**Reviewer 1 Feedback:**

We appreciate this reviewer’s high-level feedback and will address each critique line-by-line referencing the tracked changes version of the manuscript for clarity:

“*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)…”*

**Our response:** Thank you for highlighting this concern. We acknowledge emotional intensity is a measured variable in Study 1. Our comparisons to Sheppes’s work aim to contextualize our results within the broader field, acknowledging both similarities and differences. To clarify this distinction, we have added additional statements on pages 6, 9, 20, 27, 31, and 53 to emphasize our approach is observational.

*“…and in this particular study context, it's very likely an outcome of regulation just as much as it's a predictor of regulation.*”

**Our response:** We agree that the measured variable could be both an outcome and a predictor of regulation. We have updated our manuscript on pages 28 and 53 to reflect this limitation and included the precise language used for self-report on page 17.

To attempt to address the reviewer’s concerns regarding experimental control and limitations of our primary predictor, we conducted an additional analysis examining regulation usage as predicted by the section of the haunted house in which the regulation occurred – a proxy for emotion intensity. The organizers of the haunted house designed two sections to be high-intensity and two sections to be low-intensity (See Cliver et al., 2024 for more details). These constitute a more controlled, albeit a lower resolution, representation of emotion intensity than our primary analysis. An assessment of self-reported fear collected immediately after each section mirrored this design structure (See pg. 27 for details). Using these sections as categorical predictors allowed us to explore associations with strategy usage in a similar fashion to the Sheppes and Sheppes-inspired work, but still resulted in no observed association between intensity and usage. This is of course an imperfect, post-hoc solution to this criticism, especially given that the majority of regulated events occurred in high-intensity sections, but we believe it at least adds credence to the notion that our null results in Study 1 are not simply a product of lack of control. These changes are covered on pages 19 and 27.

*“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.”*

**Our response:** We appreciate this alternative interpretation. While we intended to target emotion as a precursor, we acknowledge the potential for post-regulation assessments. We have revised our discussion on page 29 to reflect this and included an additional analysis to clarify our point.

*“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.”*

**Our response:** We agree that this was incorrectly stated. Distraction should not be interpreted to grow in efficacy as intensity increases (some events may be too intense to regulate at all), but it is relatively more efficacious than reappraisal as high intensities. As previously noted, we corrected this language, but also contextualized our finding by reminding the reader that there are methodological differences between our approach and the approaches that we are benchmarking ourselves against.

*“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.”*

**Our response:** Mixed interpretations of our question could potentially obfuscate a true effect, but such confusion would likely occur on the subject-level (i.e., different subjects may interpret the question differently, but the same subject would likely interpret the question the same every time they answer it). This sort of idiosyncratic effect is precisely what statistical approaches such as the hierarchical modeling we used are designed to adjust for. We agree with the reviewer that “… If emotional intensity drives use of distraction vs. reappraisal, we'd expect a positive association between intensity and distraction use” but feel that the logic behind “ … if [lower] emotional intensity is also the \*outcome\* of successful regulation … we'd expect a negative association between intensity and distraction use” is making the same mistake noted earlier: that distraction is so effective at attenuation that post-regulation intensity of a pre-regulation high intensity stimulus would look indistinguishable, if not lower, than pre-regulation low intensity stimuli. I believe that a more accurate expectation, based upon the cited literature, might be a positive association with a lower intercept or smaller logistic coefficient. We lack the ability to conclusively explain the null with this study design.

As we view it, the value of reporting this null is simply that when assess these variables in an ecologically valid way (i.e., minimizing manipulation) and in a context with features (i.e., highly stimulating, high intensity, complex) which mirror other circumstances in which self-regulation could be of vital importance, we do not find this relationship, despite an impressively substantial effect size and consistent replication in more controlled contexts. In considering this point, we made modifications to our abstract and significance statement to more accurately reflect that how we adjusted experimental control is likely an important component to what we observed across all studies; not just Study 1.

*“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.”*

**Our response:** Study 2 was designed to mirror aspects of Sheppes’s work, demonstrating the expected patterns under controlled conditions. We intended for Study 2 to complement Study 1 by showing the importance of experimental control. The comparison, while imperfect, is valuable in highlighting the boundaries of the observed effects.

*“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.”*

**Our response:** We conducted Study 3 to compare forecasting and actual usage of regulation strategies. While it replicated Sheppes’s canonical relationship under controlled conditions, it aimed to illustrate the effect of reduced experimental control. We have revised the discussion on page 39 to clarify this motivation and the limitations of our design.

**Reviewer 4 Feedback:**

No comments