

Journal of Experimental Psychology: General

Emotion regulation strategy use and forecasting in response to dynamic, multimodal stimuli

--Manuscript Draft--

Manuscript Number:	XGE-2023-1118R3
Full Title:	Emotion regulation strategy use and forecasting in response to dynamic, multimodal stimuli
Abstract:	<p>Successful emotion regulation (ER) requires effective strategy selection. Research suggests that disengagement strategies (e.g., distraction) are more often selected than engagement strategies (e.g., reappraisal) as emotional experiences intensify. However, the extent to which ER strategy choice in controlled circumstances reflects strategy usage during complex, multimodal events is not well understood. The present research uses dynamic, multimodal stimuli (i.e., a haunted house, horror movies) to examine the association between affective intensity and regulatory strategy usage among untrained participants – individuals given no prior regulation instructions or direction. Both a preliminary study (n = 54) and Study 1 (n = 118) failed to find relationships between emotional intensity and strategy usage to downregulate emotions as participants navigated a haunted house. Distraction was self-reported to be less successful than reappraisal at high intensities, contrary to expectations. Participants in Study 2 (n = 152) forecasted regulation strategy usage based upon descriptions of emotionally-regulated experiences from the preliminary haunted house study. Affective intensity predicted which strategies forecasters predicted they would use; though, forecasters overpredicted how often distraction was used in practice. Study 3 (n = 242) incorporated strategy usage and forecasting within the same design by showing untrained participants video stimuli of varying-intensity and capturing their regulatory responses. Forecasters again predicted using distraction more often than strategy users did in practice. Forecasters also overpredicted how effectively distraction reduced negative affective intensity relative to what strategy users reported. These results may highlight a disconnect between strategy fittedness when self-regulation occurs in uncontrolled, highly-intense, or complex circumstances.</p>
Article Type:	Unmasked Article
Keywords:	Emotion; self-regulation; naturalistic stimuli; decision making
Corresponding Author:	William John Mitchell, M.A. Temple University Philadelphia, PA UNITED STATES
Corresponding Author E-Mail:	billy.mitchell@temple.edu
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Temple University
Other Authors:	Joanne Stasiak Steven Martinez Katelyn Cliver David Gregory Samantha Reisman Helen Schmidt Vishnu P Murty, PhD Chelsea Helion, PhD
Author Comments:	If the figures do not translate well to grayscale, I can submit alternatives with patterns for regression lines to more easily distinguish between conditions.
Corresponding Author's Secondary	

Institution:	
First Author:	William John Mitchell
Order of Authors Secondary Information:	
Manuscript Region of Origin:	UNITED STATES
Suggested Reviewers:	<p>Bryan Denny, PhD Assistant Professor, Rice University bryan.denny@rice.edu Dr. Denny has conducted extensive research on the use of emotion regulation strategies in both lab and naturalistic contexts and would provide invaluable insights towards the sufficiency of our approach to measuring our variables of interest in the field.</p> <p>Kristen A Lindquist, PhD Professor, The University of North Carolina at Chapel Hill Kristen.lindquist@unc.edu Dr. Lindquist is one of the foremost experts on constructivist theories of emotion and would provide important feedback regarding our methodology for measuring emotion.</p> <p>Ajay Satpute, PhD Assistant Professor, Northeastern University a.satpute@northeastern.edu Dr. Satpute's research often focuses on the influence of context upon emotional experiences (most notably fear) and, as such, he could critically evaluate the role that our unique study setting played in shaping our results.</p> <p>Brett Q. Ford, PhD Assistant Professor, University of Toronto brett.ford@utoronto.ca Dr. Ford has written extensively upon naturalistic emotion regulation behaviors and the relative strengths and weaknesses inherent to regulation strategies (most notably reappraisal). As such, she would have valuable feedback about our results and conclusions.</p>
Opposed Reviewers:	
Order of Authors:	<p>William John Mitchell</p> <p>Joanne Stasiak</p> <p>Steven Martinez</p> <p>Katelyn Cliver</p> <p>David Gregory</p> <p>Samantha Reisman</p> <p>Helen Schmidt</p> <p>Vishnu P Murty, PhD</p> <p>Chelsea Helion, PhD</p>

Reviewer 1 Feedback:

We appreciate the high-level feedback provided by Reviewer 1 and have provided responses to each critique. In some cases, we were not sure that we had interpreted the reviewer's request as intended. In these instances, we start our response with our interpretation of the critique. All following references to page numbers will refer to the clean version of the updated manuscript:

“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: This critique, as we have read it, raises a concern that we have incorrectly framed our predictor variable of interest as experimentally manipulated rather than observed. We thank the reviewer for bringing up this important point, and for helping us to clarify how our work contributes to existing emotion regulation research. In response, we have modified language throughout the manuscript to further delineate our work and distinguish it from Sheppes's work where appropriate. However, we have conducted additional post-hoc analyses using predictors from the study design that more closely mirror the original Sheppes work (this is discussed in more detail in the next response).

We make comparisons to “Sheppes[’s] foundational work” not because we believe that standardized ratings represent a direct comparison to idiographic self-report metrics, but because, conceptually, it is the preeminent standard to which almost all emotion intensity and regulation strategy research is currently being contrasted to. Few other studies in the extant literature, as recently noted by Specker, Sheppes, and Nickerson (2024), use subjects' self-reported emotional experiences as predictors of emotion regulation. Thus, the Sheppes paradigms remain the closest conceptual contrast to contextualize our results, even if we deviate from one another regarding experimental control of emotional intensity. As we view it, the models are still similar in the phenomena and associations they intend to represent – using emotional intensity (whether that be standardized or non-standardized) to predict emotion regulation (whether that be usage or choice). Our research primarily differs from prior work in the prioritization external and internal validity when modeling the target phenomena and attending to the effects of this difference. This was a central goal of this manuscript, and we therefore think that the comparison between this and prior work is valid and important.

Confusion about experimental control was a concern cited by previous reviewers (See our exchanges with Reviewer 3). We had previously modified our manuscript to reflect that the relationship in Study 1 was not experimentally manipulated (e.g., *“It must be noted that we did not directly manipulate emotional intensity within this design...”*, pg. 53) and to avoid language suggesting causality or experimental control (*“Study 1 tested whether the emotional intensity of negatively-valenced events was associated with the likelihood ...”*, pg. 10). In response to this critique, we have again carefully reviewed the language we used to describe Study 1. To ensure that we are not giving the impression that our predictor in Study 1 was experimentally manipulated, we have added additional statements on pages 5 to 10, 20 to 21, 26 to 28, 30 to 32, 39, and 53. We have also added an explicit statement acknowledging that having access to standardized ratings would have substantially strengthened the contribution of the present research, and note reasons why we weren't able to get these ratings in our present design on page 53. There may still be some examples of inaccurate language, which if noted, we would be happy to correct.

“...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: This is also an important point that was not previously discussed. We agree that we are unable to disentangle the extent to which our measured variable is a product of or precursor to emotion regulation in this design. While we intended to target emotion as a precursor and chose language trying to capture it as such, affective or regulatory assessments may have been biased or

misreported. If subjects did interpret our post-exposure questions (i.e., how intense was this experience) in different ways or if self-report values were in some way a representation of both pre- and post-regulation intensity, it may have had an effect upon the relationship between strategy usage and emotional intensity. There are two relevant pieces of information which may mitigate this concern: Specker et al., 2024 – one of the only studies to assess emotion intensity before and after regulation using distraction or reappraisal - found that pre- and post-regulation intensity scores were highly correlated ($r = 0.809$, $p < 0.001$). Additionally, we used hierarchical modeling to account for idiosyncratic subject-level effects and we believe that confusion or bias between pre- and post-regulation intensity 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). Regardless, we have updated our text on pages 20 and 53 to reflect this limitation and addressing it has improved the manuscript. We have also added the precise language used to capture self-report in the interest of transparency (pg. 17).

To attempt to address the reviewer's concerns, we have conducted two additional analyses. The first examines 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 confirmed the intended design structure (See pgs. 15, 26 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. The second analysis computes an aggregate rating of intensity for each unique observation using a leave-one-out approach. This attempts to predict regulation behavior using relatively more normalized representations of the emotion intensity of each event. With this technique, the predictor becomes the average intensity reported by all *other* subjects (i.e., excluding the 'experiencer's' self-reported intensity) experiencing that same event in each observation. This is a higher resolution approach than using section to predict regulation, but, again, no observed association between intensity and usage was observed. These analyses are of course imperfect, post-hoc solutions to this criticism, but we believe they at least add credence to the notion that our null results in Study 1 are not simply a product of lack of measurement control. These changes are covered on pages 20 and 26 to 28.

*“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 believe that the previously noted post-hoc analysis does address this precursor/product concern to some degree, but we would also like to emphasize that the result of interest in Figure 4 is not the association between intensity and success, but that at most intensity values reappraisal was reported as more successful than distraction. Finding reappraisal to be a more successful strategy at higher intensities is atypical in the context of the extant literature (Shafir et al., 2016; Sauer et al., 2016; Specker et al., 2024) whether subjects reported pre- or post-regulation intensities. We included the results of an additional analysis - a respecified version of the simple slopes model with intensity as the moderator - to hopefully make our intended point clearer. We do deviate from the previously cited work in that this success metric is subjectively assessed; we do not have pre- and post-regulation intensity to compare, which would be a relatively more objective means of measuring this. We have updated pg. 30 to reflect that. In considering this critique, we reworded our interpretation of this result. Page 30 now further emphasizes that: 1) we have no evidence to suggest that this generalizes beyond the highly specific features of this context and 2) this evidence in itself does not conclusively explain what we observed within this context.

“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 with the reviewer that this was incorrectly stated and thank them for bringing this to our attention. 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 at high intensities. As previously noted, we corrected this language, but also contextualized our finding by reminding the reader that there are important 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: We interpret the reviewer's response to suggest that we should expect an interaction between differences in pre- or post-regulation intensity by the strategy used. In other words, if the same high intensity event were to be regulated by two identical people, one using distraction and one using reappraisal, we should expect distraction to produce a lower post-regulation intensity than reappraisal. If responding to an identical low intensity stimulus, we might expect the opposite: a lower post-regulation intensity for reappraisal relative to distraction. Because our design cannot separate pre- and post-regulation intensity, instances in which subjects report pre-regulation intensity might offset instances in which subjects report post-regulation intensity, resulting in a null effect. If we have misunderstood, please let us know.

If interpreted correctly, this seems to be describing the relationship between strategy *effectiveness* at different emotion intensities. This is undoubtedly a critically important question to answer in naturalistic settings, but one that we think is outside the scope of the present work. The present work focuses on strategy usage *likelihood* at different emotion intensities. Fully addressing this point would likely require running a new set of studies wherein this question is the primary focus, more standardized stimuli that could be repeatedly used over time, the controlled and equal use of multiple strategies across participants, and potentially a more extensive longitudinal design.

We do lack the ability to conclusively explain the null with this study design and acknowledge that. As we view it, the value of reporting this null is simply that when assessing these variables in an ecologically valid way (i.e., minimizing manipulation) and in a context with features which mirror other circumstances in which self-regulation could be of vital importance (i.e., highly stimulating, high intensity, complex), we do not find this relationship. This is despite an impressively substantial effect size and consistent replication in more controlled contexts. As such, we believe that the present research helps to identify boundary conditions which should be more iteratively tested. Studies 2 and 3 contribute somewhat to this goal, but are far from exhaustive. In considering this point, we have 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 intended to more closely mirror aspects of the structure of a traditional Sheppes design. We expected to find the canonical relationship because of these additional experimental constraints. We have made this expectation clearer in the present manuscript (pgs 31 to 32) to reduce confusion. We intended to demonstrate that conceptually similar but less-complex versions of the same stimuli could elicit the expected patterns. We do not intend for Study 2 to explain the results of Study 1 (e.g., *"The different results observed in these studies are difficult to compare, though, as many features differ between the approaches..."*, pg. 38), but for them to be considered in parallel, as complementary to one another, and to situate the present work in the context of the extant work on regulatory strategy choice/usage. The comparison is imperfect (hence the need for Study 3), but, we argue, still valuable as a demonstration that manipulation is, in part, essential to the existence of this effect, which conflicts with Study 1's priority of external validity. This concern was noted by Reviewer 3 as well and our previous correspondence may be of relevance.

"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: The goal of Study 3, as well as Study 2, was to conceptually replicate Sheppes's original work rather than to "disprove" the effect. We intended to illustrate the boundaries of observing the effect by altering aspects of experimental control. We frame this as an "extension of previous work" and have changed how we discuss the study goals to further reiterate this point on page 39.

Notably, the design of Study 3 was motivated by Reviewer 3's request to try to more directly compare Studies 1 and 2. In an ideal world, Study 3 would have been conducted in another quasi-naturalistic setting (ideally the same Study 1 setting) with some relatively more controlled design decisions (*"Because we could not incorporate an immersive experiential component such as in Study 1 in this experimental design..."* pg. 39), but timing (the haunted house only runs one month of the year) and resources made this impossible. Our team decided to use video stimuli instead and to contrast forecasting and usage. The hypothesis was that regulation decisions made within the idealized simulations people engage in when forecasting (which represents almost an additional level of control) should more closely resemble the canonical relationships documented elsewhere, while actual usage and the realities that complicate regulation would make the relationship between intensity and strategy 'messier', perhaps not enough to completely dissipate the effect between intensity and usage but to mitigate it.

Ultimately, our predictions were not supported by the results of Study 3, as noted on page 39. However, we still found notable differences in line with Studies 1 and 2 which contradict expectations set by the existing literature. The categorical intensity of our stimuli was balanced between high and low intensity and yet we observed reappraisal to be used more often than was forecasted. We also found reappraisal to reduce negative emotion intensity more than distraction reduces negative emotion intensity, and that distraction reduced negative emotion less than subjects forecasted it would. We have revised our discussion of these results to emphasize that Studies 2 and 3 contained greater experimental control than Study 1.



Department of Psychology

19 August 2024

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 feedback contained within your decision letter dated May 31st, 2024 and the opportunity to resubmit our manuscript (XGE-2023-1118). We have revised our manuscript in accordance with the reviewer comment and organized those changes in the appended document. Because the reviewer only provided high-level commentary, we did our best to divide that into sub-topics and respond to those as appropriate.

We believe that our revision addresses all concerns cited by the reviewers and has certainly increased the quality of the manuscript. 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

Word Count: 14746

Emotion regulation strategy use and forecasting in response to dynamic, multimodal stimuli

William J. Mitchell^a, billy.mitchell@temple.edu *
Joanne Stasiak^b, joanne.stasiak@psych.ucsb.edu
Steven Martinez^a, stevent.martinez@temple.edu
Katelyn Cliver^c, katelyn.cliver@temple.edu
David Gregory^a, david.gregory@temple.edu
Samantha Reisman^d, Reisman@brown.edu
Helen Schmidt^a, helen_schmidt@temple.edu
Vishnu P. Murty^a, vishnu.murty@temple.edu
Chelsea Helion^a, chelsea.helion@temple.edu

- a) Department of Psychology & Neuroscience**
Weiss Hall, Temple University, 1701 N 13th St. Philadelphia, PA, USA 19122
- b) Department of Psychological & Brain Sciences**
Building 251, University of California, Santa Barbara, Santa Barbara, CA, 93106
- c) Department of Psychological & Brain Sciences**
Disque Hall, Drexel University, 3201 Chestnut St, Philadelphia, PA 19104
- d) Department of Cognitive, Linguistic, and Psychological Sciences**
Box 1821, Brown University, Providence, RI, 02912

* Corresponding author.

E-mail address: billy.mitchell@temple.edu

Address: 717 Weiss Hall, Temple University,
1701 N 13th St. Philadelphia, PA 19122

ABSTRACT (249 / 250 Words): Successful emotion regulation (ER) requires effective strategy selection. Research suggests that disengagement strategies (e.g., distraction) are more often selected than engagement strategies (e.g., reappraisal) as emotional experiences intensify. However, the extent to which ER strategy *choice* in controlled circumstances reflects strategy *usage* during complex, multimodal events is not well understood. The present research uses dynamic, multimodal stimuli (i.e., a haunted house, horror movies) to examine the association between affective intensity and regulatory strategy usage among untrained participants – individuals given no prior regulation instructions or direction. Both a preliminary study ($n = 54$) and Study 1 ($n = 118$) failed to find relationships between emotional intensity and strategy usage to downregulate emotions as participants navigated a haunted house. Distraction was self-reported to be less successful than reappraisal at high intensities, contrary to expectations. Participants in Study 2 ($n = 152$) forecasted regulation strategy usage based upon descriptions of emotionally-regulated experiences from the preliminary haunted house study. Affective intensity predicted which strategies forecasters predicted they would use; though, forecasters overpredicted how often distraction was used in practice. Study 3 ($n = 242$) incorporated strategy usage and forecasting within the same design by showing untrained participants video stimuli of varying-intensity and capturing their regulatory responses. Forecasters again predicted using distraction more often than strategy users did in practice. Forecasters also overpredicted how effectively distraction reduced negative affective intensity relative to what strategy users reported. These results may highlight a disconnect between strategy fittedness when self-regulation occurs in uncontrolled, highly-intense, or complex circumstances.

SIGNIFICANCE: How individuals manage their emotions during dynamic, complex experiences is an important question for anxiety- and trauma-based disorder prevention and treatment. This study found distraction to be less effective than reappraisal at high negative emotional intensity in these difficult-to-regulate environments, which contrasts findings from relatively more controlled studies.

CONSTRAINTS ON GENERALITY: We aimed to generalize our findings to the downregulation of negative emotions by non-clinical US populations in high-intensity, dynamic, and/or multimodal emotional situations. However, the setting of Study 1 (i.e., a haunted house) likely will not generalize to all dynamic, multimodal emotional situations. Studies 2 and 3 contain reasonably representative samples in terms of education, socioeconomic status, and race, but were collected online, thus limiting generalizability to non-optimal settings. Our recruitment added additional constraints as well. Using horror-related stimuli likely resulted in a self-selection bias. Participants in Study 1 skewed young, falling between the ages of 18 and 34, and were recruited from an urban setting, which may limit the applicability of our conclusions to children, older age groups, or individuals living in non-urban settings. Participants skewed more female than male in Studies 1 and 2 and more male than female in Study 3, which may result in an underrepresentation of men and women, respectively. Lastly, racial demographic information was not captured in Study 1, which leaves questions of racial representation unanswered. Future research should explore these phenomena in other novel settings with a wider age range and balanced gender ratio to enhance the external validity of our findings.

