methods which need to be resolved, and some areas where the Introduction could be strengthened. I hope that my comments below will be useful for the authors.
General comments

1.   The term "obese" is now considered to be stigmatising language, and it's important that we don't perpetuate stigma in our research. Please use person-first language when referring to members of this group i.e., "people with obesity". The authors may also want to consider rephrasing "AIDS" to "people with HIV", particularly given that HIV is now a very manageable virus and most people with HIV do not develop AIDS.

Thank you for bringing this up. We changed all instances of "obese people" to "individuals/people with obesity". We plan to use the term "individuals/people with AIDS, as changing to "HIV" changes the meaning of the label.

We noted these changes in paragraph 7 in the section "Deviations from the original" as follows:
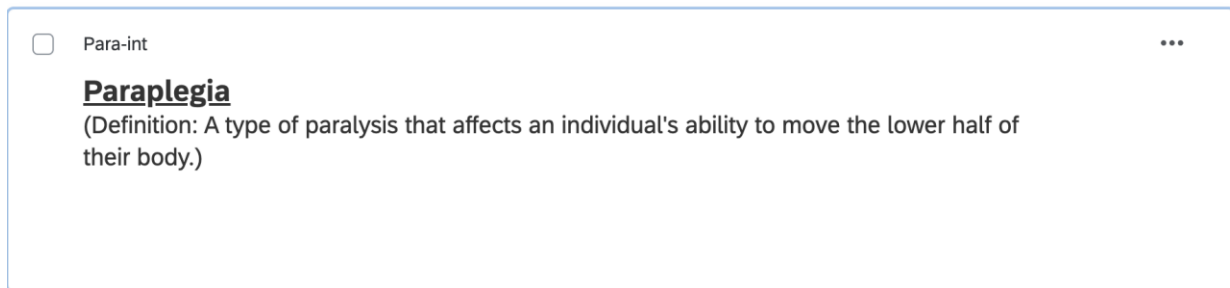
"Lastly, we made changes to the wording. "Sympathy" was used in our survey to replace"pity" as the former was more relevant to the current context. We also changed the wording of different stigma labels. As "Vietnam War syndrome" seems to be outdated, we changed the stigma label to "Post-traumatic stress disorder (PTSD)," which we thought was more generalizable and fitting for current times. In addition, we used the terms "individuals/people with obesity" and 'individuals/people with AIDS" to not perpetuate stigma in research."

2. There are several grammar and spelling errors throughout the report. Please check it over carefully and correct where needed.

Thank you for pointing this out. We tried to address all grammar and spelling errors in our revision.

7. I recommend providing a definition of paraplegia in the survey for participants who may not be familiar with the term.

Excellent suggestion, we agree. We added a definition of paraplegia in the survey as follows:

☐ Para-int   ···

Paraplegia
(Definition: A type of paralysis that affects an individual's ability to move the lower half of their body.)

Choice of study for replication: Weiner et al. (1988)
3. The phrase "phenomenon updating extensions" on page 10 is a bit clunky. I suggest revising to something with less jargon, like "the potential for extensions to the model" or similar. That sentence also can't be a paragraph on its own – I recommend combining it with the next paragraph that starts with "The article has had…".

Thank you for the suggestion. We combined the sentence with the next paragraph.

We made the following changes to paragraph 1 under the section "Choice of study for replication: Weiner et al. (1988)":

"We chose the article by Weiner et al. (1988) for replication based on several factors: impact, topic importance, absence of direct replications, and the potential for adding extensions to test the robustness of the phenomenon. First, the article has had much impact on scholarly research in social cognition, and judgment and decision-making…"

5.   In paragraph 3 of page 11 the authors refer to the increasing prevalence of mental health conditions in their discussion of the implications and importance of replication. But the stigmas investigated in this study are not all mental health conditions, and so this sentence seems a bit out of place, particularly in the absence of discussion regarding the prevalence of the non- mental health conditions investigated (e.g., AIDS or HIV, diabetes).

Thank you for the feedback, we agree.

