**Summary of the Article:**

The authors, Lu and Feldman, present a Stage 1 Registered Replication Report for the foundational Lerner & Keltner paper by 2001., which investigated associations between dispositional fear, anger, happiness and risk optimism or risk preference. They adjust the original design to fit for an online study and propose to combine studies 1-3 of Lerner & Keltner into one design. Furthermore, they extend the original paper by adding hope as additional dispositional emotion.

**Introduction of Review:**

Before coming to the review part I want to thank the authors for their efforts to increase the quality of psychological research and want to laud the general research programme by Dr. Feldman aimed at replicating foundational work in judgement and decision making research. I also want to laud the fact that the authors decided to submit their work as a registered report and that they recognize that direct replications need not repeat mistakes of original studies.

**Summary of Review:**

In summary, I recommend acceptance with revisions. I will detail major points to be addressed first and minor points and stylistic issues to be addressed after. Briefly put, the authors need to a) better justify some of the choices they make regarding study selection, sample size justification, and measure selection and b) make it clearer how possible findings will be interpreted.

**Major Points:**

1. The authors argue for their choice study by stating that there are a) no direct replications, b) they can improve on the methodology used by the original authors and c) the impact of the original study. While I agree that the original study was quite impactful and am also not aware of any direct replications, I`d ask the authors to provide stronger justification for the need of a direct replication.
There are three reasons for this: Firstly, there are many impactful papers which have not yet been directly replicated. Secondly, the original studies were for the time quite well powered, particularly Study 2. Thirdly, there are likely several conceptual replications and extensions of their work among those 4211 citations. Impact and the lack of a direct replication are not sufficient justification for why a direct replication is necessary. One potential way to justify the importance would be highlighting the theoretical implications of the different potential findings. I`d recommend the authors to check Isager et al., 2021 (<https://psycnet.apa.org/record/2022-14587-001>) for some additional guidance and discussion. Relatedly, the authors write that there are practical implications for other domains without spelling out what these implications are – I`d recommend that the authors are more straightforward about what those implications are.

2. Table 1 details the hypotheses but does not explain how the different potential findings would be interpreted in light of theory being evaluated. The table at the very beginning of the manuscript does a better job, at least mentioning for which theories these findings could be relevant. I`d ask the authors to add a table that details A) The Hypothesis B) The Analysis being used, e.g., as a regression equation and C) The interpretation of the different potential results in light of the theories being evaluated. It is important to detail what a significant finding or a non-significant finding will mean in terms of theory.

3. On Page 14/15 you write “In our main replication analyses, we aimed to use the categorization of the target article as is, yet to improve on the methods of the original we also opted to directly assess participants’ perceived controllability and certainty of the events. With these measurements we sought to revisit the ambiguity categorization of the target article and also conduct analyses of ambiguity as the two continuous measures rather than a dichotomy of an aggregate.” If I understand this correctly, the authors will use both the operationalisation of the original authors and their own improved version. What will they do if results of both operationalisations do not converge?

4. On Page 16 the authors explain that “Our sensitivity analysis indicated this sample would allow for the detection of r = .12 and Cohen’s f2= 0.015 (4 predictors; both 95% power, alpha = 5%, one-tail), considered weak effects in social psychology (Lovakov & Agadullina, 2021), and therefore reasonable as our Smallest Effect Size of Interest (SESOI)”.
In my opinion, such a justification is not sufficient, because effect sizes cannot be interpreted out of contexts and claims that effects matter must be accompanied by empirical evidence for this claim. By determining a SESOI, the authors are effectively arguing that this is the smallest effect size that matters. My question would be why? One way to make an argument would be to explain how a particular effect size corresponds to real life outcomes. For example, if there is a correlation of *r* = .1 between anger and risk preference, what does that mean for the likelihood to have a car crash? One could also think about what effect sizes would be meaningful for or in light of the theory. The authors can find some guidance and discussion in Primbs et al., 2022 (Note: My own work), Anvari et al., 2022, and Anvari & Lakens, 2021.

5. For several of the trait questionnaires on page 21ff you write that you were unable to determine the exact items used. I`d ask to contact the original authors and check if they can share this information (if you did not already do this of course). Jennifer Lerner is still active: https://www.hks.harvard.edu/faculty/jennifer-lerner

6. I believe the authors miss a strategy for analysing and interpreting non-significant results. I`d recommend to check out equivalence testing – if the aim is to replicate a finding, it is paramount that also null results can be interpreted properly.

 **Minor points:**
1. The authors wrote on page 10 that there is no direct replication and state on page 11 that they will conduct a close replication. I later saw your Table detailing how they arrive at the fact that this is a close replication, but for the reader on page 11 this is a confusing difference.
2. Lerner & Keltner used American Undergraduate Students. The authors seem to use a different sample: Please justify why these samples are comparable.
3. Page 17: “We also employed the Qualtrics fraud and spam prevention measures: reCAPTCHA, prevent multiple submission, prevent ballot stuffing, bot detection, security scanmonitor, relevantID, etc.”. Please name all measures you use.
4. I applaud the authors for ensuring a fair payment of participants with the procedures lined out on page 17.
5. Table 4 is difficult to understand and contains information that is presented elsewhere in the manuscript. I`d suggest removing or restructuring the table. For example, the item phrasing is included only for some of the variables but not for all, and is partially discussed again later on. I think you could remove those.
6. On Page 21 the authors write: “Given the high Pearson correlation reported in the original between the two scales (r = .54), we followed the original’s method in combining the two scales into an aggregate score”. What do the authors plan to do if they do not observe such a high correlation in their sample?
7. I`d ask the authors to include regression equations in the manuscript. They remove all uncertainty about the analyses you will conduct. For example, now I`m not clear whether you will conduct separate regressions for each dispositional emotion or combine the IVs in the same analysis (see also Major Point 2)
8. On Page 28 right before the last “insert table” is a free-flying “N”

With Kind Regards,

Maximilian Primbs
(max.primbs@ru.nl)