KEYWORDS: emotion, self-regulation, naturalistic stimuli, decision making

INTRODUCTION

Overwhelming situations can generate profoundly intense emotions which often do not match our ideal emotional states. We can try to change our emotional responses in loud or crowded spaces, unexpected confrontations, and circumstances that we lack control over, but our ability to successfully regulate our emotions in these situations may be limited. A key factor in determining regulation efficacy is the strategy someone uses or chooses. Regulation strategy efficacy can be highly context-dependent, leading many to emphasize studying emotion regulation (ER) in different contexts as the next crucial direction for the field (Aldao, 2013; Dixon-Gordon et al., 2015; English et al., 2017; Rottweiler et al., 2018; Tang & Huang, 2019).

The Process Model of Emotion Regulation. Foundational ER research (i.e., the Process Model) has identified categories of common strategies that people use to regulate their emotions (Gross, 1998). Characteristics of the regulator and the context can determine the effectiveness of these strategies (Young & Suri, 2020) and people often choose strategies to match their present circumstances (Opitz et al., 2015; Sheppes et al., 2011). For example, engagement strategies, like altering the meaning or interpretation of the emotion-eliciting stimulus (i.e., reappraisal), often have high cognitive demands. These can be contrasted against disengagement strategies, like diverting attention (i.e., distraction) or inhibiting the expression of emotion (i.e., suppression), which typically have relatively low cognitive demands (Sheppes & Gross, 2011).

How might these regulatory patterns be reflected in strategy choice? Although not explicitly directed to do so, the anticipation of a scary moment during a horror movie might prompt a person to look away from the screen (distraction), think about the actors in a different light (reappraisal), or limit the expression of their fear, all in an effort to reduce, or downregulate, an unwanted feeling. People in situations with few cognitive resources may compensate by selecting

strategies that demand less cognitive effort according to the selection, optimization, and compensation (ER-SOC) hypothesis (Opitz et al., 2012). For example, if someone were watching a horror movie while they were tired or under high cognitive load, it may be easier to look away from the screen than to generate alternative, less fear-inducing reinterpretations by which to view the film's events. Attempting to reappraise would constitute a high-risk strategy in this context because it might be less likely to work (Ford & Troy, 2019) than a comparatively low-effort strategy like distraction.

This supposition is bolstered by the especially robust influence of emotional intensity upon strategy choice, as distraction is chosen more often (Hay et al., 2015; Orejuela-Dávila et al., 2019; Sheppes et al., 2011; Young & Suri, 2020) and is more effective (Shafir et al., 2016) than reappraisal in response to high intensity stimuli, at least in part because it can be deployed before relevant emotional information has been represented in working memory (Sheppes et al., 2014; Sheppes & Gross, 2011) and requires fewer cognitive resources (Dorman Ilan et al., 2019). This effect has been thoroughly replicated in lab studies and ecological momentary assessment (EMA) studies (Colombo et al., 2020; Heiy & Cheavens, 2014). However, this association may not readily translate to more complex, demanding, and less controlled environments (Sheppes, 2020), like that which we test directly here, due to differences in how we commonly measure and manipulate ER in these studies.

External Validity in Extant Emotion Regulation Paradigms. Lab ER paradigms differ from the everyday experience of ER in a few key ways that may limit generalizability. Lab ER paradigms (e.g., Sheppes et al., 2011, 2014) usually train participants to use regulatory strategies before a task begins, which may prime more introspection and metacognition (Carver & Scheier, 1981) than what occurs in the typical ER experience. Lab ER paradigms also often necessarily

show previews of emotional stimuli to allow participants to prepare their regulatory responses, but dynamic, high-intensity events in our everyday lives are often unexpected or difficult to anticipate. Following stimulus previews, lab ER paradigms also frequently prompt individuals to select a strategy, but aversive experiences in everyday life, like sharing the road with an erratic driver or being bullied, may not explicitly prompt the implementation of self-regulation.

Such study design decisions prioritize internal validity over external validity. Self-report capture of emotion in ER studies is often either assumed based upon standardized ratings associated with the stimuli (e.g., the IAPS picture set) (Bradley & Lang, 2007), measured through unidimensional Likert scales (e.g., valence) (Shafir et al., 2016), or is captured through established measures (e.g., the Positive and Negative Affect Schedule) (Watson et al., 1988; Weiss et al., 2021). These approaches offer an efficient, reliable, and standardized means of assessing self-regulation but make assumptions about the emotional states of subjects that might not accurately reflect the complexity of emotional experience. As such, ER choice may more accurately reflect ER capacity than actualized ER behaviors (McRae & Gross, 2020). By relying upon these standardized manipulations, researchers are able to make a stronger casual claim regarding the relationship between emotion intensity and regulation, but this relationship appears complicated when the predictor is self-reported or disaggregated (Specker et al., 2024).

EMA studies – another common means of studying ER – do more directly assess the external validity of ER strategy choice relationships by capturing emotionally evocative events within the everyday lives of trained research participants (e.g., Haines et al., 2016; Heiy & Cheavens, 2014). However, most people are not trained to consider their ER strategies in their daily lives and are not prompted or primed to engage regulatory control before an emotional event occurs (Friedman & Gustavson, 2022). As such, training participants in these studies may also

introduce important but often underappreciated deviations in regulatory behaviors from how untrained counterparts might respond in the same situation.

Contemporary constructivist theories posit that emotions are not fixed reactions but are instead shaped dynamically by cognitive and social processes, encompassing one's personal interpretations, beliefs, and social interactions (Lindquist et al., 2012). Study designs that prioritize external validity promote participant- or stimulus-level idiosyncratic experiences that are pivotal to developing working generative models of how the world around us works (Lee et al., 2021). Free-response capture may require fewer a priori assumptions from researchers about a participant's emotional experience than pre-determined self-report options or standardized scales, thus reducing unintended researcher influence (Gendron et al., 2012; Lindquist et al., 2006) and improving generalizability (Miller et al., 2019). Applying natural language processing techniques to free-response data can improve our ability to capture the multifaceted, idiographic emotional experiences that constructivist frameworks emphasize (Mohammad, 2018) without sacrificing accuracy (Diamond et al., 2020) or quantifiability.

Research designs that incorporate more features of naturalistic regulation, such as not instructing or prompting participants to regulate (e.g., Heiy & Cheavens, 2014; Opitz et al., 2015) often find people explore and flexibly apply multiple strategies that blur the boundaries of typical strategy classifications (Aldao & Nolen-Hoeksema, 2013; Ford et al., 2019; Heiy & Cheavens, 2014; Opitz et al., 2015; Szasz et al., 2018). These approaches also capture meaningful variance in self-regulatory behaviors that more controlled designs cannot (Friedman & Gustavson, 2022; Kamradt et al., 2014; Malanchini et al., 2019). For example, overstimulation from complex, multimodal contexts may simultaneously be aversive and more cognitively demanding (i.e., better suited for disengagement strategies). However, attention may also be challenging to control in a

context with so much attention-demanding stimuli, reducing the likelihood of observing the high-intensity-distraction association characterized in controlled laboratory studies (Draheim et al., 2022). Spontaneous or untrained ER in these contexts may rely more heavily on person-specific features, like habits, than on how emotionally arousing the situation may be (Christou-Champi et al., 2015; Koole et al., 2015; Norem, 2008). Yet, challenging situations may be precisely when adaptive regulatory control (i.e., the pattern between intensity and regulation choice observed in most laboratory studies) may be most valuable, as maladaptive ER tendencies predict more severe manifestations of post-traumatic stress disorders (Hannan & Orcutt, 2020) and related post-traumatic stress disorder outcomes (Specker & Nickerson, 2022). Thus, identifying whether the established association between intensity and effort-related strategy usage occurs in high stress contexts is important for the development of potential interventions.

The Present Study. To our knowledge, no study has examined whether affective intensity predicts strategy usage in aversive, high-intensity, multimodal contexts using untrained participants (i.e., individuals given no prior training in how to use or classify ER strategies and no explicit direction to regulate their experiences). While affective intensity represents a particularly prominent predictor of ER behavior (i.e., $r^+ = 0.46 - 0.61$; “a very large effect size”, according to recent meta-analyses; Matthews et al., 2021), the extent to which features of a dynamic, multimodal situation could overshadow this effect when experimental control is minimized remain unclear. The goal of the present research is to examine whether these well-established regulatory patterns emerge in samples of untrained participants exposed to a dynamic, feature-rich stimuli high in emotional variability, such as a haunted house. Haunted house experiences have been used with marked success in recent research to study emotion and self-regulation (Clasen et al., 2019; Stasiak et al., 2023; Tashjian et al., 2022). While haunted houses only represent a small proportion

of the variability which emotionally-relevant experiences could materialize as, they nonetheless offer safe and consistent but high-arousal events to elicit a wide range of emotional experiences (i.e., positive and negative emotions), intensities, and responses (i.e., regulation behaviors). This variance would be difficult to generate in a complex multimodal context outside of the lab while offering the same level of safety to the participants and control to the researchers.

Because the comparison between reappraisal and distraction is nearly ubiquitous in the extant regulation literature, as noted by Heiy & Cheavens (2014), Study 1 aimed to test whether this effect extends to untrained subjects in high-intensity circumstances by having untrained participants navigate a haunted house and report their undirected emotional and regulatory behaviors in a surprise recall task immediately after and one week after exposure. This granted high levels of fidelity in capturing subjective participant emotion and experience at the expense of greater experimental control. We additionally anticipated that participants experiencing high cognitive load as a result of exposure would more often use distraction, in line with the ER-SOC hypothesis. Surprisingly, we did not find that intensity or cognitive load was associated with strategy usage. To minimize researcher bias, we then applied a multiverse approach to the data, systematically expanding and constraining our data inclusion criteria and adjusting for important covariates such as regulation tendencies and emotional goals, in an effort to identify conditions in which the relationship between affective intensity and strategy usage might emerge as statistically significant. We again failed to find an association.

This surprising finding motivated Study 2, which aimed to determine whether participants exposed to similar experiences as the haunted house but in a less dynamic, more controlled context (i.e., forecasting rather than experiencing) would more often forecast, or predict, using distraction in response to descriptions of high intensity regulated events and reappraisal in response to

descriptions of low intensity regulated events from the haunted house. We hypothesized that the decontextualized self-regulation choices motivated by stimulus-response paradigms might be closer in practice to simulating or forecasting self-regulation than regulation usage, as forecasting may be more controlled and idealized than the complicated realities of regulation usage. We did observe the canonical association between emotional intensity and regulatory strategy selection with this design, though with an effect size smaller than the typical range found by Matthews et al. (2021). However, many study design differences between Studies 1 and 2 limited our ability to make direct comparisons and draw conclusions about why those differences in outcome may have occurred.

To attempt to close this design difference gap, Study 3 tasked untrained participants with watching videos of varying negative intensity and subsequently either reporting the regulatory strategies that they used (strategy usage) or the strategies that they predict might be used (strategy forecasting) to downregulate their negative emotions. This study again featured greater experimental control than Study 1, but retained some of Study 1 design features to promote ecological validity. For both forecasters and strategy users, increasing self-reported affective intensity predicted a greater likelihood of using distraction. However, forecasters were less likely than regulators to use reappraisal at low intensities. A trend that emerged across all three studies was an inconsistency between how effective participants predicted distraction might be within these situations and how effective distraction was actually reported to be. Methods and analyses supporting these findings are discussed.

STUDY 1 METHODS

Study 1 tested whether the emotional intensity of negatively-valenced events was associated with the likelihood of using a low-effort or high-effort regulatory strategy in a dynamic,

feature-rich setting with an untrained sample of participants. To assess emotional intensity, participants performed a surprise free recall task immediately after exposure during which they self-reported the emotions they experienced via free-response and the intensity of those emotions during self-selected events from the haunted house. Participants also noted the direction of their regulation attempts (i.e., upregulation, downregulation, no regulation), how effective those regulation attempts were, and the means by which they attempted to regulate in their own words. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

PARTICIPANTS:

Preliminary Study. A preliminary study was conducted in October 2019 consisting of 54 participants (age: $\bar{x}_{age} = 24.22$ yrs, range = 18 – 34 yrs, $sd_{age} = 3.97$ yrs; gender: 26 female, 27 male, 1 non-binary; Ethnicity: 18.51% Hispanic, 81.49% Not Hispanic) who were recruited from a large northeastern city via flyers for an IRB-approved fear and memory study. This study aimed to characterize the memory of (Cliver et al., 2024) and physiological responses to (Stasiak et al., 2023) high-intensity, naturalistic events. However, this preliminary study was also designed to allow us to determine the distribution of ER strategy usage in this context and determine a more appropriate sample size for our primary study, as a review of the extant literature did not yield analogous study designs within the ER space. *A priori* power analyses for this preliminary study were conducted using the smallest effect size reported by Sheppes et al.’s 2011 examination of emotional intensity and regulatory choice ($\eta_p^2 = 0.43$). Using “WebPower” (Zhang & Mai, 2019) in R 3.6.1 (R Core Team, 2022), we determined 18 participants would sufficiently power our main effect in a typical lab context. Given the additional complications our study design introduces which would likely reduce the effect size between our variables of interest and the resources we

had available, we surmised a sample of 54 participants, three times the minimum sample size, might be sufficient to identify an effect. This analysis failed to find a statistically significant association between affective intensity and strategy usage, but our best performing multilevel binary logistic regression model [$strategy \sim intensity + (1 / Participant)$], as determined by an AIC comparison information theoretic approach, produced an odds ratio of $OR = 1.83$ ($95\% CI = [0.65, 3.2]$, $p = 0.079$) (See **Supplementary Materials** for more preliminary study design and analyses details). This observed model from the preliminary study was used to generate an *a priori* power curve for Study 1 via “simr” (Green & MacLeod, 2016) in R 3.6.1, which estimates the proportion of simulated datasets in which the null hypothesis is rejected given the target model. This approach to calculating power is preferable for hierarchically-structured data because it does not assume independence of observations. We determined that at least 76 participants with an average of 3 observations each (228 observations total) would be required to sufficiently power our study based upon simulations with the observed data (Two-tailed, $\alpha = 0.05$, $1-\beta = 0.80$, $Pr(Y=1/X=1) H0 = 0.615$). We increased the target sample size to 120 participants, due to attrition concerns and the needs of the concurrently-ran experiment on fear and memory (Cliver et al., 2024). However, we were only able to recruit 98.33% of our recruitment goal due to time constraints (i.e., all data collection must occur before the haunted house closes for the season).

Study 1. In October 2021, 118 participants (age: $\bar{x}_{age} = 20.80$ yrs, $range = 18 - 34$ yrs, $sd_{age} = 2.87$ yrs; gender: 73 female, 40 male, 5 non-binary) were recruited from a large northeastern city via flyers for an IRB-approved study on fear processing at a local haunted house attraction. A description of Study 1’s design is illustrated in **Figure 1**. Eligible participants were native English speakers between the ages of 18 and 35, had normal or corrected-to-normal vision, were not pregnant, had no history of seizures, cardiovascular issues, or neurological disorders, could

comfortably walk for at least one hour, and had not been to this haunted house in the past. On average, participants were more educated ($\bar{x}_{\text{Years of Education}} = 16.90$ yrs, $sd_{\text{Years of Education}} = 2.75$ yrs) than the average US adult, who according the U.S. Census Bureau's American Community Survey, has completed 13.7 years of education. Categorically, 81.9% reported having completed some college (58.6%), a 4-year degree (12.9%), some post-graduate studies (03.4%), or a post-graduate degree (06.9%). Socioeconomic status was slightly negatively skewed, with 14.5% of respondents reporting making less than \$15,000 per year, 07.7% reporting between \$15,001 and \$25,000, 07.7% reporting \$25,001 to \$35,000, 05.1% reporting between \$35,001 and \$50,000, 22.2% reporting between \$50,001 and \$75,000, 12.8% reporting between \$75,001 and \$100,000, 17.9% reporting between \$100,001 and \$150,000, and 12.0% reporting greater than \$150,000. The racial and ethnic identities of participants were not assessed. Participants were compensated \$60.00 in Visa debit cards.

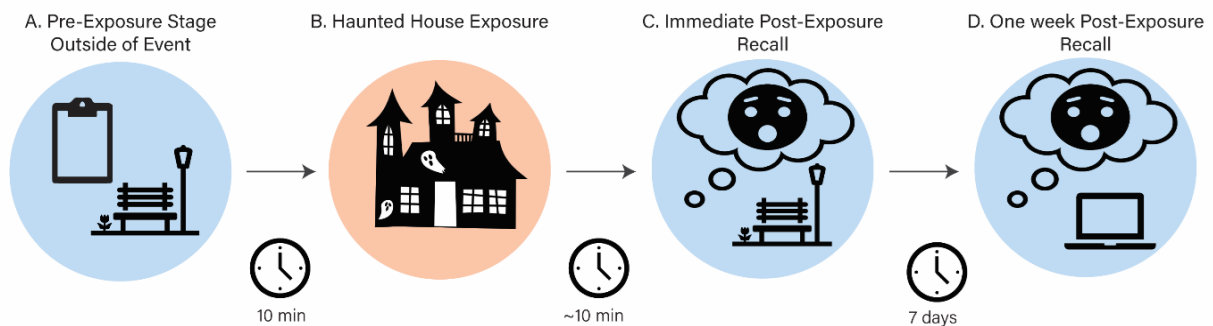


Fig 1. Study 1: Task Overview - One hundred and eighteen (118) participants traversed a haunted house in small groups. **(A)** Prior to the haunted house, participants completed baseline questionnaires outside of the event at a local park. **(B)** The haunted house lasted for ~37 minutes. **(C)** Participants then immediately recalled three events, and their attempts to regulate them post exposure. **(D)** They then again recalled the same three events and an additional six events at an online follow-up session.

