

8 November 2023

Dr. Jessica D. Payne, Associate Editor

*Journal of Experimental Psychology: General*

Manuscript No. XGE-2023-1118

Dear Dr. Payne and reviewers,

Thank you for the thoughtful feedback contained within your decision letter dated August 09th, 2023 and the opportunity to resubmit our manuscript (XGE-2023-1118). We have revised our manuscript in accordance with the reviewer comments and organized those changes in the appended document. All comments are numbered and organized by reviewer and gravity, for ease of identification. Both clean and tracked-change copies of our manuscript are included with this most recent submission.

We believe that our revision addresses all concerns cited by the reviewers and has certainly increased the quality of the manuscript. Of course, if there are any questions or concerns, please do not hesitate to contact me. Again, we appreciate both your time and the time of the reviewers and we look forward to your assessments of this most recent version of our manuscript.

All the best,

William Mitchell

Doctoral Candidate, Psychology & Neuroscience Department

Temple University

William.J.Mitchell@temple.edu

**Reviewer 2 Feedback:**

**Major Comments**

1. "These approaches offer strong-validation and unparalleled control but might not accurately reflect the multidimensionality of emotional experience." — I don't quite follow this conclusion. The approaches do not offer "strong validation" or "unparalleled control" - they are just people rating images.”

**Our response:** We agree with this point and have updated our language to be more precise about what these approaches offer the field: “*These approaches offer an efficient, reliable, and standardized means of assessing self-regulation but might not …*”

1. “In the methods section it states, "Of the 182 self-reported events… 30.7% used reappraisal and 61.5% used distraction…" and later, a small proportion used suppression, and a negligible proportion used situation modification and situation selection. This information seems like it would be more useful to the reader if placed in the results section. In general, the results section could use more organization to make sure the key points sink in and have their space.”

**Our response:** We believe this was a misunderstanding stemming from unclear language. The 182 events we refer to here were not observed during Study 1, but rather during our pilot. We cite this statistic here as a justification to focus primarily on reappraisal and distraction in Study 1. However, we did update the language to hopefully make this more clear. We also do take the point about rearranging information in the results more generally and have made some modifications to further emphasize the key points, including breaking larger sections into smaller ones with more specific headings.

1. “In the "covariates" subsection, it is stated, "to assess each variable's potential covariation with affective intensity in predicting regulation usage" — do you mean each variable was tested as being a moderator? And later, it is stated that each variable was tested for "collinearity with affective intensity". But the only result reported is that cognitive load failed to predict the type of emotion regulation strategy. Where is everything else? It is also seems unusual, given the rest of the subheadings in the results section, to use "covariates" as a subsection when the inferential statistic is testing a specific hypothesis. In "Experiment 2 Methods", the purpose becomes clearer - "…even when adjusting for noted moderators like cognitive load." Perhaps that goal can be made more explicit above. ”

**Our response:** Thank you for highlighting this. The intention was to highlight that we did not find evidence to suggest nuisance variables that commonly might confound the relationship between intensity and usage, such as time of time or cognitive load, had any statistically significant relationships to our predictor or outcome variables. Earlier versions of the manuscript did include a detailed description of the analyses conducted and results for each nuisance variable, but these were cut both for word limit concerns and to direct audience attention to the details and analyses more central to the primary purpose of the study. However, these analyses, justification for these analyses, and the results of these analyses remained present in the markdown script at the time of submission.

Our preregistration for Study 1 specifically noted that Cognitive Load may moderate the relationship between intensity and usage and, thus, was specifically mentioned, though we did not find the hypothesized association. This statement felt out of place and we agree that this section demands general revisions. As a result, we have altered the title, framing, and briefly summarized the results of each analysis. We also separated cognitive load from this section to its own section to more closely reflect our preregistration. If the reviewers feel further specific corrections are required, we would be happy to do so in a future revision at their request.

1. “In study 1, the descriptions were used to code for regulation strategies. In study 2, were the same descriptions given to a new set of subjects, who then indicated which regulation strategy they would use? If so, is there a concern that subjects from study 2 could determine the regulation strategy that subjects in study 1 were using, and relying on that information to make their judgments? ”

**Our response:** Thanks for highlighting this, as it is a point that we should better clarify. Participants in Study 1 were asked to first describe the emotionally salient event in as much detail as they could, and to then separately describe in their own words how they attempted to regulate that event. Participants in Study 2 were presented with the event descriptions, but not the regulation descriptions. The event descriptions were screened for any indications of regulation strategy usage (e.g., “… so I looked at my shoes”, “… so I imagined that this was a movie”, “… I closed my eyes and ran”) prior to being used in Study 2 and none were identified. As such, we do not believe that the event descriptions in and of themselves unduly directed Study 2 participants to choosing one of the available options, but I have revised the description of Study 2 to make that more apparent.

