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. 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.

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.

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

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…".

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)?

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).

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.

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

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.

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.

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.)?

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).

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.

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.