MATERIALS AND PROCEDURE:

Pre-Exposure. Participants reported to a provisional headquarters (i.e., tent, tables, chairs, computers, etc.) constructed just outside the haunted house property to complete individual difference questionnaires, questionnaires assessing prior knowledge of the haunted house, expectations, and motivations for participating, as well as a measure of cognitive load. To assess prior knowledge, participants indicated the number of times they had attended other haunted houses, how much information they felt they knew about this specific haunted house, and from which sources had they learned information about this specific haunted house (e.g., advertisements, friends, news reports, etc.). To assess expectations, participants were asked how much positive and negative emotion they anticipated experiencing prior to exposure on a 5-point Likert scale. Labels for the 5-point Likert scales included “None at all”, “A little”, “A moderate amount”, “A lot”, and “A great deal”. Participants were also asked how fearful they felt, how sensitive they were to startling stimuli, how much they enjoy haunted houses, and how much they enjoy feeling fear (all assessed on Likert scales). Motivations to participate were assessed using 100-point sliding scales and included payment, thrills, novel experiences, challenges, social pressure, scientific interest or duty, and boredom. The poles and central position were labelled as “Not at all motivating”, “Moderately motivating”, and “Extremely motivating”. Cognitive load was assessed prior to exposure, immediately after exposure and at a later follow-up using a 15-item Remote Associates Test (RAT). Forty-five RAT items were selected for their difficulty as measured by Bowden’s 15-second trials, such that each item had two equally difficult counterparts which could be randomly assigned across the three timepoints (Bowden & Jung-Beeman, 2003). Following instructions, participants completed three practice trials with feedback. During the RAT task, participants had 15 seconds to identify the target word and did not receive feedback. Participants were then fitted with heartrate monitors and escorted to the haunted house entrance.

All pre-exposure questionnaires and materials are available within our OSF repository (*See **Open Practices***).

Exposure (Haunted House). Participants experienced the haunted house in 31 groups across 11 nights ($\bar{x}_{size} = 3.81$ participants; $sd_{size} = 1.12$ participants). This specific haunted house was chosen because: 1) it uses professional actors renowned for eliciting a range of affective responses; 2) it contained four themed sections designed to be either low- or high-intensity, each with a unique aesthetic providing variability to the stimuli; 3) it provides a remarkably consistent experience across sessions; and 4) coordination with the facility granted us special privileges to use equipment (i.e., computers, heart-rate monitors) and better guarantee consistency across exposures (i.e., entering before other guests, keeping groups together).

Participants received minimal instructions to promote ecological validity. Participants were to walk through the haunted house in a single file line and avoid sharing thoughts, reactions, and experiences with other participants. Limiting communication prevented participants from skewing each other's memories and perceptions of the events before they could be recorded. Participants were encouraged to act and react as naturally as possible without interacting with each other (i.e., grabbing, holding, touching, etc.). While verbal and physical social support are recognized and effective forms of emotion regulation, they were beyond the purview of this study's focus, which was based upon common Process Model strategies. However, perceptions of social support were captured after exposure to assess their potential effect upon our outcomes of interest. As part of a concurrently-run memory experiment, some participants ($n = 58$) did receive additional instruction to navigate specific sections as if they would later have to complete a memory test or write a review of that section. However, these groups did not statistically differ significantly from the control group in reported affective intensity ($F(2,81) = 1.41, p = 0.25$), regulation strategy usage ($F(2,73)$

= 0.05, $p = 0.95$), or regulation strategy success ($F(2,81) = 1.93$, $p = 0.15$) as determined by ANOVA. The data relevant to this memory experiment is beyond the purview of this study and is better captured within another forthcoming manuscript. The accompanying research assistant led the group through each section. The approximate exposure time was 37.4 minutes.

Post-Exposure. Following exposure, participants completed immediate post-exposure assessments at a remote site outside of the haunted house. Participants were tasked with identifying three emotionally salient events that occurred from a randomly selected haunted house section and reporting affective and regulatory details of that event. Events from other haunted house sections were not tested during the immediate exposure session to avoid conflicts with the aforementioned concurrently-running memory experiment. Participants were not trained in ER strategies, nor were they primed to consider their ER strategies prior to these questionnaires. Though experiential sampling methods (ESM) and ecological momentary assessments (EMA) are often applied in naturalistic settings to capture moment-to-moment fluctuations in emotion and regulation with considerable success (*See* Colombo et al., 2020; Shahane et al., 2023), we chose post-exposure assessments as our means of assessment in part for the following reasons: 1) assessing participant reactivity mid-exposure can lead participants to alter their affective and regulatory behaviors in response to the assessment prompts (Stone et al., 2003), 2) training participants to use ESM or EMA technology effectively in this context would require training in ER strategy categorization in violation of the goals of this study, and 3) the use of such technology during exposure would violate the immersive, high-intensity nature of the context (Shiffman et al., 2008). Furthermore, studies which have used cued recalls to capture ratings of emotion experiences after exposure have found high consistency between experience and recall (Hutcherson et al., 2005; Levenson et al., 1983). For each of the three events, participants wrote a detailed description of what occurred.

Participants then noted which emotions (up to a total of five) they felt during this event via free response. Participants rated how intense their emotion felt ("During [this event], how intense was the [emotion] you experienced?") on a 5-point Likert scale (labels included, in order of increasing intensity: 'None at all', 'A little', 'A moderate amount', 'A lot', and 'A great deal'). Participants were also tasked with describing how they tried to regulate each emotion, if at all ("Did you attempt to reduce or intensify this [emotion]?"), via free response ("Please describe how you attempted to reduce or intensify this [emotion]"). People's accuracy in recalling details of similar real-world experiences via free response was surprisingly high in recent investigations using similar methodology (Diamond et al., 2020). Participants were also asked directly whether they attempted to down- or up-regulate their experiences, how successful their regulatory efforts were, and regulatory responses were assessed in response to each emotion rather than each event. Thus, rather than exploring this phenomenon 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. We refer to data captured at this time point as being "immediately reported". Following completion of immediate post-exposure measures, participants were dismissed, instructed to not discuss their experiences, and to remain in contact with researchers for a follow-up session which was conducted remotely (*time since exposure: $\bar{x}_{delay} = 7.01$ days; $sd_{delay} = 0.91$ days*) to assess how memory of self-regulation and memory accuracy in this context alters over time. At this delayed follow-up session, participants were tasked with identifying six additional events and their affective and regulatory responses to each.

Strategy Usage Coding. During the preliminary study, two hypotheses-blind raters classified strategy descriptions into one or more strategy categories: Reappraisal, Distraction, Suppression, Situation Selection, Situation Modification, or ‘None of the above’ (IRR Agreement = 0.880). Raters were also provided the participant’s description of the event and the emotions they experienced which they indicated having downregulated. Raters were undergraduate research assistants who trained by first reviewing examples of landmark literature which defined the strategies of interest as commonly used in the field (Gross, 1998, 2002). Raters then reviewed select methodological excerpts from experimental papers to see how cognitive reappraisal, attention deployment, and other Process Model strategies were defined within past studies (Shafir et al., 2016; Sheppes et al., 2011). Lastly, raters independently completed classification exercises using examples of regulation strategy descriptions from the same context but which were collected during the preliminary study. Through the training and classification process, raters were instructed not to collaborate or discuss their ratings with each other during the rating process. After individually classifying each description, a moderator (author W.M.) compared the ratings for disagreements (i.e., cases in which raters disagreed on how a regulation event should be classified). The moderator then met with both raters remotely using a digital video conferencing platform and moderated a review of the classifications, asking raters to compromise in cases of classification disagreement. The moderator’s role was to facilitate discussion of classifications and document their conclusions, but was not involved in the discussion and disconnected during them (i.e., muted their microphone; turned off camera) to avoid unduly influencing the outcome.

This approach revealed that distraction and reappraisal were by far the most commonly used strategies in this context. Of the 182 self-reported downregulated events recorded in the preliminary study, 30.7% used reappraisal and 61.5% used distraction, with the other three

strategies (i.e., suppression, situation modification, situation selection) combined appearing in fewer than 20% of events (Note that the total percentage sum is greater than 100% due to the occurrence of multi-strategy experiences). Two hypotheses-blind raters classified each observation's strategy description in Study 1 into one or more strategy categories: Reappraisal, Distraction, Suppression, a combination of the three, or none of the above (IRR Agreement = 0.877). Situation Modification and Selection were excluded due to the infrequency of their use. Suppression was also used infrequently in Study 1, but was categorized by Study 1 raters due to its large presence within the broader literature. Raters were undergraduate research assistants who were trained using the same methodology described in the preliminary study, but were not the same raters from the preliminary study. The training materials and instructions generated for this purpose have been made available within our OSF repository (*See Open Practices*).

Event & Location Coding. Prior to launching our study, a key was generated by research staff documenting each of the unique events contained within the haunted house, including detailed descriptions, their temporal order, and the approximate locations of each event. This key was used as a reference by two additional hypothesis-blind independent raters who had not experienced the haunted house. They identified the approximate locations of events reported by participants using the detailed descriptions that participants provided. Each event was labeled as occurring either within one of the four sections or as “not applicable” in cases of high ambiguity. Descriptions of the same event experienced by different subjects were identified and binned together. The training and rating process otherwise mirrored the procedure outlined for strategy usage coding. Agreement between raters was high (IRR = 0.918).

Questionnaire response processing and coding. Emotion responses were processed by: 1) removing entries lacking intelligible affective information (e.g., “-“, “nothing”, “idk man”), 2)

removing unnecessary punctuation, hyphenation, and qualitative modifiers (e.g., “very sad” becomes “sad”, 3) splitting compound emotion response (e.g., “sad / angry” becomes “sad” and “angry”, 4) correcting spelling errors according to the top suggestions recommended by R’s native spell checking software, 5) lemmatization (e.g., “annoyance”, “annoying”, and “annoyed” become “annoy”). These modified emotion responses were then merged with the NRC lexicon which contains over 20,000 English emotion words human rated by valence, arousal, and dominance (Mohammad, 2018). Valence was determined using NRC lexicon valence scores. Observations without an associated NRC lexicon entry were dropped due to lack of valence data.

Analysis. To explore our primary question, the effect of emotional intensity upon regulatory strategy usage, we specified mixed effect binary logistic regressions accounting for the random effect of participants using the “lme4” package (Bates et al., 2015) in R (R Core Team, 2022) and followed an information theoretic approach via AIC comparison, which prioritizes model parsimony and penalizes models with excessive variables.

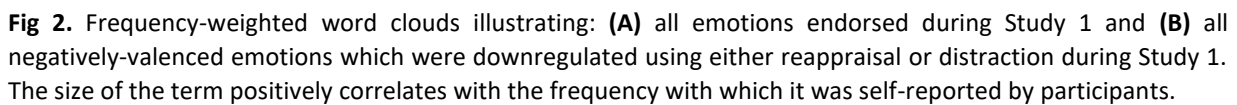
This observational approach, using the idiographic self-reported emotional intensity of each subject to predict regulation behaviors, differs substantively from its experimentally-controlled predecessors, which instead generally manipulate emotion using standardized stimuli sets and use the standardized intensity of each stimulus as the predictor of strategy choice rather than the intensity subjects experienced. Such an approach assumes that, over a sufficiently large sample, the average experienced intensity of any given stimulus will match the standardized value for that stimulus. In exchange for making this assumption, researchers likely reduce potential confounds and are better able to address causality. By taking the observational idiographic approach that we had, we assume that the personal experiences reported represent emotion pre-regulation, though it may be the case that these reports were also influenced by post-regulation

emotion. While pre- and post-regulation emotion intensities appear highly correlated (Specker et al., 2024), important differences may be obfuscated. By not making the same assumptions as more experimentally controlled studies, our analytic approach seeks a pragmatic understanding of the relationship between intensity and emotion regulation, accepting potential confounds as they may naturally cooccur and aiming to determine whether the relationship exists without the degree of experimental control that prior studies established.

All data and scripts used to produce this analysis are publicly available at OSF (*See **Open Practices***). Preregistration for Study 1 methods and hypotheses is publicly available at As Predicted (https://aspredicted.org/DP1_453).

STUDY 1 RESULTS

A subset of 298 observations in which a negative emotion was downregulated by either distraction or reappraisal was used for analyses related to our primary hypothesis. These observations were reported by 77 participants. Of the 298 total observations, 175 (58.72%) reported using distraction to regulate their emotions. The average emotional intensity of observations was 2.41 ($range = 0 - 4$, $sd = 0.932$). **Figure 2b** illustrates the variance in all emotions and **Figure 2a** illustrates the emotions endorsed within the dataset of 298 observations.



Thrill-seeking motivated study participation. Although horror attractions may feature aspects of emotional experiences that are difficult to capture in more traditional paradigms, they may have limitations as well. For example, such a setting may generate self-selection biases in our sample. To monitor this, participants were asked about their expectations and motivations to participate in this study. Participants were asked how much they enjoyed fear on a 0 to 6 scale, with the average response sitting above the midpoint of the scale ($\bar{x} = 3.40$, $median = 4$, $sd = 1.82$). Participants were also asked, “To what extent are the following items motivating your participation?” on a 0 to 100 scale with items including: the payment received for participating ($\bar{x} = 59.2$, $median = 59$, $sd = 31.1$), the thrills they may feel in the haunted house ($\bar{x} = 63.3$, $median = 73$, $sd = 31.7$), the opportunity for new experiences ($\bar{x} = 70.1$, $median = 78$, $sd = 28.0$), the opportunity for challenging experiences ($\bar{x} = 53.3$, $median = 50$, $sd = 30.3$), social pressures ($\bar{x} = 15.5$, $median = 02$, $sd = 25.7$), a desire to help science ($\bar{x} = 59.4$, $median = 58$, $sd = 29.9$), and boredom ($\bar{x} = 28.0$, $median = 15$, $sd = 31.4$). Thrill-seeking motivations strongly correlated with fear enjoyment ($r = 0.582$, $p < 0.001$) and Bonferroni-adjusted contrasts did determine that thrill-seeking as a motivation was significantly greater than boredom ($t(144) = 7.17$, $p < 0.001$), social

pressure ($t(144) = 9.70, p < 0.001$), and the pooled average of all motivations ($t(504) = 4.19, p < 0.001$), but not challenge-seeking ($t(144) = 2.05, p = 0.289$), novelty-seeking ($t(144) = -1.38, p = 1.000$), payment ($t(144) = 0.84, p = 1.000$), or science participation ($t(144) = 0.81, p = 1.000$). This may suggest that our participants were slightly higher in thrill-seeking motivations relative to what we might expect to find in an average population.

Study participants regulated their emotions without prompt. Another limitation may be a lack of motivation to self-regulate. Because haunted houses are entertainment, participants may be less inclined to self-regulate, regardless of which strategy they use. Within the preliminary study, we found a positive association between state anxiety measured immediately before the haunted house and the average extent to which participants attempted to regulate their emotions (assessed through the question, “To what extent did you attempt to change or regulate how you felt during this event?”) across events ($\beta = 0.34, 95\% CI = [0.054, 0.626], p = 0.021$), as well as a positive association between the average negative affective intensity of an event and extent to which participants attempted to regulate their emotions during that event ($\beta = 0.31, 95\% CI = [0.17, 0.46], p < 0.001$) (See **Supplementary Materials** for more preliminary study design and analyses details). Both suggest, as expected, that participants within this setting experiencing negative emotions do self-regulate these emotions without researcher-prompting. Self-regulation extent or efforts were not assessed within Study 1 to reassess this relationship.

Intensity did not predict regulatory strategy usage. To test our primary hypotheses, models using either z-scored emotional intensity or person-centered emotional intensity as the primary predictor were constructed iteratively (i.e., adding one fixed effect to the model at a time), but across all model comparisons, no model performed better than our null model which did not feature any fixed effects ($ICC = 0.40$). Our best performing non-null model -- including only

intensity as a fixed effect ($p = 0.10$ when compared to null) -- did not find a relationship between emotional intensity and strategy usage ($OR = 1.36$, $95\% CI = [0.95, 1.95]$, $p = 0.10$) (**Fig. 3**). An odds ratio of 1.36 suggests that for every one standard deviation unit increase in emotional intensity, the odds of choosing distraction to regulate increase by approximately 36%, but importantly, this association is not statistically significant and markedly smaller than what might be observed in more controlled strategy selection paradigms. As such, we did not find strong evidence to support that emotional intensity predicts strategy usage in this dynamic, high-intensity setting.

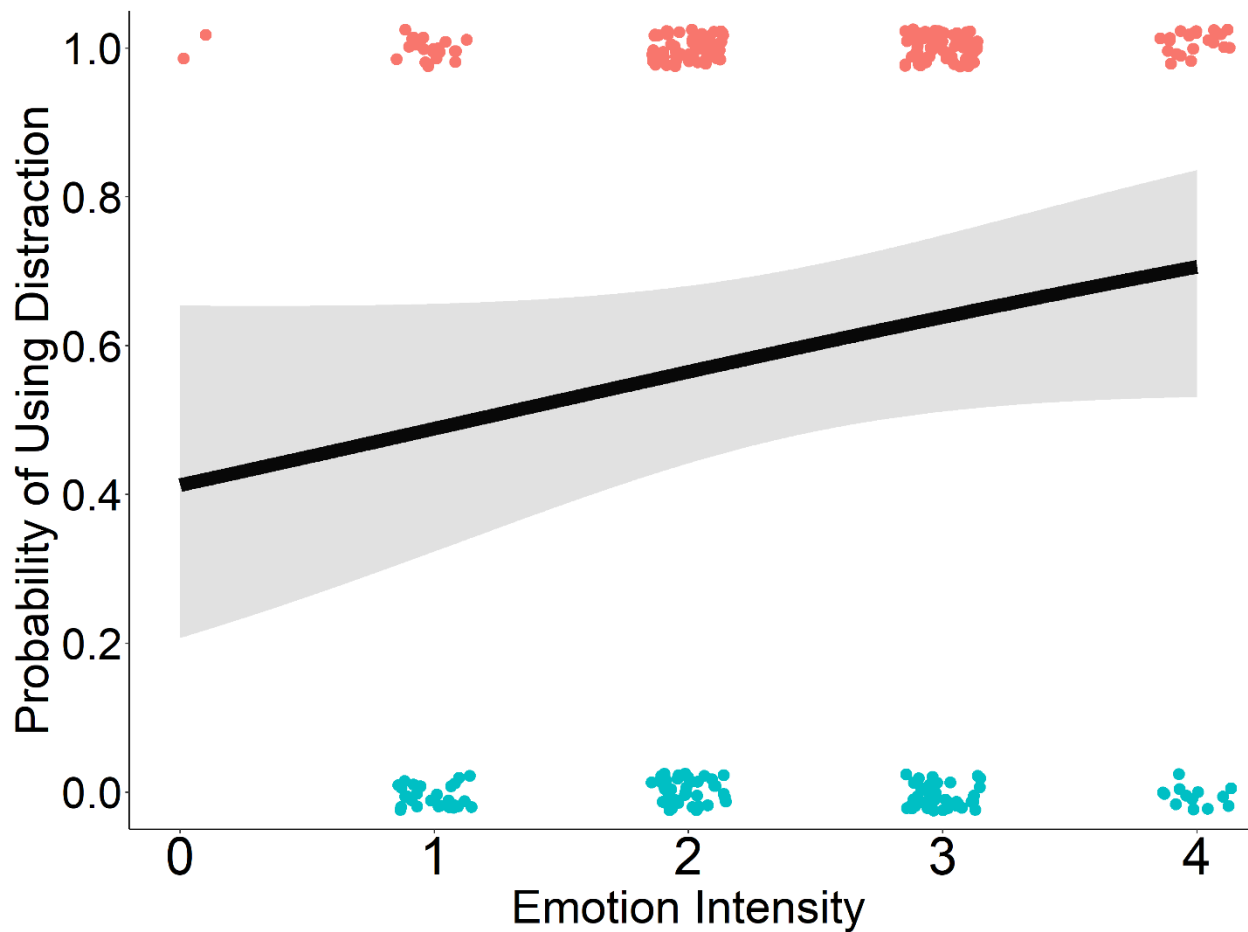


Fig 3. Across almost all tested mixed effects binary logistic regression models, emotional intensity failed to predict strategy usage. Visualized is our model using only emotional intensity ($\bar{x} = 2.41$, $sd = 0.932$) to predict regulation strategy choice between reappraisal and distraction among negative emotions. Regression line represents likelihood of selecting distraction as opposed to reappraisal at any given emotional intensity value. Points represent individual observations. Regression ribbon represents standard error.