1. “The authors refer to environmental affordances as a potential factor that explains which regulation strategies end up being selected in a given situation. I think this is a nice way of interpreting the results. I think the section focusses too much on a passive viewing perspective (i.e. on how the environment grabs attention). The situation may also dictate the action affordances one has available. From Figure 4 examples, there are clearly actions available like running to the next room. It's unclear how reappraisal and distraction interact when these "escape" opportunities are there, too, in naturalistic contexts. Action affordances in general seem like an important aspect of what separates emotion regulation in everyday life v. in decontexualized experimental settings, which could be developed a bit further in the manuscript.”

**Our response:** Again, we agreed that this was an underdeveloped point in our original manuscript and have expanded our discussion to incorporate it. We have additionally added **WHAT WE ADDED**.

1. “I'm wondering if the introduction might also benefit by more explicitly framing the argument that the study is testing the external validity and generalizability of conclusions from laboratory paradigms to situations that can occur in everyday life more generally. The current framing emphasizes more narrowly focusses in on certain dimensions (multimodality of experience), but I think it may help to nest this notion as a subset of the broader theoretical question at hand, which concerns external validity. The authors might also consider the following articles to help situate the theoretical space:

Miller, L. C., Shaikh, S. J., Jeong, D. C., Wang, L., Gillig, T. K., Godoy, C. G., ... & Read, S. J. (2019). Causal inference in generalizable environments: systematic representative design. Psychological inquiry, 30(4), 173-202. — which argues for the importance of representative design in psychology.

Lee, K. M., Ferreira-Santos, F., & Satpute, A. B. (2021). Predictive processing models and affective neuroscience. Neuroscience & Biobehavioral Reviews, 131, 211-228. — which argues in section 5 the importance of external validity and representative design based on predictive processing models of the brain (and integrates with constructionist theory insofar as emotions are constructed by volleys of an integration of predictions and prediction errors) ”

**Our response:** Again, we agree with this point and earlier versions of the manuscript did attempt to incorporate this framing, though, feedback we had received in the process resulted in it being removed. We have reincorporated that language, specifically WHAT WE DISCUSSED.

1. “I'm wondering if the discussion section could also broaden to highlight other studies that challenge external validity - might there be other studies in emotion or in memory wherein findings in the lab do not converge with those conducted in everyday life settings? If so, it would bolster support for this study in contributing to a broader movement that underscores the importance of external validity and generalization to everyday life. ”

**Our response:** As noted in the response to Comment 5, we expanded the discussion section to include WHATEVER WE INCLUDED, which we believe also satisfies this critique.

**Minor Comments**

1. “I'm wondering why the plots show Emotion Intensity as z-scores (Fig 2, 3, e..g) - might the actual scale participants completed be informative here? Also, the endpoints of the scale for emotion intensity are unclear (even in the methods section; page 14). Is Emotion Intensity calculated by averaging over each emotion indicated per person and haunted house situation? I'm guessing not… but in general, this measure could be clarified in the writing. Relatedly, is it in principle possible that someone could have indicated "calm" or "sleepy"? In that case, would emotional intensity mean intensely calm or sleepy? How about intensely sad? While intensely afraid or excited seems to make sense when thinking about emotional intensity as arousal, in some of these other cases I'm a bit confused. Ultimately, it may just help to clarify what, exactly, emotional intensity is referring to. ”

**Our response:** We apologize for the confusion. The range of possible intensity scores (0-4), as well as the average score (2.44) is listed on page 18. However, I have updated the description on page 14 to include the actual labels participants saw to hopefully better represent the experience.

We believed that using z-scored scales rather than the original raw scales would improve interpretation, but we are open to feedback and criticism regarding this supposition. Our rationale was that the original scale may have limited generalizability beyond the scope of this study, but a standardized scale has a readily-accessible and universally-accepted interpretation, which could facilitate cross-model comparisons. By using standardized units, we could also minimize skewed presentation, as the resultant plot becomes a visualization of the standardized effect size. However, this question did bring attention to an important oversight: we had omitted the standard deviation from our descriptive statistics of affective intensity, limiting interpretation. To perhaps find a ‘happy medium’, we’ve modified our plots to show both the raw and z-scored values. Again, though, we are open to feedback if the reviewers feel an alternative style of presentation might be a better fit.  
  