We amended the manuscript to be less about mental health and to refer to stigmas more broadly in paragraph 3 under the section "Choice of study for replication: Weiner et al. (1988)":

"The original study had practical implications for healthcare and social welfare. Weiner et al. (1988) noted that associations between the perceived controllability of stigma onset and help-giving tendencies might lead to difficulties in soliciting public funding and support for mental health or behavioral issues and stigmatized populations. Several studies demonstrated that the perceived causality of stigmas was associated with fundraising outcomes (Berkowitz, 1975), patients' treatment participation and adherence (Corrigan, 2004), and clinicians' attitudes towards patients (Boekel et al., 2013). As the coverage and prevalence of the stigmas included in Weiner et al. (1988) may have changed since publication, revisiting the study allowed us to examine current public attitudes and social support for stigmatized individuals."

Table 1
4.   I know Table 1 provides a summary, but could the authors please provide at least 1 example of a reproducibility issue and how it will be addressed in the main text (page 11)?

Thank you for the suggestion. We provided an example of a reproducibility issue in paragraph 2 of the section "Choice of study for replication: Weiner et al. (1988)":

"We summarized the main reproducibility issues and our decisions to address them in Table 1. To illustrate, one of the issues we identified involved the scale of measurement. The original article stated that ratings for the dependent variables were made on a 9-point scale (meaning 0 to 8 or 1 to 9). However, some mean values for the original ratings on the 10 stigmas were less than 1 and more than 8: mean ratings for the perceived responsibility and assistance on blindness in Experiment 1 were 0.9 and 8.5, respectively, in Table 1 (p. 740). In Experiment 2, the mean ratings on perceived responsibility and pity for AIDS were 0.6 and 8.2, respectively, as shown in Table 5 (p. 746). Such indicated either a reporting error or that the original study used a 10-point scale that ranged from 0 to 9."

> **Table 6 (updated to Table 5 in Version 2 of the manuscript)**
> **6.   There doesn't seem to be a hypothesised controllability condition x stigma source interaction in Table 6.**
> **Would you expect the effect of onset controllability information to be more pronounced for physical stigmas than mental-behavioural stigmas? Hypothesis 8 is also very vague, please be more specific as to how you predict perceived controllability to change affective reactions and helping judgements.**

We added the comparison of the effects of controllability information between the two types of stigmas under the section "Exploratory directions", after the section "Comparison between ratings in the control and experimental condition".

Following Dr. Pennington's suggestion, we also added an exploratory direction to assess which stigma were perceived to be the most onset-controllable, elicited the most negative affective reactions, and the least help-giving intent (and vice versa). Please refer to our response to Dr. Pennington's suggestion on assessing which conditions may be most-least stigmatized above for more details on these two exploratory directions.

We changed the wording of Hypothesis 8 in Table 5 for clarity:

> "Higher perceived controllability of a stigma leads to increased anger and decreased liking and sympathy towards the stigmatized individual, and decreased tendencies to provide personal assistance and donations."

> **Table 9 (updated to Table 8 in Version 2 of the manuscript)**
> **8.   Table 9 – the original study also included Canadians. The details of deviation for population needs to be revised so that the similarities and differences in population between the two studies are clearer. An explanation for how contextual variables are different in the replication relative to the original study also appears to be missing.**

Thank you for catching that. Yes, you are right.

We made the following changes to "Population (e.g., age)" in Table 8:

> "The original study recruited college students from UCLA in the United States and the University of Manitoba in Canada in return for class credit in introductory psychology. We sampled US American respondents."

We provided details of deviation for the design facet "Contextual variables":

> The original study was conducted in the 1980s, whereas we conducted our replication study in 2022.

Power analysis

9.   I think the sensitivity analysis should be done for a sample of 600, not 800, as 800 is only being collected to ensure that a minimum of 600 participants are retained once exclusions have been made through the data cleaning process.

Sensitivity analyses should be made on the final sample size. We aimed for 800, regardless of exclusions, and we do not know at the moment what the exclusions would be and what the final sample will be. We plan to update the sensitivity analysis after data collection is completed. Sensitivity analysis is better suited for the final sample, rather than the targeted sample.

We added a side note in paragraph 2 of the subsection "Power analysis" as follows:

"[Sidenote: The sensitivity analysis will be updated after data collection is completed.]"

Participants

10.   The simulated data includes an age range of 0-100, which isn't realistic. And the authors are not intending to recruit a sample of 1000. Is this just a default in Qualtrics? I am not very familiar with data simulation in Qualtrics so this may not be correct, but I would have thought that the authors would want the simulated data to be as similar as possible to the data they intend to collect (i.e., age range, sample size etc.)?