Multiverse approach also failed to explain strategy usage. We expanded the scope of our primary analyses and conducted additional exploratory analyses to determine whether a stronger association between strategy usage and affective intensity could be found using different inclusion criteria, comparing engagement strategies (i.e., reappraisal) to disengagement strategies (i.e., suppression, distraction) as defined in the broader literature (e.g., Dixon-Gordon et al., 2015) and including contra-hedonic regulation activity (i.e., downregulating positive emotion). We found only a single model which surpassed nominal statistical thresholds of significance in model fit

($ICC = 0.37$; $p = 0.04$ when compared to null), but which did not maintain significance after adjusting to maintain a family-wise error rate ($p = 0.32$). Details and results of this analysis can be found in the **Supplementary Materials**.

Manipulated emotion intensity did not predict regulation strategy usage. Although we lack explicit standardized ratings for each event within the haunted house, each section was designed to either elicit low- or high-intensity emotional reactions. As such, by specifying the intensity category of the section in which each event occurred as a predictor predicting regulation outcome, we were able to conduct another exploratory analysis with a design that imperfectly-but-more-closely resembled the emotion manipulation exhibited in lab studies. A paired t-test - using ratings of fear that subjects self-reported on a 5-point numeric scale after each section during exposure - confirmed that subjects experienced the high-intensity sections with significantly more fear than the low-intensity sections ($\bar{x}_{high} = 3.38$ $\bar{x}_{low} = 1.90$, $95\% CI = [1.35, 1.61]$, $t(235.0) = 22.9$, $p < 0.001$). The same trend was observed when using the self-reported intensity of individuals emotions reported by subjects in a Welch's t-test, though it fell short of statistical significance ($\bar{x}_{high} = 2.41$ $\bar{x}_{low} = 2.18$, $95\% CI = [-0.02, 0.48]$, $t(47.6) = 1.82$, $p = 0.075$). Using a Pearson's Chi-squared test with Yates' continuity correction, we again found no association between the categorical intensity of the section in which each event occurred and the regulation behavior that followed each event ($\chi^2(1, N = 231) = 0.002$, $p = 0.964$). Among events in low-intensity sections, forty percent (40%) were regulated via reappraisal and sixty percent (60%) were regulated via distraction. Among high-intensity section events, 42.7% of events were regulated via reappraisal and 57.3% of events were regulated via distraction. Although this approach is low in resolution, it at least suggests that this null relationship is not simply because our hypothesized predictor (i.e., self-reported emotion intensity) is a product of, rather than a precursor to, self-regulation.

Aggregate emotion intensity did not predict regulation strategy usage. One means of approximating normative ratings for events in the haunted house may be to take a ‘leave-one-out’ approach to emotion intensity. With this technique, the downregulation behavior of each observation is regressed upon the average negative emotion intensity reported by all other participants who experienced and reported that same event. Excluding each subject’s self-reported emotion intensity yields a predictor which is independent from the outcome, mirroring the relationship between standardized stimulus ratings and emotion regulation choice in more controlled settings. Two-hundred ninety (290) observations were described well enough by subjects to be reliably categorized by our independent raters as one of 36 unique events that subjects experienced within the haunted house. Twenty-four (24) of the 36 unique events were reported by two or fewer participants and excluded from this analysis to ensure enough observations for a mean rating to be calculated ($\bar{x}_{obs} = 9.92$, $median_{obs} = 8.0$, $sd_{obs} = 6.76$, $max_{obs} = 22$). We conducted an additional analysis excluding any event with fewer than five observations as well to be more conservative in our aggregation (events = 7, $\bar{x}_{obs} = 14.3$, $median_{obs} = 13.0$, $sd_{obs} = 5.5$, $max_{obs} = 22$). Because emotions captured in our study were multidimensional (i.e., subjects could report intensities for multiple emotions for each event), we also took two approaches to compute aggregate ratings: using the average of all emotion intensities that a subject reported for any given event, and using the maximum of all emotion intensities that a subject reported for any given event. At a minimum of three observations per event, a mixed effects binary logistic regression did not find ratings computed using either approach to predict regulation usage (max: $OR = 1.20$, $95\% CI = [0.21, 7.03]$, $p = 0.96$; mean: $OR = 1.43$, $95\% CI = [0.17, 12.40]$, $p = 0.96$). A null effect was also observed using a more conservative five-observation minimum (max: $OR = 0.52$, $95\% CI = [0.06, 15.30]$, $p = 0.96$; mean: $OR = 0.91$, $95\% CI = [0.04, 23.8]$, $p = 0.96$). This

exploratory analytic approach is of a higher resolution than our analysis using section as a predictor and came to the same conclusion as our other analyses. However, this approach may also be underpowered due to excluding ambiguous observations and events with too few analogous observations.

Cognitive load did not predict regulatory strategy usage. The association between strategy choice and cognitive load has been well documented in the literature (Dorman Ilan et al., 2019), but attempts to replicate it beyond lab settings have been limited. Given the high-intensity, often overwhelming nature of this setting, we had preregistered the hypothesis that cognitive load should demonstrate a positive association with the probability of using distraction as an ER strategy. However, we failed to find evidence that cognitive load post-exposure -- as assessed by the RAT test -- predicted strategy usage during exposure ($b = -0.02$, 95% $CI = [-0.044, 0.010]$, $p = 0.21$), even when adjusting for baseline cognitive load as assessed both prior to exposure and one-week later.

Nuisance variables did not predict regulatory strategy usage. In addition to our primary analyses, we evaluated the impact of several nuisance variables. These are variables that, according to existing research, might confound the relationship between affective intensity and strategy usage, but were not of primary theoretical focus for these experiments. Examples measured in this study include emotion expectations (Denny et al., 2014), motivations to participate (Tamir, 2016), attitudes towards fear and haunted houses (Argyriou & Lee, 2020), participant sex (McRae et al., 2008), age (Blanchard-Fields et al., 2004), depression (BDI-II), anxiety (STAI), intolerance of uncertainty (IUS) (Aldao et al., 2010), regulation tendencies (ERQ) (Gross & John, 2003), time of day, and presence of peers.

We did not find significant associations between the proportion of events in which distraction was used and how positively ($b = 0.035$, 95% $CI = [-0.032, 0.102]$, $p = 0.30$) or negatively ($b = 0.047$, 95% $CI = [-0.017, 0.112]$, $p = 0.14$) participants expected to feel during the study. We also did not find an association between distraction versus reappraisal usage and the motivations cited for participation, including payment ($b = -0.000$, 95% $CI = [-0.003, 0.002]$, $p = 0.629$), thrill-seeking ($b = -0.000$, 95% $CI = [-0.003, 0.002]$, $p = 0.801$), novelty-seeking ($b = 0.001$, 95% $CI = [-0.002, 0.004]$, $p = 0.454$), peer influence ($b = 0.001$, 95% $CI = [-0.002, 0.003]$, $p = 0.595$), boredom ($b = -0.001$, 95% $CI = [-0.004, 0.001]$, $p = 0.341$), contributing to science ($b = 0.000$, 95% $CI = [-0.002, 0.002]$, $p = 0.858$), and seeking a challenge ($b = 0.000$, 95% $CI = [-0.003, 0.002]$, $p = 0.935$). We additionally did not find any associations between distraction usage and how much participants self-reported enjoying fear ($b = -0.012$, 95% $CI = [-0.050, 0.027]$, $p = 0.55$) or haunted houses ($b = 0.021$, 95% $CI = [-0.024, 0.066]$, $p = 0.35$), participant sex ($t(73) = 1.54$, 95% $CI = [-0.029, 0.225]$, $p = 0.13$), age ($b = 0.000$, 95% $CI = [-0.024, 0.023]$, $p = 0.98$), depression ($b = -0.004$, 95% $CI = [-0.024, 0.016]$, $p = 0.70$), anxiety ($b = 0.000$, 95% $CI = [-0.009, 0.008]$, $p = 0.97$), intolerance of uncertainty ($b = 0.000$, 95% $CI = [-0.005, 0.006]$, $p = 0.92$), tendency to use reappraisal ($b = 0.000$, 95% $CI = [-0.012, 0.013]$, $p = 0.97$) or suppression ($b = 0.002$, 95% $CI = [-0.011, 0.015]$, $p = 0.76$), time of day ($F(2,73) = 0.04$, $p = 0.96$), presence of familiar peers ($t(60) = -0.40$, $p = 0.700$), or which group participants traversed the haunted house with ($F(30,45) = 0.93$, $p = 0.57$). Despite the lack of association, these variables were included as covariates in some models as part of our subsequent multiverse analysis.

Regulatory strategy usage and intensity interact to predict regulatory success.

Following our emotional intensity analyses, we explored how strategy usage moderated the relationship between intensity and success, as using distraction during high-intensity emotional

states should more successfully regulate emotions than using reappraisal (Sheppes et al., 2011). After constructing a series of multilevel linear models and again following an information theoretic approach, we found that our best-performing model did indeed include an interaction between strategy usage and emotional intensity ($ICC = 0.42$, $p = 0.003$) and found that interaction to be significant ($\beta = 0.25$, $95\% CI = [0.09, 0.42]$, $p = 0.003$). However, a simple slopes analysis revealed a surprising finding: no relationship was observed between regulatory success and emotional intensity for events regulated via reappraisal ($\beta = -0.03$, $95\% CI = [-0.16, 0.10]$, $p = 0.70$), but regulatory success was *negatively* associated with emotional intensity for distraction-regulated events ($\beta = -0.28$, $95\% CI = [-0.40, -0.16]$, $p < 0.001$) (**Fig. 4**). When respecifying the simple slopes analysis to designate intensity as the moderator, we find that reappraisal was self-reported as significantly more successful than distraction at high ($\beta = -0.676$, $95\% CI = [-0.975, -0.377]$, $p < 0.001$) and average ($\beta = -0.447$, $95\% CI = [-0.667, 0.227]$, $p < 0.001$) intensities, but not low ($\beta = -0.218$, $95\% CI = [-0.508, 0.072]$, $p = 0.141$) intensities. In traditional, experimentally controlled paradigms, we often observe distraction to be more effective than reappraisal at high emotion intensity (Sauer et al., 2016; Specker et al., 2024). This data seems to document a deviation - that distraction appeared to be less – not more – successful than reappraisal - from this pattern when utilizing an observational, idiographic methodology in a high-intensity, quasi-naturalistic setting, though an important caveat is that we are capturing *subjective* assessments of success and not assessing success via measurement of intensity attenuation.

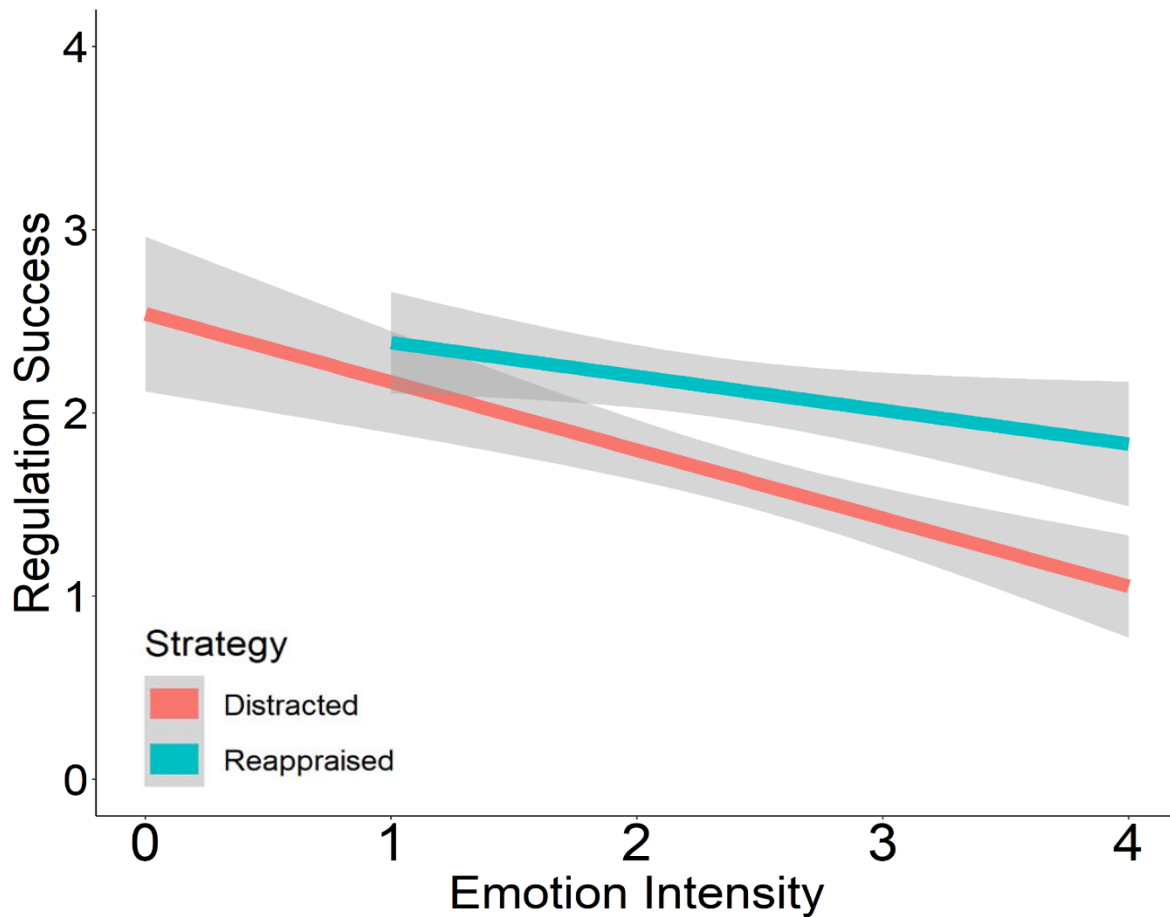


Fig 4. Strategy moderated the relationship between emotional intensity and regulatory success ($\beta = 0.25$, $p = 0.003$). While the success of reappraisal was relatively unrelated to emotional intensity, distraction demonstrated a negative association with emotional intensity, contrary to what extant literature might suggest. Given the frequency with which distraction was reported, the relative underperformance of distraction at high emotional intensities may partially explain the absence of an association between strategy choice and emotional intensity within our study.

STUDY 2 METHODS

A preliminary study and Study 1 both failed to find an association between affective intensity and strategy usage in a complex, high-intensity, multimodal environment using untrained, undirected participants, even when adjusting for noted moderators like cognitive load. Exploratory models found that the relationship may appear, albeit weakly, when including positive and negative emotions, but the effect did not survive adjustments for multiple comparisons. We also found that distraction may have been less successful at regulating higher intensity emotions than

reappraisal in this specific circumstance – a surprising finding given the extant literature. However, lack of experimental control limits interpretations of this null effect. We theorized that participants exposed to versions of the same stimuli could demonstrate ER patterns in line with the extant literature if made less complicated and given greater experimental control. Although using audiovisual recordings from the experience would have been ideal, we were unable to obtain permission to record such data during the previous studies. Instead, we generated decontextualized representations of preliminary study subjects' self-reported idiographic experiences with only the relevant information present (i.e., description of event, emotions felt, intensity of emotions) and asked a new sample of participants to simulate or forecast how they might self-regulate in the same circumstance given the information provided (events available within OSF repository). These subjects had no access to the regulatory behaviors subjects actually used. This decontextualized manipulation of emotional experiences in scenarios unburdened by the complications of reality more closely mirrors the design of a stimulus-response paradigm while retaining content similar to the haunted house. If there is a difference between participants simulating (i.e., strategy forecasting) self-regulation and participants performing (i.e., strategy using) self-regulation, it may further emphasize the effects that complexity and lack of control introduces to the ER space. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

PARTICIPANTS: In July 2021, 170 participants (age: $\bar{x}_{age} = 34.34$ yrs, $range = 18 - 75$ yrs, $sd_{age} = 14.31$ yrs; gender: 100 female, 68 male, 2 non-binary) consented to an IRB-approved online study described as measuring individual differences in choice predictions. Participants completed the study on Qualtrics and were recruited/filtered via Prolific. Eligible participants were native English speakers residing in the US between the ages of 18 and 85, had normal or corrected-

to-normal vision, had no history of reading-related disorders or literacy difficulties, as well as no history of mild cognitive impairment, head injury leading to unconsciousness, or unregulated mental health diagnosis. The racial identity of participants were as follows: 13.6% Asian, 06.8% Black, 04.3% Mixed, 03.7% Other, and 71.6% White. Although socioeconomic status data is not available, 45.2% of participants reported working full-time, 19.2% reported working part-time, 24.7 % reported not working full- or part-time, and 11.0% did not specify their work status.

