**Reviewer 1 Feedback:**

This reviewer provided high-level feedback regarding our manuscript, which makes it difficult to respond to point-by-point as I had with other reviewers. Therefore, below I include a portion of the reviewer’s response with points that we interpreted as especially important underlined:

“…In Study 1, the authors discuss their hypotheses related to 'intensity' as if intensity was an independent variable that's being manipulated (as it is in Gal Sheppes foundational work) but this may not be appropriate in Study 1. In this study, intensity is a measured variable (not a manipulated one) and in this particular study context, it's very likely an outcome of regulation just as much as it's a predictor of regulation. For example, another way to interpret figure 4 is that - rather than this being an unexpected finding - it might be exactly what one would predict if emotional intensity was the \*outcome\* of regulation success, which it very well could be given that these variables were measured at approximately the same time: as regulation success decreased, emotional intensity increased.”

“For this reason, I wouldn't necessarily feel comfortable with this take-home message, which makes it sound like the present results are the opposite of what prior work has demonstrated: "Though the extant literature from comparable lab studies should motivate us to expect the efficacy of distraction to increase and reappraisal to decrease as affective intensity increases, our data seems to document a deviation from this pattern in a high-intensity, quasi-naturalistic setting: distraction appeared to be less - not more - successful as affective intensity increased.”

“Aside from this moderation effect, the main effect between intensity and strategy choice was the primary analysis in this study, and there was no reliable/significant association found. That could be informative but, given the nature of the study, I'm not sure how informative this null association is. If emotional intensity drives use of distraction vs. reappraisal, we'd expect a positive association between intensity and distraction use. But if [lower] emotional intensity is also the \*outcome\* of successful regulation - especially distraction, which was the modal strategy used - we'd expect a negative association between intensity and distraction use. These two patterns operating at once could yield a null result, which could explain Study 1 findings.”

“This alternative interpretation also tracks with what the authors found in Study 2: when new participants are told that a given event is higher vs. lower intensity (i.e., intensity is manipulated here, rather than measured like in Study 1), they choose distraction (vs. reappraisal) more often. This is essentially a conceptual replication of the Sheppes work because intensity is manipulated (i.e., given to participants) and isn't really comparable to the intensity variable in study 1, which is a complex experience that is likely being affected by regulation as much as it's affecting regulation. For this reason, Study 2 can't effectively be used to help explain the pattern of results from Study 1.”

“The authors then conducted Study 3 to learn whether the link between intensity and distraction choice (vs. reappraisal choice) would be present in forecasted regulation contexts (like Study 2) but not in executed regulation contexts (like Study 1). But by my read, this isn't the core difference in the findings between Study 1 and 2 and so when I saw that the experimenters were again manipulating intensity in Study 3 (this time with pre-piloted lower vs. higher intensity film clips), it seemed fully reasonable for them to replicate the 'canonical relationship' between intensity and distraction, which they did. This pattern makes good sense if Study 2 and 3 are interpreted as solid conceptual replications of the original Sheppes work, where intensity is carefully manipulated for participants.”

“Altogether, this manuscript describes a very cool study (Study 1) that seems likely to be published across multiple papers (given that it seems to be a larger multi-purpose dataset), but the particular question tested in this manuscript yielded unexpected findings (due, to my mind, to the core 'predictor' being operationalized in a way that isn't comparable to prior work). And then two additional studies were run to try to explain the unexpected findings. But, if we reinterpret Study 1 as I suggest above, then follow-up studies aren't solving an unresolved question from Study 1 but rather, are (nicely!) replicating the effects we'd expect from Sheppes' model…”

Based upon the highlighted passages, we interpreted this feedback as containing three primary but related points, which we will response to here:

1. The authors discussed intensity in Study 1 as if it were a manipulated variable rather than an observed variable. The use of observed emotion intensity is a substantive deviation from the Sheppes paradigm, limiting the extent to which results from the two can be compared.
2. Because this study does not attempt to manipulate emotion, we cannot know whether self-reported emotion intensity captured post-exposure is a product of or precursor to self-regulation.
3. A null association is inherently ambiguous. The results of the supplementary analyses contained within could be explained by emotion as either a product or precursor. Therefore, our conclusions may be overstated.