Regarding how emotional intensity was calculated, we have clarified the association between usage and intensity on page 15: “*Thus, rather than exploring this phenomena at the event-level, which might require regressing the probability of using a strategy upon the average intensity of all emotions experienced in that event – an assumption we would not make in confidence - we draw associations between regulation strategy usage and the emotions that participants identify as directly motivating them.*”

In principle, this design would allow for participants to select intensity-emotion pairings which might be unexpected, such as ‘very intensely sleepy’. In practice, though, such an indication may be slightly more interpretable, as although I keep referring to this variable as ‘intensity’ for linguistic convenience, the labels participants actually selected to describe each emotion are less arousal-specific. For example, a participant who indicated feeling sleepy would have the options to select being “not at all”, “a little”, “a moderate amount”, “a lot”, or “a great deal” sleepy. The emotion responses that participants wrote were piped into the subsequent questions about that emotion so that participants could assess that emotion on this scale. Again, while perhaps not without flaws, I do think these labels provide a better description of what we’re capturing when we talk about intensity. Additionally, as you suggested, we rarely saw low arousal words used by participants in practice. The following plot is the distribution of arousal values associated with each emotional response participants provided, as determined by Mohammad (2018)’s NRC lexicon. The distribution appears positively skewed (*x̄* = 0.698, *median* = 0.773, *sd* = 0.175), indicating many of our responses were much closer to ‘afraid’ or ‘excited’ than calm. To perhaps better illustrate the types of responses that this design elicited, we added a new figure containing a word cloud of all responses participants gave, as well as those responses that met eligibility for our primary analysis. Hopefully this helps alleviate confusion and adds clarity to this aspect of our study.

A graph of a distribution of emotional values

Description automatically generated

1. “Several models were run to address contra-hedonic regulation activity (page 20). The exploratory findings are presented in Table 1. One model surpassed a "traditional" threshold, which I think means an uncorrected alpha level. As a side note, perhaps the term "nominal" might be better here. This model seems like it would involve a lot of parameters. Does the modeling approach penalize for model complexity? Separately, I think the motivation to address contra-hedonic regulation and the broad conclusion that it is not the case is useful to state in the main manuscript. However, the details of the method and results (null findings), including Table 1, might be better suited to the supplementary materials (unless there is a bigger point to this table that I'm missing). ”

**Our response:** preemptively might be intentionally obfuscating an otherwise well-established effect

1. “For study 2 methods, "reviewed examples of how both strategies might be employed" - it may be informative to include concrete descriptions of the examples that were used to train participants since these may influence the priors for the test items. ”

**Our response:** We had originally thought to not do so since the training materials are available in our public repository for the project, but have since reconsidered and added examples within the manuscript as well.

1. “In the discussion, it states, "Affective intensity predicted regulation extent…" I couldn't connect "regulation extent" to anything in the results section. ”

**Our response:** This was an oversight on my part. I have corrected this language to be more consistent and expounded upon what we meant by and how we measured regulation effort in this section.

1. “Some primary results are presented in the discussion, which seems a bit unusual. To focus the discussion, I suggest developing these alternative account in the results section. In particular, "Although participants were instructed to not discuss their experiences, the group context in which the experience occurred may have influenced behavior choices and cognitive perceptions. However, post-hoc analyses failed to find any association between group membership and strategy usage (F(30,45) = 0.93, p = 0.57). The presence and strength of friendship among group members was also assessed and was not predictive of regulation (t(60) = -0.4, p = 0.70”

**Our response:** This was another situation in which I worried noting so many additional auxiliary statistical tests that we ran for a specific purpose (e.g., assessing sample or study limitations) some distance away from the context in which they become relevant (e.g., the limitations section) might become confusing or overwhelming, but it seems that it had the opposite effect. I have tried to centralize all of these types of analyses within the results section as requested.

1. “Is there a typo in Figure 4A, "jot" should be jolt?”

**Our response:** Thanks for highlighting this; we have corrected it.

**Reviewer 3 Feedback:**

**Major Comments:**

1. “On p. 28, the authors distinguish between the effects of emotional intensity on the choice of emotion regulation strategies in "low-stimulation" and "high-stimulation" paradigms (concluding that intensity does inform choice in low, but not high stimulation settings). It took us a while to understand what was being referred to here as the authors previously framed the difference between Study 1 and Study 2 in terms of experiencing (Study 1) versus reviewing (Study 2) emotional experiences. This may be a more appropriate framing, partly from the perspective of consistency, but also because the authors do not present evidence that forecasters (in Study 2) found the reviewed events less emotional / stimulating than participants who experienced those events in Study 1. ”

“The authors also need to be careful in this section not to slip back into asserting that Study 1 measured participants choice of regulation strategies. Indeed, it might be prudent to consider the limitations of comparing findings between two studies which differ in a number of other ways (as the authors' point out in their introduction), including the way that emotion regulation was measured. In short, it may not be differences in (overall level of) emotional intensity between Study 1 and Study 2 that drive the difference in the way that people say they would choose to regulate (Study 2) or report regulating (Study 1) their emotions between specific situations that differ in emotional intensity.”

“The best way forward might be to conduct and include an additional study that directly manipulates whichever variable (or variables in a factorial design) the authors believe accounts for the difference in the effect of emotional intensity on emotion regulation (e.g., experiencing vs. reviewing emotional events, being trained in specific strategies vs. being untrained). An additional study would prevent the authors having to draw comparisons between studies (both within their manuscript and others' research) and provide confidence in conclusions.”