Sample size was determined via *a priori* power analyses assuming an attrition rate of 10%, $r^2 \leq 0.10$ for covariates, and a small effect size ($OR = 1.68$), which suggested we must recruit at least 163 participants to achieve $1-\beta = 0.80$ ($\alpha = 0.05$, two-tailed); however, this approach had not taken into account the hierarchical nature of our observations and likely underestimates our true power. Eighteen participants were excluded for failing attention checks ($n = 7$), failing to complete the study ($n = 9$), and scoring a Q Recaptcha Score lower than 0.7, indicating significant bot activity ($n = 2$). Participants were paid at a rate of \$10.25/hr.

MATERIALS AND PROCEDURE: Details from seventy-eight (78) negatively-valenced preliminary study events regulated through either reappraisal or distraction were presented to online participants. Participants first read definitions of both reappraisal (thinking about the experience in a way that reduces the intensity of the negative emotions) and distraction (looking or thinking about something else that is emotionally neutral) and reviewed examples of how both strategies might be employed. These haunted house-specific examples included: a.) making an effort to remind oneself that the people are just actors who are using props, rather than zombies trying to hurt them, and b.) choosing to look down at one's feet or focusing on what one ate for lunch rather than focusing on the zombies coming after you, for reappraisal and distraction specifically. Participants performed a brief practice task which required successfully defining and

applying both categories before the primary task began (**Fig. 5**). Participants were provided an opportunity to pause participation and contact research staff if they had questions about definitions or strategy application before proceeding. All 78 events were randomized and serially presented. These events were screened for information pertinent to reappraisal and distraction that may unduly influence participant decisions. For each event, the emotions experienced, the intensity of each emotion, how the original regulation strategy user described the event, and definitions for both strategies were displayed. Participants were then asked to predict which strategy they would choose to reduce the emotional intensity of the situation. Participants were granted as much time as needed to complete the task. Following the primary task, participants completed individual difference measures, including the Emotion Regulation Questionnaire (ERQ), the Difficulties with Emotion Regulation Survey (DERS), and the Intolerance of Uncertainty Scale (IUS). Participants

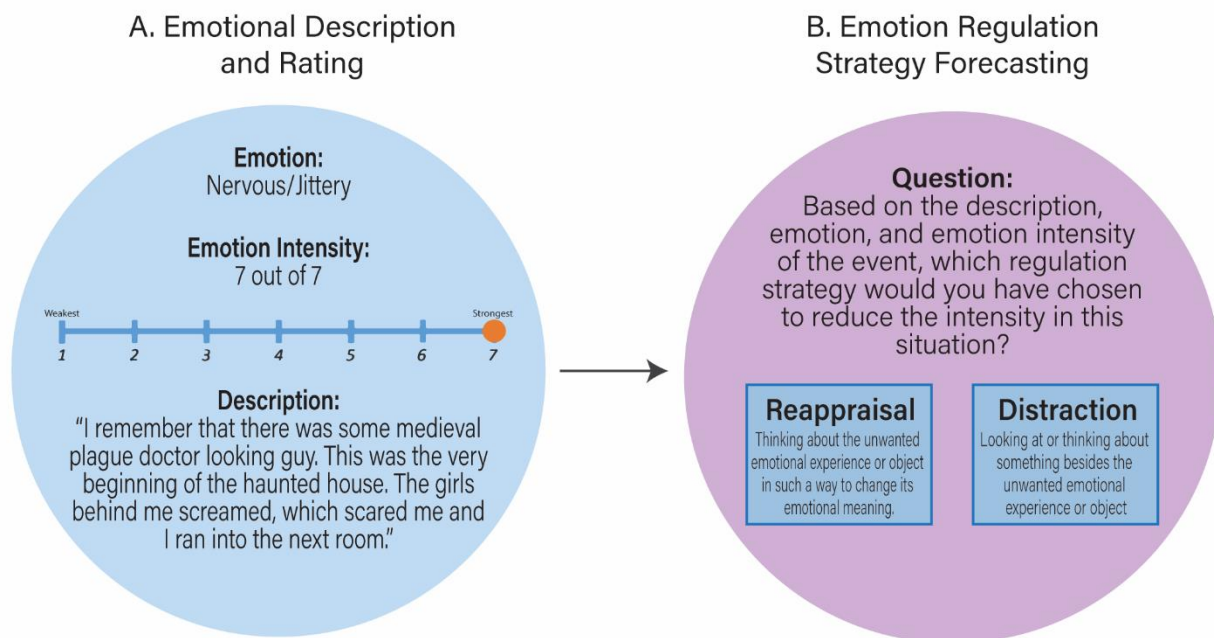


Fig 5. Study 2: Task Overview - One hundred and seventy (170) participants (strategy forecasters) read the descriptions that preliminary study participants (strategy users) wrote about their emotional experience in the haunted house. **(A)** Strategy forecasters read the strategy users' emotional descriptions and intensity rating. **(B)** Strategy forecasters indicated what regulation strategy (distraction or reappraisal) they predict they would use to regulate their emotions in the described event.

completed the study in 33.7 minutes on average (*median* = 31.5 minutes, *sd* = 14.2 minutes, *range* = 10.9 – 88.4 minutes).

Analysis. To explore whether the affective intensity that strategy users reported influenced the strategies that strategy forecasters predicted, we again specified mixed effect binary logistic regressions accounting for the random effect of participant (both the strategy forecaster and strategy user of each observation) using the “lme4” package (Bates et al., 2015) in R (R Core Team, 2022) and followed an information theoretic approach via AIC comparison. Fixed effects models were built iteratively from and compared to our null model (*ICC* = 0.14). All data and scripts used to produce this analysis are publicly available at OSF (*See Open Practices*). The design and hypotheses of Study 2 were preregistered with AsPredicted (https://aspredicted.org/XXH_W1V), though please note a discrepancy exists in the number of events used, as fewer events met our outlined criteria than initially determined.

STUDY 2 RESULTS

Study samples were similar across individual difference measures. Our first analyses aimed to determine whether relevant trait differences existed between the sample of strategy forecasters and the sample of strategy users. If such differences exist, they would limit our ability to associate differences in regulatory forecasting or usage to differences in presentation and context. Both preliminary and Study 2 participants completed the ERQ and IUS. If differences exist in ERQ scores, the groups may differ in their underlying propensity to choose specific strategies. Using a Welch’s Two Sample T-Test, we did not find significant differences between the groups in their likelihood of using reappraisal ($\bar{x}_{users} = 29.9$, $\bar{x}_{forecasters} = 31.0$, 95% *CI* = [-3.22, 1.03], $t(44.6) = -1.04$, $p = 0.305$) according to the ERQ reappraisal subscale. However, significant differences were observed between the two groups for the expressive suppression ERQ subscale ($\bar{x}_{users} = 12.5$, $\bar{x}_{forecasters} = 15.7$, 95% *CI* = [-5.03, -1.32]. $t(48.1) = -3.44$, $p = 0.001$). The relevancy

of the suppression subscale is unclear in this context, as suppression is not directly tested and neither subscale proved to be predictive of strategy usage during our preliminary study. Furthermore, a bivariate linear model found that neither the ERQ reappraisal subscale ($b = -0.001$, $se = 0.002$, $p = 0.58$) nor the suppression subscale ($b = 0.002$, $se = 0.002$, $p = 0.35$) predicted the proportion of trials in which participants selected reappraisal rather than distraction in this study as well. Additionally, if differences exist in IUS scores, the groups may differ in how they respond to ambiguity or uncertain situations; perhaps especially relevant given that “confused” was the most cited negative emotion in Study 1. However, no differences were observed in IUS scores across groups ($\bar{x}_{users} = 33.6$, $\bar{x}_{forecasters} = 34.3$, $95\% CI = [-7.07, 5.59]$, $t(34.9) = -0.238$, $p = 0.813$). Assessing differences in difficulties in applying emotion regulation strategies via DERS subscales was not possible because it was not administered to preliminary study participants.

Intensity predicts regulatory strategy forecasts. Our first model -- containing only the affective intensity of strategy users as a predictor to predict strategy forecasters’ predictions - performed better than our null model ($\chi^2(1, N = 25232) = 8.39$), demonstrating a small positive effect ($OR = 1.06$, $95\% CI = [1.02, 1.10]$, $p = 0.004$) such that more intense events were associated with an increased probability that participants predicted choosing distraction to regulate them. Additional models included age, gender, IUS score, the DERS limited access to strategies subscale, and a combination of all four as covariates. However, each model failed to outperform our initial model in chi square tests of the models (age: $\chi^2(1, N = 25232) = 0.84$, $p = 0.36$; gender: $\chi^2(2, N = 25232) = 2.81$, $p = 0.25$; IUS: $\chi^2(1, N = 25232) = 1.25$, $p = 0.26$; DERS: $\chi^2(1, N = 25232) = 0.54$, $p = 0.46$; combo: $\chi^2(5, N = 25232) = 4.52$, $p = 0.48$; respectively).

Strategy forecasters matched strategy users less than chance when selecting distraction. In signal detection theory, d' prime (d') is a measure of sensitivity that quantifies the

ability to distinguish between signal and noise in a binary decision task, and thus, can be used to determine whether congruency in strategy between strategy forecasters and strategy users is greater than chance. D' prime is calculated as the normalized value of the proportion of hits (i.e., when a strategy was used by a strategy use participant and predicted by a strategy forecasting participant) minus the proportion of false alarms (i.e., when a strategy was not used by a strategy use participant but was predicted by strategy forecasting participant). Importantly, d' is robust to unequal prior probabilities in binary outcomes, as occurs in our strategy selection (73.1% of preliminary study observations used distraction) through the incorporation of a bias parameter. Using this approach, we found that strategy forecasting participants matched the experiences of strategy using participants in their selection of reappraisal nearly at chance ($d' = -0.08$) but were below chance in matching distraction ($d' = -0.41$) (**Fig. 6**). The difference in selection congruency between these strategies was significant as determined by a paired samples t-test ($\bar{x}_{diff} = 0.328$, 95% $CI = [0.313, 0.342]$, $t(149) = 43$, $p < 0.001$), suggesting that differences in the deployment of distraction between the two groups may be driving differences in how predictive affective intensity was towards strategy forecast or usage across these two contexts. The distribution of strategy selection differed between our strategy forecasting sample and strategy use sample, as distraction was only selected in 48.8% of observations for Study 2. Taken together, we found that participants used emotional intensity to inform their ER strategy predictions but not strategy usage and that the predicted utility, and thus frequency, of using distraction in such a context might not reflect what is experienced in practice.

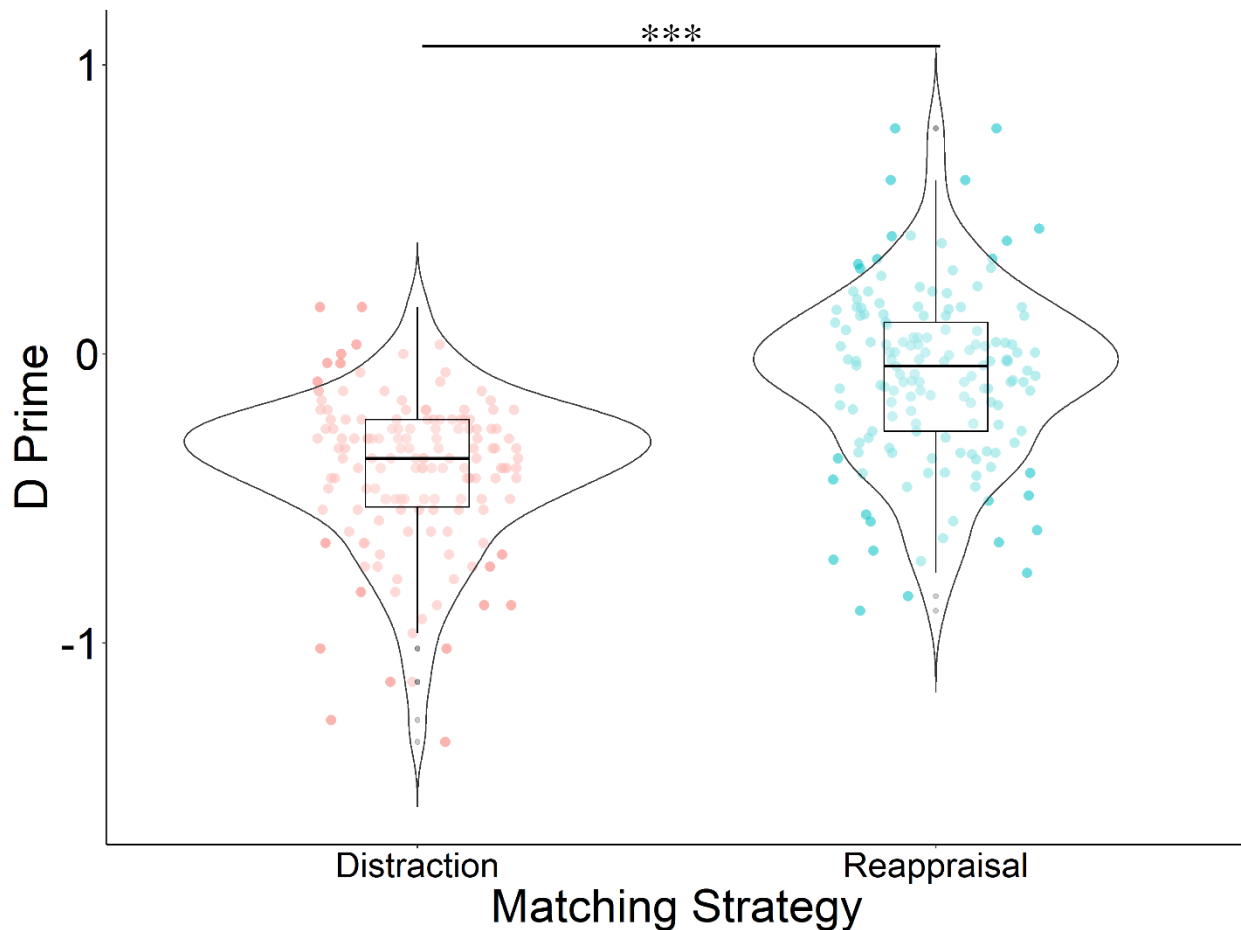


Fig 6. The strategies participants predicted they would use were assessed against the strategies that were reported as used in each situation. While neither strategy was reported greater than chance, participants were significantly more likely to predict using reappraisal on trials when distraction was actually used when compared to predicting distraction when reappraisal was actually used.

STUDY 3 METHODS

Though affective intensity failed to predict ER usage in a high-intensity, quasi-naturalistic setting, participants presented with descriptions of events that were regulated by ‘strategy users’ were more likely to forecast using distraction to regulate events described as high intensity and reappraisal for events described as being of a lower intensity. The different results observed in these studies are difficult to directly compare, though, as many features differ between the approaches. Participants in Study 1 experienced the haunted house, were not trained or prompted to self-regulate, experienced the emotions that they reported, and reported the regulatory strategies

that they used. Participants in Study 2 simulated the experiences of the haunted house, were instructed in ER strategies and prompted to make a forecast, did not experience the emotions associated with the event, and reported what strategy that they forecasted they might use in the given situation. Furthermore, Study 1 did not attempt to manipulate the emotions of participants, instead prioritizing ecological validity by observing self-report after exposure. A study which keeps instruction and stimulus exposure constant while differing the type of regulation (i.e., strategy forecasting v. usage) measured would be more conducive for direct comparisons of these two regulation processes. Study 3 attempted to rectify these discrepancies using a 2 between (condition: forecast v. experience) x 2 within (intensity: low v. high) mixed study design. Because we could not incorporate an immersive experiential component such as in Study 1 in this experimental design¹, we instead used clips from lesser-known horror films to elicit regulatory response to feature-rich, dynamic representations of negative emotion. This design allows us to directly compare the differences in regulatory behavior among untrained and unprompted participants when either reporting actual usage or forecasted selection in response to high and low intensity stimuli.

We originally hypothesized that we would replicate the results of Study 1 and Study 2: that participants placed within the “strategy user” condition would not demonstrate a relationship between stimulus intensity and regulation strategy usage due to the added complications and design decisions which reduce experimental control relative to similar designs, while participants in the “strategy forecaster” condition would demonstrate the canonical relationship between stimulus intensity and regulation strategy usage (i.e., as affective intensity increases the likelihood of relying upon distraction also increases). Counter to our predictions, we found that both

¹ The haunted house has a limited seasonal run time, and we cannot modify the intensity of the events in the haunted house as it is run by a private company.

conditions demonstrated characteristics of this canonical relationship; however, important differences in ER strategy and effectiveness were also observed across the two conditions. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

PARTICIPANTS: In October 2023, 247 participants (age: $\bar{x}_{age} = 38.57$ yrs, $range = 20 - 76$ yrs, $sd_{age} = 12.24$ yrs; gender: 94 female, 152 male, 1 non-binary) consented to an IRB-approved online study described as using clips from horror films to explore emotions and behavior. Participants completed the study on Pavlovia and were recruited/filtered via Prolific. Eligible participants were native English speakers residing in the US between the ages of 18 and 85, had normal or corrected-to-normal vision, had no history of reading-related disorders or literacy difficulties, as well as no history of mild cognitive impairment, head injury leading to unconsciousness, or unregulated mental health diagnosis. The racial identity of participants was as follows: 08.9% Asian, 06.9% Black, 05.7% Hispanic, 09.3% Mixed, 04.6% Other, and 64.6% White. Socioeconomic status was well distributed with 09.3% reporting a household income of under \$15,000 per year, 09.3% between \$15,001 and \$25,000, 10.5% between \$25,001 and \$35,000, 13.4% between \$35,001 and \$50,000, 19.0% between \$50,001 and \$75,000, 13.8% between \$75,001 and \$100,00, 08.1% between \$100,001 and \$150,000, 06.9% above \$150,000 and an additional 04.9% who did prefer not to say. In response to the question, “How much do you enjoy watching horror movies”, the average participant response was a 2.95 on a scale from 0 to 6, with 0 corresponding to “Not at all” and 6 corresponding to “Extremely” ($median = 3$, $sd = 2.16$).