**Our response:** If Figure 4 represents emotion intensity as a product rather than a precursor, then

**Our response:** Why is it very likely? It’s perhaps equally as likely, but I have no reason to believe that it is more likely. There are two dissociable components to this critique. We agree that the analysis that we used did not manipulate and did not seek to manipulate emotions. Why would manipulation of an emotion differ from a non-manipulated one? They make more assumptions than we do, as they often use the standardized measure of emotion intensity rather than the experienced emotion. Though not perfect, some degree of affective manipulation did occur in the study. Subjects traversed the haunted house through 4 sections: two of which were deemed high intensity by the designers and two of which were deemed to be of lower intensity. When we analyze regulation behavior using these metrics, we still find the same regulatory behaviors.

Acknowledge that one means of accomplishing this could have been to have independent raters quantify the intensity of each event and to use those. This was difficult in our case given the time scale we had to work on.

Make the point that by taking an observational approach and capturing personal experience, we're forced to assume that the personal experience that they report is pre-regulation. However, if we were to take an IAPs approach, we'd also be forced to assume that the standardized value as we have represent personal experience.  
  
It is true that intensity could be either a product of or precursor to self-regulation in this study. While we argue that the language that we used (and "did you attempt to reduce or intensify this emotion?") suggests that we intended to target the emotional precursor (i.e., you may be able to prepare regulatory responses, but you cannot regulate something that hasn’t happened).   
  
. We [will change language in the manuscript to reflect this].

Specker et al., 2023 found that pre- and post-regulation emotion ratings were highly correlated (r = 0.81, p < 0.001). As best as I can tell, no one has tried to assess whether post-regulation intensity is a significant predictor of affective intensity. I could not find any articles which have explored whether there are still statistically significant differences in the post-regulation intensity of high and low intensity images after regulation. Shafir et al. 2016 did find a main effect of emotion category upon attenuation with a considerable effect size (ηp2 = 0.83), but this was calculated as the difference between the self-reported post-regulation rating and the standardized rating, so the baseline may be somewhat unreliable for this purpose.   
  
I tried searching for an emotion regulation dataset in which to test this theory, but none were publicly available.

**Our response:** It’s not intended to be read as the opposite, but it is intended to document the same phenomena. If the only want to comment upon Sheppes’s reappraisal and distraction is by using standardized stimuli in controlled settings with trained individuals whose experience we can dissect then no naturalistic study will ever comment upon them. We will be doomed to forever entertain a pristine, bottled version of emotion regulation that psychologists fawn over and the kind of self-regulation that everyone else actually has to use.

**Our response:**

**Our response:** I think it’s a large assumption to assume that this study design manipulates intensity. We’re just telling them intensity was. We can’t say Study 1 doesn’t match Sheppes but Study 2 does just because Study 2’s conclusions better fit Sheppes if we squint. There are many aspects of Study 2 that still differ from Sheppes, such that I would never call it a straight forward replication. It’s simply more similar in the specific dimension of controllability.

**Our response:** Make the point that studies 2 and 3 were not straight forward Sheppes replications. we get some patterns that are consistent with sheppes, specifically in participants that are either extremely psychologically distant from the experience (study 2) or making predictions about what should be normatively done for others (study 3). Emphasize the ways in which StuThe none option, measuring intensity before and after, using independently assessed intensity to predict self-regulationdy 3 deviated pretty substantially from Sheppes in both design and outcome.

Examine whether the other study that used individual intensity ratings also deviated from Sheppes [[SPECKER 2023]]

Claimed to be the first to measure pre- and post- regulation intensity in order to assess the importance of regulation flexibility (Specker 2023). Had Sheppes as a middle author. Their language is very confusing, but I think what happened is they aggregated trials and found, across the sample, an association between intensity and choice, but when they disaggregated, they found no association between intensity and choice. They kind of dismiss the latter finding as not having enough experimental control.

**Reviewer 4 Feedback:**

No comments