The purpose of the data simulation is to help show what the final report will look like after data collection, the actual numbers from random noise have no meaning, and in this specific case we are not using age for any of the analyses.

We simulated using Qualtrics for the age range allowed as text entry (0-100). We would rather not restrict this Qualtrics text entry field further.

Design and procedure

11.   I share the authors' concern regarding the design used in the original study. Table 8 presents the controllability information IV as a 3-level between-participants variable. However this is not accurate. If the original design is retained, the control condition (and the participants assigned to it) are analysed separately to the participants in the two controllability information conditions in a one-way ANOVA testing the effect of stigma origin on outcomes when no information about onset controllability is provided (like the authors do in Table 12). The two experimental conditions are then analysed as a 2-level variable. If participants in both experimental conditions receive information about controllability and uncontrollability of onset, then this variable becomes a 2-level within-participants variable. The analysis is then a two-way within- participants ANOVA (not mixed) testing the effect of onset controllability information and stigma origin on outcomes. True comparisons between the outcomes from these separate ANOVAs cannot be made, so your first exploratory direction (page 18) is not achievable.

I am not convinced by the original paper's justification for providing information about both onset controllability and uncontrollability in both experimental conditions, particularly given that this information pertains to individual cases, not to the health condition as a whole. Seen as this replication is already correcting other limitations and errors in the original paper, I would recommend changing the design so that the controllability information IV is truly between- participants (i.e., participants either receive no information, they receive information that indicates all conditions were onset controllable, or all conditions were onset uncontrollable). If the authors choose to stick with the original design, Table 8 needs to be corrected so that the IV is not shown as a 3-level between-participants variable, and then potential limitations should be discussed in stage 2 and perhaps also mentioned in the methods section (e.g., original design means that comparisons between control condition and treatment conditions can't be made or should be made with caution).

Thank you for the comments, and we agree with the suggested changes to make it truly a between-subject design.

We changed our design to fit with your suggestion. Participants now either 1) receive no information, 2) receive information that indicates conditions were onset controllable, or 3) conditions were onset uncontrollable.

We deleted the second paragraph in version 1 of the manuscript within the section "Design and procedure", and made the following changes to paragraph 1:

> "We set up the survey online using Qualtrics, and participants were first asked to complete the consent form and pass verifications checks. Our replication experimental design manipulated controllability information with three outcomes: 1) neutral control condition with no information on stigma onset controllability, 2) stigma controllability information, and 3) stigma uncontrollability information. Participants were randomly assigned to one of the three conditions. The order of dependent variables…"

We also deleted the row in Table 7 detailing the display of stigmas in each condition.

In "Deviations from the original", we added descriptions of the change to a between-subject design before describing the simplification of the questionnaire format, as follows:

> "In the original study's Experiment 2, the authors manipulated controllability information and created three outcomes: i) neutral control condition with no information on stigma onset controllability, ii) stigma controllability information and iii) stigma uncontrollability information. Thus, researchers could conduct a between-subject comparison and contrast the three conditions against one another in testing the effects of controllability information on the participants' perceptions of stigmas.
>
> However, the original study balanced the controllability/uncontrollability conditions by evenly mixing controllability and uncontrollability stigma information in the two information conditions. This meant that in practice, participants were randomly assigned to one of the following three conditions: 1) neutral control condition, 2) seven stigmas described as controllable and the other seven described as uncontrollable, and 3) the stigmas described in #2 as controllable described as uncontrollable and vice versa. Such created a within-subjects design in which participants in the experimental condition were exposed to two outcomes from manipulation, i.e., stigma controllability information and stigma uncontrollability information. Stigma ratings in the control condition were therefore analyzed separately from the two experimental conditions with controllability information. In sum, there was no true comparison between the control condition and experimental conditions, as the display of the stigmas in each condition did not reflect the between-subjects design created from manipulating controllability information.
>
> We improved the experimental design per reviewer feedback to test for the causal effects of controllability information on perceptions of stigmas. In our study, the IV controllability information was truly between-subjects (i.e., participants either received no information, they received information that depicted all conditions as onset controllable, or information that depicted all conditions as onset uncontrollable).

The authors of the target article generated twelve varying questionnaire formats….”