Sample size was determined via *a priori* power analyses assuming an attrition rate of 10%, $r^2 \leq 0.10$ for covariates, and the most conservative effect size observed in our previous studies (*OR*

= 1.30). To achieve $1-\beta = 0.80$ ($\alpha = 0.05$, two-tailed), at least 240 participants were required, or 264 when accounting for attrition. Up to 300 slots were made available on Prolific to account for returned or incomplete research participation. Five participants were excluded for failing attention checks ($n = 1$) and familiarity with the stimuli ($n = 4$). Though we did not meet our recruitment goal of $n = 264$, our sample ($n = 242$) still surpassed the threshold calculated to achieve sufficient power. Participants were paid at a rate of \$12.00/hr.

MATERIALS AND PROCEDURE: Participants completed a web browser-based task built primarily with jsPsych v7.0 (De Leeuw et al., 2023). Following consent and a disclaimer regarding the content of the stimuli (i.e., gore, violence, mature language, harm to self or others), participants were instructed to complete the study in an isolated, distraction free space without others present to improve immersion. Participants were also instructed to set audio to a comfortable but audible level, to silence their phones, and to minimize or close other programs. To further standardize the experience and minimize distractions, the task was programmed to automatically run the browser in full screen mode and the task was programmed to stop if full screen mode was exited. Mobile devices, tablets, and screens with a resolution of less than 700 x 1250 pixels were eligible to participate.

Participants were instructed to watch and react to each of the four video clips as they naturally would. These instructions appeared as a reminder before each video played. Before each stimulus, participants were asked, “How intense are the negative emotions that you feel before starting this clip?” with responses captured on a linear sliding scale ranging in values from 0 (labelled “Not at all Intense”) to 100 (labelled “Extremely Intense”). Which side of the scale corresponded to which label was counter-balanced across participants. The default value of the slider thumb was set to 50 and participants were required to interact with the slider thumb before

progressing. As participants moved the slider thumb, the value corresponding to its position was visualized and updated accordingly. Video order was randomized. After each video, participants were asked, “How intense were the negative emotions that you felt while watching this clip?”. Responses were captured using the same scale as previously described. However, participants were also reminded of the value that they had selected prior to starting the video (e.g., a participant who indicated a 77 on the scale before watching the video saw the message: “Before this video, you reported your negative emotions were a: **77**”).

All video clips were 120 seconds in length and pulled from relatively lesser-known independent horror films, including *Vicious Fun* (Particular Crowd, Black Fawn Films), *The Marshes* (28 Productions), *Head Count* (Godmother Industries), and *Superhost* (Superchill). Forty hypothesis-blind independent raters watched and rated these clips, as well as six other films on a scale from 0 to 100 across metrics including arousal, valence, narrative coherence, and familiarity. These four clips were chosen due to their low average familiarity values ($range \bar{x} = 0.25 - 6.05$) - indicating that few participants would have likely seen them before -- and negative median valence scores (i.e., values below 50; $\bar{x}_{Vicious\ Fun} = 15$, $\bar{x}_{The\ Marshes} = 01.5$, $\bar{x}_{Head\ Count} = 33.5$, $\bar{x}_{Superhost} = 25.5$). They were also chosen for having either relatively high or relatively low median arousal scores ($\bar{x}_{Vicious\ Fun} = 77.5$, $\bar{x}_{The\ Marshes} = 70$, $\bar{x}_{Head\ Count} = 50.7$, $\bar{x}_{Superhost} = 58.5$) and minimal variance around those values. This allowed us to attempt to manipulate the affective experiences of participants to influence self-regulation behaviors.

After viewing and rating all videos, Study 3 participants were assigned to either a “strategy use” ($n = 130$) or “strategy forecast” ($n = 112$) condition. Both conditions were given descriptions and examples of distraction and reappraisal that mirrored those used in Study 2 (though, ‘haunted house’ was replaced with ‘horror movie’). Whether participants received the

description and examples of distraction or reappraisal first was counterbalanced across participants. However, following these descriptions participants were asked either, “Which strategy, if any, did you use to regulate your negative emotions while watching this clip?” (strategy use condition) or “Which strategy, if any, would you predict the average person would use to regulate their negative emotions while watching this clip?” (strategy forecast condition). Participants responded to this question by selecting either “Reappraisal”, “Distraction”, or “Neither”. The order in which these options appeared were randomized for each participant. Reminders of what the three options referred to appeared below these options. Reminders for participants in the strategy use condition read, “CHOOSE REAPPRAISAL if you changed how you thought about the video; CHOOSE DISTRACTION if you diverted your attention from the video; CHOOSE NEITHER if you did not implement either of the other two strategies”. Reminders for participants in the strategy forecast condition read, “CHOOSE REAPPRAISAL if they should change how they thought about the video; CHOOSE DISTRACTION if they should divert their attention from the video; CHOOSE NEITHER if they should not implement either of the other two strategies”. Note that a potential discrepancy exists between the phrasing of our strategy forecast instructions (“Which strategy ... *would* you predict...”) and reminders (“...if they *should* ...”) which we discuss further in the limitations. If participants selected neither, they were not asked any further questions for that stimulus. If participants selected a non-neither option, they were asked how much the strategy they used reduced or the strategy they forecasted would reduce negative emotions while watching the stimulus on a 0 to 100 scale as previously described. Participants also either answered how effortful it was to use a strategy and how successfully they could use it, or how effortful that they predict it might be and how successfully they predict it could be used. These were each captured on 7-point Likert scales. Labels included “Not at all”,

“Slightly”, “Somewhat”, “Moderately”, “Considerably”, “Very Much”, and “Extremely”. Participants responded to questions for each clip in the order that they clips were watched and a still from each clip was centrally-placed on the screen for each question to ensure participants understood which stimulus that they question was referring to. Following all questions, participants were asked to provide a brief example of both reappraisal and distraction in a free-response textbox to further assess task comprehension. Following the primary task, participants completed individual difference measures, including the Emotion Regulation Questionnaire, the Difficulties with Emotion Regulation Survey, and the Intolerance of Uncertainty Scale. Participants completed the study in 21.9 minutes on average (*median* = 20.5 minutes, *sd* = 56.0 minutes, *range* = 13.6 – 50.7 minutes). For additional information, see our OSF repository containing all task materials, including instructions and texts (*See Open Practices*).

Analysis. To explore whether affective intensity and condition influenced the strategies participants used or predicted, we again specified mixed effect binary logistic regressions accounting for the random effect of participant and stimulus using the “lme4” package (Bates et al., 2015) in R (R Core Team, 2022) and followed an information theoretic approach via AIC comparison. Fixed effects models were iteratively built from and compared to our null model ($ICC = 0.162$), which did not contain any fixed effects. All data and scripts used to produce this analysis are publicly available at OSF (*See Open Practices*). The design and hypotheses of Study 3 were preregistered with AsPredicted (<https://aspredicted.org/n3ne3.pdf>).

STUDY 3 RESULTS

Video stimuli elicited predictable emotional responses. To ensure that our emotion manipulation was successful, negative affective intensity ratings following each low- and high-intensity stimuli were compared using Welch’s Two Sample t-test, which found that our manipulation was successful ($\bar{x}_{high} = 53.5$, $\bar{x}_{low} = 40.6$, 95% $CI = [9.1, 16.8]$, $t(948.1) = 6.61$, $p <$

0.001). The average baseline intensity (i.e., intensity assessed prior to each video) was 27.4 pts (*median* = 19, *sd* = 27.0) while the average post-exposure intensity was 47.1 pts (*median* = 50, *sd* = 30.9).

Strategy forecasters and users reported no differences in emotional intensity. Though participants were randomly assigned to conditions after having provided affective ratings, we wanted to ensure that any effects discovered between conditions could not be attributable to differences in experienced affective intensity. A Welch's Two Sample t-test comparing affective intensity ratings between strategy users and forecasters both before ($\bar{x}_{user} = 26.2, \bar{x}_{forecaster} = 28.8$, 95% *CI* = [-6.1, 0.8], $t(908.3) = -1.50, p = 0.134$) and after ($\bar{x}_{user} = 46.5, \bar{x}_{forecaster} = 48.0$, 95% *CI* = [-5.4, 2.4], $t(917.9) = -0.73, p = 0.460$) stimulus exposure found no statistically significant differences in affective ratings. Strategy forecasters and users also did not differ across any relevant individual difference measures or demographics, including the reappraisal ($t(198.6) = 0.14, p = 0.890$) and suppression ($t(222.57) = -1.94, p = 0.054$) ERQ subscales, both Factor 1 ($t(235.0) = -0.94, p = 0.349$) and Factor 2 ($t(220.4) = -0.96, p = 0.336$) of the IUS scale, the limited access to strategies subscale of the DERS ($t(221.4) = 0.08, p = 0.938$), age ($t(220.0) = -0.14, p = 0.888$), or horror enjoyment ($t(225.9) = 0.12, p = 0.908$).

Intense emotional responses were more likely to be regulated. To add greater ecological validity to a forced choice paradigm, we provided participants the option to indicate whether a stimulus should be regulated (i.e., reappraisal or distraction) or not regulated (i.e., neither). In line with hedonic motivations to self-regulate, we expect that videos of greater self-reported negative affective intensity should be more likely to be regulated; if that is not the case, then our design may generate noisy or counterintuitive results. To that end, we did find that videos in which participants selected neither distraction nor reappraisal were of a lower intensity ($\bar{x}_{regulated} = 53.3$,

$\bar{x}_{unregulated} = 36.7$, 95% $CI = [12.7, 20.]$, $t(735.6) = 8.30$, $p < 0.001$) and a mixed effect binary logistic regression found that each standard deviation unit increase in affective intensity results in a 2.86 fold increase in the odds of self-regulation occurring (95% $CI = [2.16, 3.78]$, $p < 0.001$).

Strategy forecasters anticipated regulating more often than strategy users; Strategy users reappraised more often than strategy forecasters. Overall, reappraisal was used or forecasted in 33.1% of trials, distraction was used or forecasted in 28.5% of trials, and 36.7% of trials were left unregulated or without either option being forecasted. To assess whether there were differences in the distribution of strategy usage or forecast by condition, which could have implications for our primary analysis, we conducted chi-square test on strategy use/forecast and condition, which found differences in the frequencies with which strategy forecasters and users left trials unregulated ($\chi^2(1, N = 953) = 120.35$, $p < 0.001$) and used or forecasted a regulation strategy ($\chi^2(1, N = 613) = 6.14$, $p = 0.013$). More precisely, strategy users reported regulating (52.6%) and not regulating (47.4%) about equally while forecasters heavily favored regulating

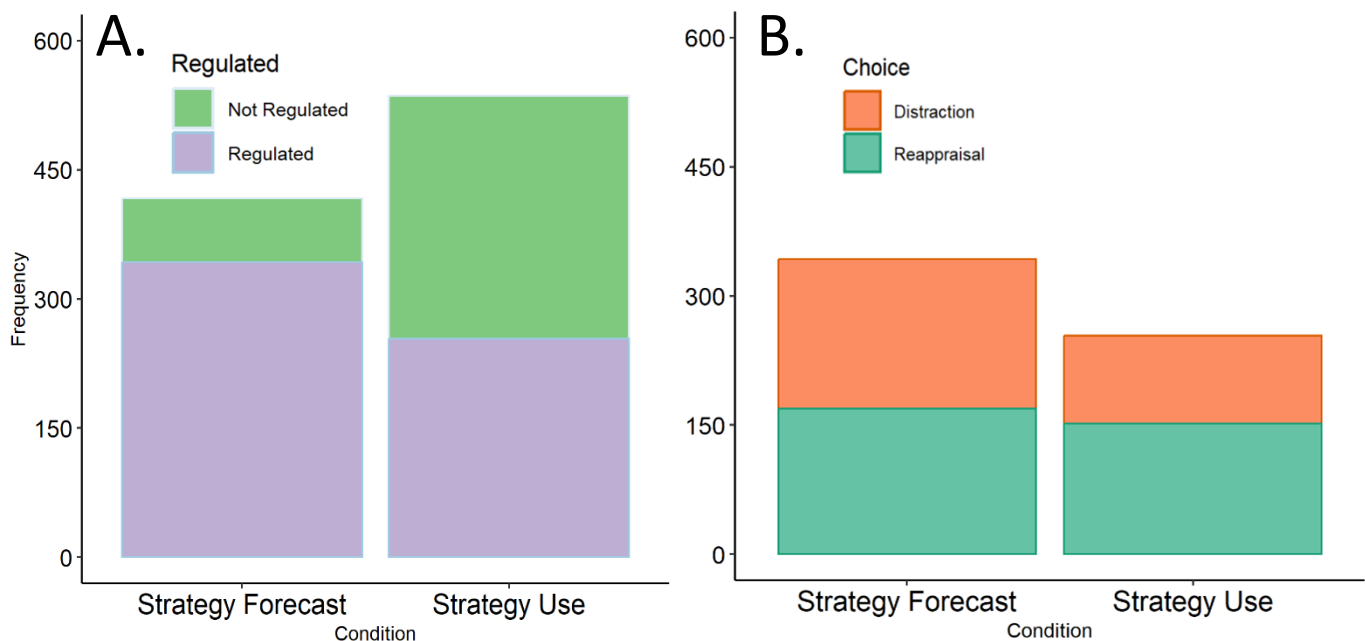


Fig 7. Chi-square tests highlighted that: **(A)** strategy users were more likely to not regulate a stimulus than strategy forecasters, but that **(B)** strategy forecasters more evenly used both strategies when predicting strategy selection.

(82.2%) over not regulating (17.7%) across trials (**Fig. 7a**). Additionally, strategy users reported using reappraisal (59.8%) more often than distraction (40.2%) while strategy forecasters predicted choosing both distraction (49.2%) and reappraisal (50.7%) about evenly (**Fig. 7b**).

Strategy forecasters and users differ in strategy selection probability at low, not high, intensities. Following the analysis plan outlined within the Study 3 methods sections, we compared models via chi-square tests that included fixed effects for post-exposure affective intensity, baseline affective intensity, condition, and an interaction between post-exposure affective intensity and condition. These models were built iteratively, adding one effect at a time. Our best fitting model included all four terms ($\chi^2(1, N = 597) = 4.205, p = 0.040$) and found the interaction was predictive of strategy usage/forecasting ($OR = 0.56, 95\% CI = [0.36, 0.87], p = 0.010$). Having tested the model that we hypothesized, we identified a selection of possible covariates (age, horror, enjoyment, ERQ subscales, IUS subscales, and the DERS limited access to strategies subscale) and iteratively added them to the model in order of most correlated with the outcome variable to pursue the model of best fit. Only the addition of the reappraisal ERQ subscale improved model fit above that of our hypothesized model ($\chi^2(1, N = 597) = 9.3587, p = 0.002$). This model yielded a significant interaction term ($OR = 0.61, 95\% CI = [0.39, 0.95], p = 0.029$) and suggests that condition is moderating the relationship between affective intensity and strategy usage/forecasting. More specifically, at high intensities both strategy forecasters and users are similar in their likelihood of predicting or using distraction, respectively. However, in congruence with forecasters predicting to choose reappraisal less than strategy users used it in practice, strategy forecasters are less likely to predict choosing reappraisal than strategy users were to use reappraisal at lower intensities (**Fig. 8**).

Strategy forecasters and users differ in regulation effort. Participants in both conditions failed to generate significant differences in how much effort they exerted or predicted that they would exert to use reappraisal and distraction ($F(1, 215) = 0.392, p = 0.532$). However, the participants in the use condition did differ from participants in the forecasting condition in how effortful regulation was across strategies, even when adjusting for strategy usage/forecast ($F(1,215) = 3.820, p = 0.045$). Bonferroni-adjusted post-hoc contrasts determined that this difference is primarily driven by reappraisal; experiencers reported that reappraisal was more

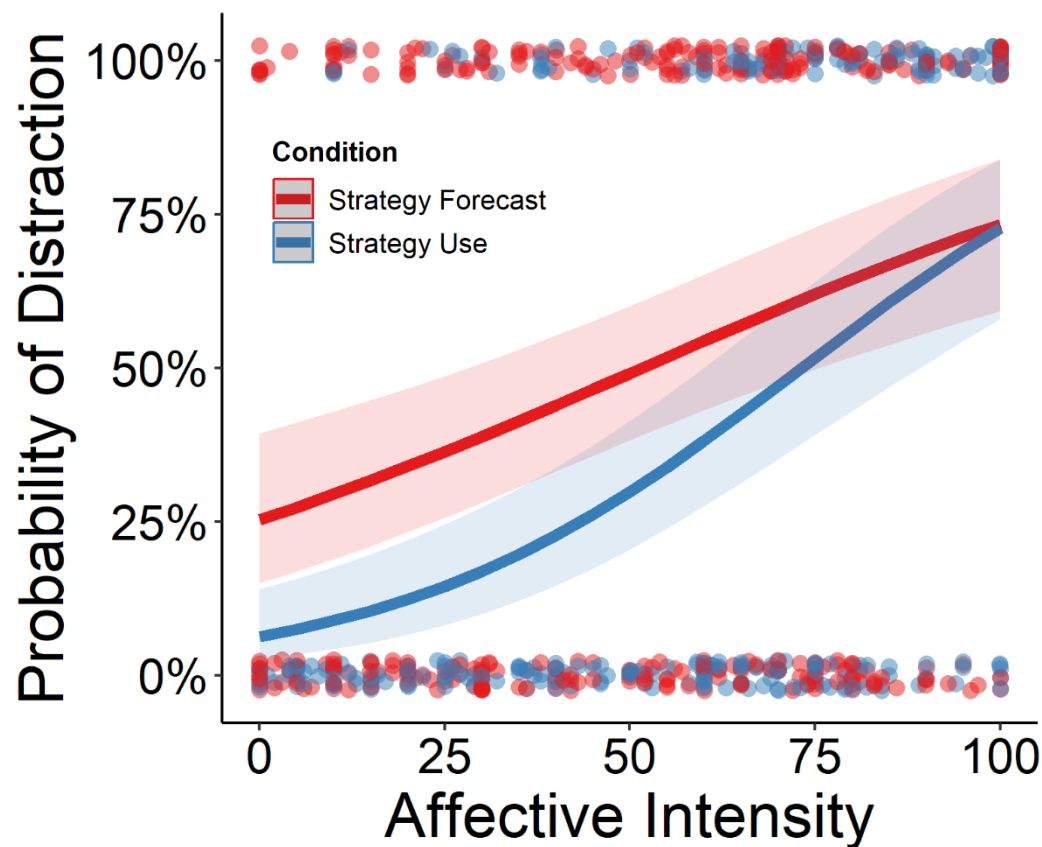


Fig 8. Mixed effects binary logistic regression found that intensity predicts strategy usage and prediction for both strategy users and forecasters, respectively. However, at lower intensities, strategy forecasters are less likely to predict selecting reappraisal than strategy users were to use reappraisal.

effortful than forecasted predicted it would be ($\bar{x}_{user} = 3.11, \bar{x}_{forecaster} = 2.60, 95\% CI = [0.13, 0.88], t(296.5) = .66, p = 0.016$).