**Our response:** We do agree with this point, also feel that the framing was inconsistent in the original submission, and have since revised it to focus on experiencing and forecasting, especially with the addition of Study 3, which we conducted to further inform the interpretation of Study 2. Given the planning and resource cost that a full field study (i.e., haunted house study) would demand, we were not able to explore the

1. “Emotional intensity was not manipulated, but rather was measured (retrospectively) using self-report. This is a significant limitation because retrospective recall of emotional experience may differ significantly from in-the-moment experience. This methodological decision is also difficult to understand, given that it might have been possible to manipulate intensity within a haunted house. If an additional study was conducted (as suggested above) then this limitation could be addressed, but it certainly needs to be considered in the section on limitations.”

**Our response:** We are not able to manipulate emotional intensity in a haunted house attraction. The set, events, and actors are chosen or designed by the organization that we had partnered with to run the experiment, but we did not have the ability to influence this. The resources required to create our own version of this at the same quality of this, which may have allowed us to exert some degree of experimental control, would be unattainable. Although in an ideal world we would capture descriptions of the events as they occurred, this would ruin the immersion and have undesired effects upon the emotional experience and would have fundamentally altered the way that participants experienced subsequent events within the haunted house. If the primary objective was to document how untrained participants self-regulate when unprompted to do so, interrupting the experience to ask self-regulation-related questions may have prompted later self-regulation. Additionally, there were many events that participants could have self-identified as emotionally salient and significant variation in which events participants did identify as emotionally salient. We could not predict beforehand which events a given participant may have chosen to report on or intervene after every event. This has already been captured in our limitations section when stating: “WHATEVER WE SAID”

1. “It was great to see that the authors considered the effectiveness of the strategies that participants chose. However, the authors conclusions may be too strong considering how effectiveness was measured (i.e., participants reports of how successful their regulatory efforts were). For example, on p. 21 the authors state that "our data seems to suggest the efficacy of using distraction within this high-intensity, quasi-naturalistic setting to be of a lesser magnitude than what had been found in lab studies wherein distraction was used". The authors need to provide references for the previous work that has looked at the effectiveness of distraction in response to high-intensity stimuli - both in the results section and in the introduction - and tone down the conclusions given the differences in the way that efficacy of regulation was measured.”

**Our response:** We have revised our language, as we did not intend to make an intercontext or objective comparison but rather a subjective one. When measured in the same experiment or context, stimulus-response laboratory paradigms like those we had highlighted consistently find that distraction tends to be more successful than reappraisal in high intensity settings. We found the opposite within this context and simply intended to highlight the relative difference.

1. “On p. 30, the authors state that "affective intensity predicted regulation extent but not usage"; however, it is not clear how "regulation extent" was measured. This should be clarified in the methods and/or results section of Experiment 1.”

**Our response:** Thank you for highlighting this. Including this statement on page 30 was an oversight and it has since been removed. Regulation Extent referred to the effort participants exerted in trying to self-regulate. It was originally addressed in the primary manuscript, but was moved to the supplemental materials in this revision. Though, we did update our language in the supplemental materials as well to hopefully make more clear how this variable was assessed.

1. “It was disappointing to read that the physiological data collected from participants in the pilot study has been reported elsewhere; namely, in a paper in Neuropsychologia. The paper in Neuropsychologia does not focus on how participants regulated their emotions, but the heart rate data provides a useful validation of the intensity of the emotion experiences and so would have been useful in this manuscript. We also note the suggestion that a third 'forthcoming manuscript' will report on a memory test given to participants (p. 13). It may be appropriate to split the data in this way, but the authors should do so having carefully considered issues around 'salami-slicing of publications' and the pros and cons of contributing to a proliferation of academic content.”

**Our response:** We had taken this criticism into consideration and conducted additional exploratory analyses where possible using the physiological and regulatory data from the pilot and Study 1. What did we briefly find?. Please see the updated manuscript for more specific details.   
  
We were not able to find enough data from Study 1 to run any analyses. There were data collection issues collecting within the field. In the pilot, we do have enough data but the events occur at the subsection level, so there is a fundamental misalignment.

1. “Finally, please simplify language where possible. For example, "had predictive utility towards" can just be "predicted". "…the affective intensity experiencers reported" (p. 26) could be "…the intensity of the emotions that experience reported" etc. The authors should also be consistent with the language used throughout the manuscript to ensure that the conclusions drawn accurately reflect what has been measured. For example, when discussing the findings of Study 1 throughout the manuscript, the authors should refer to strategy use and not choice (throughout).”

**Our response:** We have reviewed the language and corrected where appropriate. Please let us know if you still feel that the language in this revision is not sufficiently economical or consistent.