We made the following changes to the details of deviation for the facet “Procedural details” in Table 8, under the section “Replication closeness evaluation”:

“We adopted a between-subjects design in which participants either received 1) no controllability information, 2) information that depicted all conditions as onset controllable, or 3) information that depicted all conditions as onset uncontrollable. The original study used a within-subjects design by evenly mixing controllability and uncontrollability stigma information in the two experimental conditions. ”

Following the change to the between-subjects design, we made the following changes to “Replication: As in the original” under the section “Data analysis strategy”:

“Following Weiner et al. (1988), we first focused on the no information (condition) ($N =$ 267) and conducted one-way repeated measures ANOVAs to compare the ten physically-based versus mental-behavioral stigmas used in the target article on perceived stability and controllability, affective reactions, and help-giving judgment. We then conducted correlation and multiple regression analyses to test the relationships between stigma source and those dependent variables in the control condition across all stigmas.”

We added the sub-section “Replication: Improvement from the original” under “Data analysis strategy”, which follows the sub-section “Replication: As in the original”:

“We used two-way between-subjects ANOVAs to compare the controllability condition ($N = 266$) versus uncontrollability information ($N = 267$) for physically-based stigmas versus mental-behavioral stigmas on perceived controllability, affective reactions, and help-giving judgment.”

12. On a more conceptual note, I am not convinced that the conditions examined in both the original and the extensions in this replication study are all stigmas. Stigma is an attribute or socially constructed group that is devalued and categorises a person as different from "us" (Goffman, 1963). While we know that there is stigma associated with HIV, obesity, drug abuse, PTSD etc., I'm less certain whether having diabetes, a stroke, cancer, blindness, Alzheimer's disease, or heart disease constitute stigmas. Is there evidence that people with these conditions experience stigma? If they did at the time the original study was conducted, is that still the case now (and vice versa)?
I would like to see some more discussion (and references) in the introduction about why these are considered stigmas (particularly in justifying the inclusion of diabetes and stroke as new stigmas beyond the fact that they are prevalent health conditions), what may have changed since the original study was conducted, and how this relates to the importance and implications of this replication

We understand what you mean, but this is a replication project, and our main goal is to simply repeat a classic very well cited impactful study in the literature.

The original article defined stigma as "a mark or sign for perceived conditions of deviations from norm" (Jones et al., 1984) and a bearer of stigma are defined as "generally limited or undesirable" (Goffman, 1963), and summarized stigmas to "represent negative outcomes or unwanted effects." We followed the original's conceptualization, which primarily portrayed stigmas as conditions considered undesirable or limiting in society, rather than the discriminatory aspect of stigmatization. Thus, even though there is no heavy discrimination towards individuals with diabetes, stroke, and blindness etc., the conditions themselves seem to be recognized as debilitating and undesirable, which follows the original article's conceptualization of stigma.

We changed the opening to paragraph 2 under the 'Background' section for clarity:

> "Public stigmas are negative social reactions to perceived deviations from the norm (Jones et al., 1984) and are connotative of negative outcomes, or undesirable, debilitating conditions (Goffman, 1963; Weiner et al., 1988)."

We also added a paragraph in the sub-section "Limitations" under "Discussion" to discuss the possibility that people may have different perceptions and definitions of stigma as follows:

> 'Weiner et al. (1988) defined stigma as "a mark or sign for perceived conditions of deviations from the norm" (Jones et al., 1984). A bearer of stigma is viewed as "generally limited or undesirable" (Goffman, 1963), and the target article summarized stigmas to "represent negative outcomes or unwanted effects." We followed the original's

conceptualization, which primarily portrayed stigmas as conditions undesirable or limiting in society. However, there may be other ways to conceptualize stigmas, including connotating stigma with discrimination and marginalization. The current study may be better understood as "an attributional analysis of reactions to socially undesirable conditions".'

References:

- Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity.* New York: Simon & Schuster
- Jones, E. E., Farina, A., Hastorf, A. H., Markus, H., Miller, D. T., & Scott, R. A.(1984). Social stigma. San Francisco: Freeman

13.   I think the authors could draw more on previous research and theory in supporting their exploratory directions. For instance, there has been a lot of work done on the role of perceived stability and controllability in the context of intergroup relations, stigma and discrimination - social identity theory makes clear predictions about this. I'm not suggesting that the authors must draw on social identity theory specifically, but I think this section would be strengthened considerably if it made connections with current social psychological theory, referred to more recent research, and made the case for how this exploratory direction might contribute to the field.

Thank you for your suggestion. As replication is the main goal of our study, we would like to keep the focus on the core hypotheses.