Strategy forecasters and users differ in regulation intensity reduction. When examining how effective participants thought the strategies would be at reducing affective intensity, adjusting for the random effect of participant, we found a significant main effect of strategy ($\bar{x}_{reappraisal} = 61.9, \bar{x}_{distraction} = 54.5, 95\% CI = [3.31, 11.48], t(582.12) = 3.55, p < 0.001$) and condition ($\bar{x}_{users} = 55.0, \bar{x}_{forecasters} = 61.1, 95\% CI = [-10.32, -1.85], t(494.6) = -2.83, p = 0.004$), but more intriguingly, in an interaction model we found a significant interaction between the two variables ($F(1,377) = 4.31, p = 0.038$). A subsequent Bonferroni-adjusted contrast revealed a significant difference between strategy users and forecasters in how effectively distraction reduces affective intensity ($\bar{x}_{users} = 48.0, \bar{x}_{forecasters} = 58.3, 95\% CI = [-16.62, -4.13], t(198.2) = -3.28, p = 0.001$) (**Fig 9**). This finding is congruent with findings from both Study 1 and Study 2 in which individuals seem to overestimate or mispredict the effectiveness of distraction, but not reappraisal, within these settings.

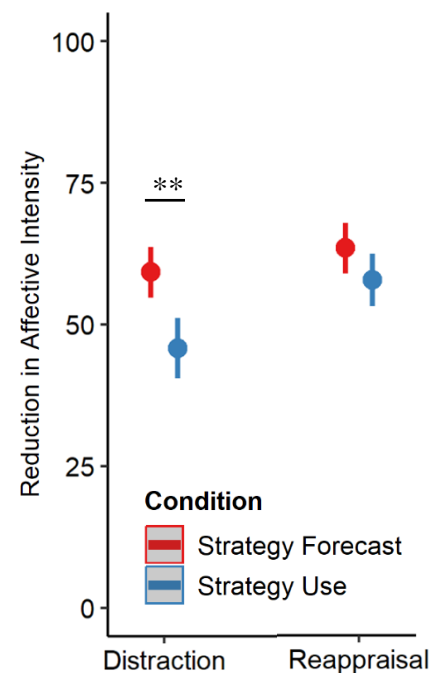


Fig 9. Strategy forecasters predicted that distraction would reduce negative affective intensity by a greater magnitude than strategy users reported that it actually had.

GENERAL DISCUSSION

Three studies examined the association between emotional intensity and regulation strategy usage in response to dynamic, multimodal stimuli of varying intensity. Study 1 tasked untrained participants with recalling emotional and spontaneous regulatory behaviors in a surprise recall task after exposure to a high-intensity multi-modal context (a haunted house). Affective intensity did not predict strategy usage. Though distraction was used more often than reappraisal, distraction was less successful at regulating in response to high affective intensities in this context. Study 2 found an association between affective intensity and strategy usage among participants forecasting regulatory behavior in response to descriptions of regulated events that haunted house participants actually experienced. However, strategy forecasters overpredicted how often distraction was used by participants who regulated within the haunted house (i.e., strategy users). Study 3 measured differences between strategy users and forecasters regarding ER within the same study and found significant differences between strategy forecasters and users in how often distraction was forecasted or used and how effective the groups predicted or reported distraction to be, respectively. The present findings highlight challenges in translating ER theory to real-world application, as uncontrolled, complex, and dynamic situations in everyday life may strongly influence the efficacy and frequency with which regulatory strategies are used.

These results add nuance to our understanding of affective intensity's influence upon regulation strategy application and complements research on environmental affordances, or the extent to which features of a situation lend themselves to either distraction or reappraisal (Suri et al., 2018; Young & Suri, 2020). It is possible that individuals initially regulated their emotions with the strategy they believed had the greatest value (i.e., the most effective strategy with the lowest effort investment; typically distraction) based upon models of regulation they had previously generated from experiences with similar media, as outlined by computational decision-

making frameworks of ER (Etkin et al., 2015). However, the unrelenting, attention-grabbing nature of challenging high-intensity situations may grant few affordances by which to distract oneself. On the other hand, reappraisals may appear to be less valuable initially, resulting in lower usage, but could be more effective when used due to a relatively greater volume of environmental affordances (i.e., things to repurpose or reconstrue (Uusberg et al., 2019); e.g., actors, props, goals, etc.). This explanation mirrors the strategy-selection relationship hypothesized by Etkin and colleagues (2015) who posited that a more explicit, higher-cost, model-based approach to ER may be applied more effectively when implicit, or model-free, ER tendencies were not arriving at their desired goal via prediction error adjustment alone. Distilling events down to a text-based representation or video, as had been done in Studies 2 and 3, may provide greater distraction affordances and rebalance the likelihood of individuals choosing or predicting reappraisal or distraction in situations when it would canonically make sense to do so.

How a person passively assesses the emotionally-relevant features of an environment may guide regulatory behaviors, but so too might the action affordances – opportunities to move, modify, or engage with the environment – that a situation presents. Action-oriented perspectives on ER emphasize that emotion regulation behaviors are a product of active processes, such as forward modelling (Bramson et al., 2023). Common associations between self-regulation, emotion, and activity in the sensorimotor and pre-motor systems illustrate how entangled action is with regulation (Bramson et al., 2018; Mobbs et al., 2007; Saarimäki et al., 2016). The degree to which one is free to physically navigate their space or interact with evocative stimuli may have important implications on situational appraisals, thus informing which strategies participants predict may or may not work (Ridderinkhof, 2017). **Figure 5** highlights one example of this within the haunted house, as a participant was able to mitigate a negative emotional reaction by physically

circumventing the stimulus. Stimulus-response paradigms which situate participants in stationary positions or lack contexts enveloping the emotionally-evocative stimulus may not be well suited for modeling the effect that these factors have upon typical or daily ER strategy usage.

Hot-cold empathy gap research, which measures forecasting differences between how people think they will feel hypothetically and how they feel in practice (Loewenstein, 1996), are relevant as well. Individuals in “cold states” consistently underpredict the challenges associated with meeting affectively-relevant goals during “hot states” (Sayette et al., 2008; Van Boven & Loewenstein, 2003). Such a pattern mirrors the differences observed between Studies 1 and 2, wherein decontextualizing events (i.e., shifting from a hot state to a cold state) yielded a predictable pattern in strategy choice not observed during hot state ER usage. They also mirror differences in strategy usage and success. Such patterns highlight that emotion self-regulation is a complex, multi-faceted construct and different proportions of its variability may be better captured by different approaches (Friedman & Gustavson, 2022). Though this study is the first to our knowledge that has utilized a high-intensity, dynamic, feature-rich paradigm to demonstrate this in the domain of ER, similar approaches have demonstrated similar discrepancies in the moral domain (FeldmanHall et al., 2012). Computerized lab tasks have been theorized to assess regulatory performance in optimal conditions (Wennerhold & Friesse, 2020) and this may contrast study designs such as Study 1, which may be closer to performance in typical conditions, and Studies 2 and 3, which may be somewhere in-between (Friedman & Gustavson, 2022). Though participants may be capable of performing self-regulation at high levels in optimal conditions, they may not feel motivated to do so in typical conditions (Grund & Carstens, 2019).

Limitations & Future Directions. There are several limitations in our experimental approach that demand attention. First, our Study 1 aims were relatively narrow in comparison to

the vast ER behavior variability captured by this dataset and resulted in excluding many observations that did not meet our inclusion criteria (*See Fig. 2*). More analyses would be required to fully explore this space. Additionally, though many features of our design mirror regulation of emotions in everyday life, haunted houses may have limited generalizability to other high-intensity settings that we commonly experience (Clasen et al., 2019; Tashjian et al., 2022) and the purpose haunted houses commonly serve (i.e., entertainment) could also circumscribe their generalizability to graver situations. The use of a haunted house as our setting also may have resulted in self-selection biases within our sample. It must be noted that we did not directly manipulate emotional intensity within this design and most of the self-report data that we did collect was captured post-exposure, not during exposure. Thus, without manipulation, it is unclear whether self-reported emotion intensity in Study 1 represents a precursor to regulation, a product of regulation, or some combination of the two. Although computing standardized emotion scores of each event within the haunted house would be ideal and theoretically possible, the limited window in which we had to collect our experimental data made it difficult to assemble an additional independent sample for this purpose. While we chose language which aimed to target emotion as a precursor, it is entirely possible that subjects may have misremembered or misinterpreted this intent. While some of our exploratory analyses suggest that the relationship between intensity and self-regulation failed to materialize even with stimulus manipulation (i.e., using section as a predictor of regulation) and an aggregate of negative emotion intensity (i.e., using a leave-one-out approach as a predictor of regulation), these analyses were limited due to low-resolution and the possibility of low-power, respectively. Because of this design, the results of Study 1 analyses are not conclusive. Additionally, we unfortunately did not have measures of startle sensitivity or fear-enjoyment from Study 2 participants. Thus, we cannot compare our Study 1 and Study 2 samples along these

metrics, which explain some of the differences in regulatory behaviors between samples. Studies 2 and 3 were also conducted entirely online. Though means of standardizing the experience were attempted, we have less control than possible in an in-person context and cannot guarantee that participants were fully focused on study tasks. Additionally, the choice to homogenize stimuli for strategy forecasters and users in Study 3 may have also unintentionally homogenized the regulation processes that we intended to target. In other words, forecasters may have just been tempted to report the strategies that they used and users may have been tempted to report what they *would* have used had they regulated. However, the inclusion of a “no regulation” option was intended to reduce the latter possibility.

Participants in Study 3 were initially instructed to predict what strategy they believed the average person *would* choose while the strategy reminders phrased strategy selection as what average persons *should* choose. Both phrasings do imply some form of forecasting, but the former may be more likely to elicit *descriptive* forecasts and the latter more likely to elicit *prescriptive* forecasts, and these may lead to different regulatory conclusions. For example, a person might know that they *should* reduce stress by seeking therapy but might instead engage in maladaptive methods of stress-reduction, such as substance abuse. Choice normativity (i.e., variance around which action one *should* take) is well-studied in health, moral, and social psychology but has been comparatively unexplored in emotion regulation. Extant ER literature highlights variance around how people regulate, but not necessarily around how they think they should regulate or how they think they would regulate. Regardless, this inconsistency may have introduced confusion into participant forecasting and the extent to which this may have had impacts upon the outcome of Study 3 is unknown.

A small but important contingent of the existing ER literature has highlighted intra-strategy heterogeneity in the regulation techniques that we had examined (Uusberg et al., 2019; Webb et al., 2012). For example, Webb and colleagues identified at least three distinct ER approaches that could be categorized as reappraisal (i.e., reappraising the stimulus, reappraising the emotional response, reappraisal via perspective-taking) and these reappraisal subtypes demonstrate varying effectiveness in modulating emotion depending upon situational and personal factors. It may be the case that specific types of reappraisal and distraction demonstrate an intensity-selection relationship that better mirrors the canonical association observed in lab studies, but testing such a question with sufficient power would necessitate far more observations than are available in this dataset.

Due to resource constraints, participants also necessarily experienced the haunted house in groups. Although participants were instructed to not discuss their experiences and post-hoc analyses failed to find any association between group membership and strategy usage, the group context in which the experience occurred could have influenced behavior and cognitive perceptions in unknown ways. The presence and strength of friendship among group members was also assessed and was not predictive of regulation. Lastly, some events, particularly in the preliminary study and multiverse analyses, had to be captured at a one-week delay due to collaboration with a memory study, which calls into question the reliability of some responses. Future research should limit the delay between experience and report as much as possible without interfering with emotional events. Future research might also take interest in the order of strategies reported in regulation-challenging situations; that is to say, does the likelihood of using a reappraisal strategy for a given person in this context increase with each failed attempt to engage distraction? Although we did not have sufficient observations to examine the data with such

granularity, examining associations between the intensity of specific emotions and regulation usage might also be of particular interest to future researchers in this space (Young & Suri, 2020). Despite its limitations, this dataset and approach may be of interest to those exploring spontaneous regulation tendencies from untrained participants in response to both positive and negative events.

Conclusion. Taken together, the present studies represent what we believe to be the first attempt to extend the association observed between affective intensity and different ER behaviors to dynamic, multimodal and high-intensity settings using untrained participants. This approach offers an alternative means of exploring ER usage while pursuing greater ecologically-valid in study design. Our data and results may be of particular interest to other emotion, self-regulation, and cognitive control researchers interested in quasi-naturalistic design. In failing to replicate lab results with Study 1 but finding a modest association in Studies 2 and 3, we may offer greater support for the importance of context in determining the fit of ER strategies, though the limitations inherent to this study leave room for other possibilities. Ultimately, though, this study highlights the importance of extending what we know about regulation in relatively mundane, controlled situations to those crowded, loud, and perhaps uncontrollable contexts in which emotion regulation success could be of dire consequence.

Acknowledgments:

We sincerely thank Ian O'Shea, Isabel Leiva, Angelique Vittone, Lauren Iglie, Kathryn Lockwood, Devlin Eckardt, Adhya Gowda, Troy Houser, Marissa Ballew, Kaitlin Dow, and Mary Nolwenn Achy for collecting and/or coding data.

Declaration Of Conflicting Interests:

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding:

This work was funded in part by the Brain and Behavioral Research Foundation's NARSAD Young Investigator Award received by Vishnu P. Murty and by the National Science Foundation (Grant No. 2123474).

Author's Note:

Earlier versions of this manuscript have been available on PsyArxiv (<https://osf.io/preprints/psyarxiv/23wtz>) since November 28th, 2022. Posters containing some of this data and analyses have been presented at the 2021, 2022, and 2024 Society for Personality & Social Psychology Conference, as well as the 2022 Society for Affective Science Conference. A symposium containing these findings was presented during the 2024 annual conference of the Association of Psychological Science. Invited presentations using some of these data and analyses occurred in 2020 at the Temple University Motivational Behavior Seminar Series and in 2023 to the March Lab at Florida State University and Satpute Lab at Northeastern University. Stasiak et al., 2023 was published using physiological and emotion data from the preliminary study. Cliver et al., 2024 was accepted for publication using memory data from the preliminary study and Study 1.

Transparency and Openness Practices:

The preregistration for Studies 1, 2, and 3 can be found at https://aspredicted.org/DP1_453, https://aspredicted.org/XXH_W1V, and <https://aspredicted.org/n3ne3.pdf>, respectively.

Deidentified data, analysis and data cleaning scripts, questionnaires, and data collection files to replicate the findings have been made publicly available via OSF and can be accessed at https://osf.io/j5sku/?view_only=89d87669e7674096819c439ca109c483. All products generated from this data, including powerpoints and posters highlighted under the Author's Note have been made available within this repository as well.

CrediT:

William J. Mitchell: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing,

Visualization, Project administration. **Joanne Stasiak:** Conceptualization, Methodology, Investigation, Resources, Data Curation, Writing – Review & Editing, Project administration.

Steven Martinez: Conceptualization, Investigation, Data Curation, Project administration.

Katelyn Cliver: Conceptualization, Investigation, Data Curation, Project administration. **David**

Gregory: Conceptualization, Investigation, Data Curation, Project administration. **Samantha**

Reisman: Conceptualization, Investigation, Data Curation, Project administration. **Helen**

Schmidt: Software, Validation, Formal analysis, Writing – Review & Editing. **Vishnu P.**

Murty: Conceptualization, Resources, Writing – Review & Editing, Supervision, Project administration, Funding acquisition. **Chelsea Helion:** Conceptualization, Methodology,

Resources, Writing – Review & Editing, Visualization, Supervision, Project administration.

References:

- Aldao, A. (2013). The Future of Emotion Regulation Research: Capturing Context. *Perspectives on Psychological Science*, 8(2), 155–172. <https://doi.org/10.1177/1745691612459518>
- Aldao, A., & Nolen-Hoeksema, S. (2013). One versus many: Capturing the use of multiple emotion regulation strategies in response to an emotion-eliciting stimulus. *Cognition & Emotion*, 27(4), 753–760. <https://doi.org/10.1080/02699931.2012.739998>
- Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, 30(2), 217–237. <https://doi.org/10.1016/j.cpr.2009.11.004>
- Argyriou, E., & Lee, T. T. C. (2020). The role of distress and fear transdiagnostic dimensions in emotion regulation choice. *Journal of Affective Disorders*, 276, 433–440. <https://doi.org/10.1016/j.jad.2020.07.060>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Blanchard-Fields, F., Stein, R., & Watson, T. L. (2004). Age Differences in Emotion-Regulation Strategies in Handling Everyday Problems. *The Journals of Gerontology: Series B*, 59(6), P261–P269. <https://doi.org/10.1093/geronb/59.6.P261>
- Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers*, 35(4), 634–639. <https://doi.org/10.3758/BF03195543>

- Bradley, M. M., & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In *Handbook of emotion elicitation and assessment*. (pp. 29–46). Oxford University Press.
- Bramson, B., Jensen, O., Toni, I., & Roelofs, K. (2018). Cortical Oscillatory Mechanisms Supporting the Control of Human Social–Emotional Actions. *The Journal of Neuroscience*, 38(25), 5739–5749. <https://doi.org/10.1523/JNEUROSCI.3382-17.2018>
- Bramson, B., Toni, I., & Roelofs, K. (2023). Emotion regulation from an action-control perspective. *Neuroscience & Biobehavioral Reviews*, 153, 105397. <https://doi.org/10.1016/j.neubiorev.2023.105397>
- Carver, C. S., & Scheier, M. F. (1981). The self-attention-induced feedback loop and social facilitation. *Journal of Experimental Social Psychology*, 17(6), 545–568. [https://doi.org/10.1016/0022-1031\(81\)90039-1](https://doi.org/10.1016/0022-1031(81)90039-1)
- Cendri A. Hutcherson, Hutcherson, C. A., Philip R Goldin, Goldin, P. R., Kevin N. Ochsner, Ochsner, K. N., John D. E. Gabrieli, Gabrieli, J. D. E., Lisa Feldman Barrett, Barrett, L. F., James J. Gross, & Gross, J. J. (2005). Attention and emotion: Does rating emotion alter neural responses to amusing and sad films? *NeuroImage*, 27(3), 656–668. <https://doi.org/10.1016/j.neuroimage.2005.04.028>
- Christou-Champi, S., Farrow, T. F. D., & Webb, T. L. (2015). Automatic control of negative emotions: Evidence that structured practice increases the efficiency of emotion regulation. *Cognition and Emotion*, 29(2), 319–331. <https://doi.org/10.1080/02699931.2014.901213>
- Clasen, M., Andersen, M., & Schjoedt, U. (2019). Adrenaline junkies and white-knucklers_ A quantitative study of fear management in haunted house visitors. *Poetics*, 73, 61–71. <https://doi.org/10.1016/j.poetic.2019.01.002>

- Cliver, K. G., Gregory, D. F., Martinez, S. A., Mitchell, W. J., Stasiak, J., Reisman, S., Helion, C., & Murty, V. P. (2024). Temporal memory for threatening events encoded in a haunted house. *Cognition & Emotion*, 1–17. <https://doi.org/10.1080/02699931.2024.2338962>
- Colombo, D., Fernández-Álvarez, J., Suso-Ribera, C., Cipresso, P., Valev, H., Leufkens, T., Sas, C., Garcia-Palacios, A., Riva, G., & Botella, C. (2020). The need for change: Understanding emotion regulation antecedents and consequences using ecological momentary assessment. *Emotion*, 20(1), 30–36. <https://doi.org/10.1037/emo0000671>
- De Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an Open-Source CollaborativeEcosystem of Behavioral Experiments. *Journal of Open Source Software*, 8(85), 5351. <https://doi.org/10.21105/joss.05351>
- Denny, B. T., Ochsner, K. N., Weber, J., & Wager, T. D. (2014). Anticipatory brain activity predicts the success or failure of subsequent emotion regulation. *Social Cognitive and Affective Neuroscience*, 9(4), 403–411. <https://doi.org/10.1093/scan/nss148>
- Diamond, N. B., Armson, M. J., & Levine, B. (2020). The Truth Is Out There: Accuracy in Recall of Verifiable Real-World Events. *Psychological Science*, 31(12), 1544–1556. <https://doi.org/10.1177/0956797620954812>
- Dixon-Gordon, K. L., Aldao, A., & De Los Reyes, A. (2015). Emotion regulation in context: Examining the spontaneous use of strategies across emotional intensity and type of emotion. *Personality and Individual Differences*, 86, 271–276. <https://doi.org/10.1016/j.paid.2015.06.011>
- Dorman Ilan, S., Tamuz, N., & Sheppes, G. (2019). The fit between emotion regulation choice and individual resources is associated with adaptive functioning among young children. *Cognition and Emotion*, 33(3), 597–605. <https://doi.org/10.1080/02699931.2018.1470494>

- Draheim, C., Pak, R., Draheim, A. A., & Engle, R. W. (2022). The role of attention control in complex real-world tasks. *Psychonomic Bulletin & Review*, 29(4), 1143–1197. <https://doi.org/10.3758/s13423-021-02052-2>
- English, T., Lee, I. A., John, O. P., & Gross, J. J. (2017). Emotion regulation strategy selection in daily life: The role of social context and goals. *Motivation and Emotion*, 41(2), 230–242. <https://doi.org/10.1007/s11031-016-9597-z>
- Etkin, A., Büchel, C., & Gross, J. J. (2015). The neural bases of emotion regulation. *Nature Reviews Neuroscience*, 16(11), 693–700. <https://doi.org/10.1038/nrn4044>
- FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition*, 123(3), 434–441. <https://doi.org/10.1016/j.cognition.2012.02.001>
- Ford, B. Q., Gross, J. J., & Gruber, J. (2019). Broadening Our Field of View: The Role of Emotion Polyregulation. *Emotion Review*, 11(3), 197–208. <https://doi.org/10.1177/1754073919850314>
- Ford, B. Q., & Troy, A. S. (2019). Reappraisal Reconsidered: A Closer Look at the Costs of an Acclaimed Emotion-Regulation Strategy. *Current Directions in Psychological Science*, 28(2), 195–203. <https://doi.org/10.1177/0963721419827526>
- Friedman, N. P., & Gustavson, D. E. (2022). Do Rating and Task Measures of Control Abilities Assess the Same Thing? *Current Directions in Psychological Science*, 31(3), 262–271. <https://doi.org/10.1177/09637214221091824>
- Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, 12(2), 314–325. <https://doi.org/10.1037/a0026007>

- Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>
- Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, 74(1), 224–237. <https://doi.org/10.1037/0022-3514.74.1.224>
- Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, 39(3), 281–291. <https://doi.org/10.1017/S0048577201393198>
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85(2), 348–362. <https://doi.org/10.1037/0022-3514.85.2.348>
- Grund, A., & Carstens, C.-A. (2019). Self-control motivationally reconsidered: “Acting” self-controlled is different to “being good” at self-control. *Motivation and Emotion*, 43(1), 63–81. <https://doi.org/10.1007/s11031-018-9721-3>
- Haines, S. J., Gleeson, J., Kuppens, P., Hollenstein, T., Ciarrochi, J., Labuschagne, I., Grace, C., & Koval, P. (2016). The wisdom to know the difference: Strategy-situation fit in emotion regulation in daily life is associated with well-being. *Psychological Science*, 27(12), 1651–1659. <https://doi.org/10.1177/0956797616669086>
- Hannan, S. M., & Orcutt, H. K. (2020). Emotion regulation in undergraduate students with posttraumatic stress symptoms: A multimethod study. *Psychological Trauma: Theory, Research, Practice, and Policy*, 12(6), 643–650. <https://doi.org/10.1037/tra0000577>

- Hay, A. C., Sheppes, G., Gross, J. J., & Gruber, J. (2015). Choosing how to feel: Emotion regulation choice in bipolar disorder. *Emotion*, 15(2), 139–145. <https://doi.org/10.1037/emo0000024>
- Heiy, J. E., & Cheavens, J. S. (2014). Back to basics: A naturalistic assessment of the experience and regulation of emotion. *Emotion*, 14(5), 878–891. <https://doi.org/10.1037/a0037231>
- Kamradt, J. M., Ullsperger, J. M., & Nikolas, M. A. (2014). Executive function assessment and adult attention-deficit/hyperactivity disorder: Tasks versus ratings on the Barkley Deficits in Executive Functioning Scale. *Psychological Assessment*, 26(4), 1095–1105. <https://doi.org/10.1037/pas0000006>
- Koole, S. L., Webb, T. L., & Sheeran, P. L. (2015). Implicit emotion regulation: Feeling better without knowing why. *Current Opinion in Psychology*, 3, 6–10. <https://doi.org/10.1016/j.copsyc.2014.12.027>
- Lee, K. M., Ferreira-Santos, F., & Satpute, A. B. (2021). Predictive processing models and affective neuroscience. *Neuroscience & Biobehavioral Reviews*, 131, 211–228. <https://doi.org/10.1016/j.neubiorev.2021.09.009>
- Lindquist, K. A., Barrett, L. F., Bliss-Moreau, E., & Russell, J. A. (2006). Language and the perception of emotion. *Emotion*, 6(1), 125–138. <https://doi.org/10.1037/1528-3542.6.1.125>
- Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: A meta-analytic review. *The Behavioral and Brain Sciences*, 35(3), 121–143. <https://doi.org/10.1017/S0140525X11000446>
- Loewenstein, G. (1996). Out of Control: Visceral Influences on Behavior. *Organizational Behavior and Human Decision Processes*, 65(3), 272–292.

- Malanchini, M., Engelhardt, L. E., Grotzinger, A. D., Harden, K. P., & Tucker-Drob, E. M. (2019). “Same but different”: Associations between multiple aspects of self-regulation, cognition, and academic abilities. *Journal of Personality and Social Psychology*, *117*(6), 1164–1188. <https://doi.org/10.1037/pspp0000224>
- Matthews, M., Webb, T. L., Shafir, R., Snow, M., & Sheppes, G. (2021). Identifying the determinants of emotion regulation choice: A systematic review with meta-analysis. *Cognition and Emotion*, *35*(6), 1056–1084. <https://doi.org/10.1080/02699931.2021.1945538>
- McRae, K., & Gross, J. J. (2020). Emotion regulation. *Emotion*, *20*(1), 1–9. <https://doi.org/10.1037/emo0000703>
- McRae, K., Ochsner, K. N., Mauss, I. B., Gabrieli, J. J. D., & Gross, J. J. (2008). Gender differences in emotion regulation: An fMRI study of cognitive reappraisal. *Group Processes & Intergroup Relations*, *11*(2), 143–162. <https://doi.org/10.1177/1368430207088035>
- Miller, L. C., Shaikh, S. J., Jeong, D. C., Wang, L., Gillig, T. K., Godoy, C. G., Appleby, P. R., Corsbie-Massay, C. L., Marsella, S., Christensen, J. L., & Read, S. J. (2019). Causal Inference in Generalizable Environments: Systematic Representative Design. *Psychological Inquiry*, *30*(4), 173–202. <https://doi.org/10.1080/1047840X.2019.1693866>
- Mobbs, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., Dolan, R. J., & Frith, C. D. (2007). *When Fear Is Near: Threat Imminence Elicits Prefrontal–Periaqueductal Gray Shifts in Humans*. *317*, 6.
- Mohammad, S. (2018). Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. *Proceedings of the 56th Annual Meeting of the Association for*

- Computational Linguistics (Volume 1: Long Papers)*, 174–184.
<https://doi.org/10.18653/v1/P18-1017>
- Norem, J. K. (2008). Defensive Pessimism, Anxiety, and the Complexity of Evaluating Self-Regulation: Defensive Pessimism, Anxiety, and Self-Regulation. *Social and Personality Psychology Compass*, 2(1), 121–134. <https://doi.org/10.1111/j.1751-9004.2007.00053.x>
- Opitz, P. C., Cavanagh, S. R., & Urry, H. L. (2015). Uninstructed emotion regulation choice in four studies of cognitive reappraisal. *Personality and Individual Differences*, 86, 455–464. <https://doi.org/10.1016/j.paid.2015.06.048>
- Opitz, P. C., Gross, J. J., & Urry, H. L. (2012). Selection, Optimization, and Compensation in the Domain of Emotion Regulation: Applications to Adolescence, Older Age, and Major Depressive Disorder: SOC-ER Applications. *Social and Personality Psychology Compass*, 6(2), 142–155. <https://doi.org/10.1111/j.1751-9004.2011.00413.x>
- Orejuela-Dávila, A. I., Levens, S. M., Sagui-Henson, S. J., Tedeschi, R. G., & Sheppes, G. (2019). The relation between emotion regulation choice and posttraumatic growth. *Cognition & Emotion*, 33(8), 1709–1717. <https://doi.org/10.1080/02699931.2019.1592117>
- R Core Team. (2022). *R: A language and environment for statistical computing*. [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ridderinkhof, K. R. (2017). Emotion in Action: A Predictive Processing Perspective and Theoretical Synthesis. *Emotion Review*, 9(4), 319–325. <https://doi.org/10.1177/1754073916661765>
- Robert W. Levenson, Levenson, R. W., John M. Gottman, & Gottman, J. M. (1983). Marital interaction: Physiological linkage and affective exchange. *Journal of Personality and Social Psychology*, 45(3), 587–597. <https://doi.org/10.1037/0022-3514.45.3.587>

- Rottweiler, A.-L., Taxer, J. L., & Nett, U. E. (2018). Context Matters in the Effectiveness of Emotion Regulation Strategies. *AERA Open*, 4(2), 233285841877884. <https://doi.org/10.1177/2332858418778849>
- Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., Sams, M., & Nummenmaa, L. (2016). Discrete Neural Signatures of Basic Emotions. *Cerebral Cortex*, 26(6), 2563–2573. <https://doi.org/10.1093/cercor/bhv086>
- Sauer, C., Sheppes, G., Lackner, H. K., Arens, E. A., Tarrasch, R., & Barnow, S. (2016). Emotion regulation choice in female patients with borderline personality disorder: Findings from self-reports and experimental measures. *Psychiatry Research*, 242, 375–384. <https://doi.org/10.1016/j.psychres.2016.04.113>
- Sayette, M. A., Loewenstein, G., Griffin, K. M., & Black, J. J. (2008). Exploring the Cold-to-Hot Empathy Gap in Smokers. *Psychological Science*, 19(9), 926–932. <https://doi.org/10.1111/j.1467-9280.2008.02178.x>
- Shafir, R., Thiruchselvam, R., Suri, G., Gross, J. J., & Sheppes, G. (2016). Neural processing of emotional-intensity predicts emotion regulation choice. *Social Cognitive and Affective Neuroscience*, 11(12), 1863–1871. <https://doi.org/10.1093/scan/nsw114>
- Shahane, A. D., Godfrey, D. A., & Denny, B. T. (2023). Predicting real-world emotion and health from spontaneously assessed linguistic distancing using novel scalable technology. *Emotion*. <https://doi.org/10.1037/emo0001211>
- Sheppes, G. (2020). Transcending the “good & bad” and “here & now” in emotion regulation: Costs and benefits of strategies across regulatory stages. In B. Gawronski (Ed.), *Advances in Experimental Social Psychology* (Vol. 61, pp. 185–236). Academic Press.

- Sheppes, G., Brady, W. J., & Samson, A. C. (2014). In (visual) search for a new distraction: The efficiency of a novel attentional deployment versus semantic meaning regulation strategies. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00346>
- Sheppes, G., & Gross, J. J. (2011). Is Timing Everything? Temporal Considerations in Emotion Regulation. *Personality and Social Psychology Review*, 15(4), 319–331. <https://doi.org/10.1177/1088868310395778>
- Sheppes, G., Scheibe, S., Suri, G., & Gross, J. J. (2011). Emotion-Regulation Choice. *Psychological Science*, 22(11), 1391–1396. <https://doi.org/10.1177/0956797611418350>
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>
- Specker, P., & Nickerson, A. (2022). An experimental investigation of spontaneous emotion regulation variability, negative affect, and posttraumatic stress disorder among traumatized refugees. *Psychological Trauma: Theory, Research, Practice, and Policy*. <https://doi.org/10.1037/tra0001217>
- Specker, P., Sheppes, G., & Nickerson, A. (2024). Does Emotion Regulation Flexibility Work? Investigating the Effectiveness of Regulatory Selection Flexibility in Managing Negative Affect. *Social Psychological and Personality Science*, 15(5), 561–569. <https://doi.org/10.1177/19485506231189002>
- Stasiak, J. E., Mitchell, W. J., Reisman, S. S., Gregory, D. F., Murty, V. P., & Helion, C. (2023). Physiological arousal guides situational appraisals and metacognitive recall for naturalistic experiences. *Neuropsychologia*, 180, 108467. <https://doi.org/10.1016/j.neuropsychologia.2023.108467>

- Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials*, 24(2), 182–199. [https://doi.org/10.1016/S0197-2456\(02\)00320-3](https://doi.org/10.1016/S0197-2456(02)00320-3)
- Suri, G., Sheppes, G., Young, G., Gerald Young, Abraham, D., McRae, K., & Gross, J. J. (2018). Emotion regulation choice: The role of environmental affordances. *Cognition & Emotion*, 32(5), 963–971. <https://doi.org/10.1080/02699931.2017.1371003>
- Szasz, P. L., Madalina Coman, Coman, M. A., Curtiss, J., Carpenter, J. K., & Hofmann, S. G. (2018). Use of Multiple Regulation Strategies in Spontaneous Emotion Regulation. *International Journal of Cognitive Therapy*, 11(3), 249–261. <https://doi.org/10.1007/s41811-018-0026-9>
- Tamir, M. (2016). Why Do People Regulate Their Emotions? A Taxonomy of Motives in Emotion Regulation. *Personality and Social Psychology Review*, 20(3), 199–222. <https://doi.org/10.1177/1088868315586325>
- Tang, Y., & Huang, Y. (2019). Contextual factors influence the selection of specific and broad types of emotion regulation strategies. *British Journal of Social Psychology*, 58(4), 1008–1033. <https://doi.org/10.1111/bjso.12313>
- Tashjian, S. M., Fedrigo, V., Molapour, T., Mobbs, D., & Camerer, C. F. (2022). Physiological responses to a haunted house threat experience: Distinct tonic and phasic effects. *Psychological Science*, 33(2), 236–248. <https://doi.org/10.1177/09567976211032231>
- Uusberg, A., Taxer, J. L., Yih, J., Uusberg, H., & Gross, J. J. (2019). Reappraising Reappraisal. *Emotion Review*, 11(4), 267–282. <https://doi.org/10.1177/1754073919862617>

- Van Boven, L., & Loewenstein, G. (2003). Social Projection of Transient Drive States. *Personality and Social Psychology Bulletin*, 29(9), 1159–1168.
<https://doi.org/10.1177/0146167203254597>
- Watson, D., Anna, L., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070.
- Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin*, 138(4), 775–808. <https://doi.org/10.1037/a0027600>
- Weiss, N. H., Schick, M. R., Waite, E. E., Haliczzer, L. A., & Dixon-Gordon, K. L. (2021). Association of positive emotion dysregulation to resting heart rate variability: The influence of positive affect intensity. *Personality and Individual Differences*, 173, 110607.
<https://doi.org/10.1016/j.paid.2020.110607>
- Wennerhold, L., & Friese, M. (2020). Why Self-Report Measures of Self-Control and Inhibition Tasks Do Not Substantially Correlate. *Collabra: Psychology*, 6(1), 9.
<https://doi.org/10.1525/collabra.276>
- Young, G., & Suri, G. (2020). Emotion regulation choice: A broad examination of external factors. *Cognition & Emotion*, 34(2), 242–261. <https://doi.org/10.1080/02699931.2019.1611544>
- Zhang, Z., & Mai, Y. (2019). *WebPower: Basic and Advanced Statistical Power Analysis* (0.5) [R]. <https://CRAN.R-project.org/package=WebPower>

