Emotion regulation strategy choice in naturalistic contexts

ABSTRACT (146 / 150 Words):

Successful emotion regulation requires effective strategy selection. Prior research suggests that low-effort relative to high-effort strategies are more likely to be selected as emotional experiences intensify. However, experiential sampling and strategy choice methods may not capture intense or realistic emotional events. The present research uses a naturalistic approach to examine strategy selection and implementation in a high-intensity real-world context. Participants navigated a haunted house and reported their emotional states, intensity, and regulation attempts. Studies 1 (n = 47) and 2 (n = 118) found no relationship between emotional intensity and use of high- vs. low-effort strategies. Study 3 participants (n = 152) predicted regulation strategy usage based on the experiences reported by Study 1 participants, and their forecasts matched the regulatory choice associations reported in prior literature. These results suggest that emotional intensity predicts regulation choices in the abstract, but is less predictive in real-world contexts.

STATEMENT OF RELEVANCE (144 / 150 Words):

Changing one's current emotional state to how one wants to feel can be challenging, but successfully doing so is important for well-being and mental health. Prior work shows individuals tend to regulate emotions using common, recognizable strategies and that strategy choice is based in part upon the emotional intensity of the regulated event. However, the approaches of previous studies make it difficult to determine if patterns in deliberative choices reflect actual usage during high-intensity events in the real-world. Our study is the first to examine strategy choice among untrained participants in a naturalistic, highly evocative context, which allows us to explore the association between emotion regulation strategy choice and usage in the real-world. Contrary to prior observations in lab studies, emotional intensity failed to predict regulation strategy choice in our paradigm, potentially reflecting that emotion regulation in real-world contexts relies on more complex computations.

KEYWORDS: emotion regulation, naturalistic stimuli, emotion, strategy selection

INTRODUCTION (0876 / 2000 Words): On comedian Nathan Fielder's HBO series *The Rehearsal*, participants are given the opportunity to prepare for upcoming, emotional life events through extensive simulation and practice. Inherent in this absurdist premise is recognition that theory often fails to materialize in practice. Such concerns plague not only the unfortunate participants in Fielder's experiments, but also every person who has predicted that an award would make them happy, that they could handle a tough critique, or that a movie wouldn't make them cry. The disconnect between prediction and practice is of particular importance in the study of emotion regulation (ER), wherein associations between strategy choice and effectiveness observed in lab settings may not generalize to how emotions are regulated in the real world (Sheppes, 2020).

Research on implementing ER often utilizes tasks in which participants are trained to use regulatory strategies, shown previews of emotional stimuli, given prompts to select preferred strategies, asked to regulate using the selected strategy during an extended stimulus presentation, and to also report their regulatory and affective responses (e.g., Sheppes et al., 2011; 2014). These experiments reliably show less cognitively-taxing strategies like distraction (focusing attention away from an emotion-eliciting stimulus) are selected more often and are more effective as affective intensity increases. Likewise, as affective intensity decreases, engagement strategies like reappraisal (changing the emotional meaning of a stimulus to modulate the emotional reaction) are selected more often and are more effective (Hay et al., 2015; Sheppes et al., 2011, 2014; Orejuela-Dávila et al., 2019; Shafir et al., 2016; Young & Suri, 2020).

However, such optimized approaches may differ from ER usage in real-world contexts, where individuals are untrained, stimuli are not previewed, and behavior is unprompted (Friedman & Gustavson, 2022). While ER choice outcomes have been replicated using auditory (Feldman & Freitas, 2021), visual (Shafir et al., 2018), and physical (Sheppes et al., 2011) stimuli, real world stimuli are multimodal and complex. When allowed, individuals also explore and flexibly apply multiple strategies that often blur the boundaries of typical strategy classifications (Opitz, Cavanagh, and Urry, 2015; Heiy & Ceavens, 2014; Szasz et al., 2018; Aldao & Nolen-Hoeksema, 2013). These discrepancies have prompted researchers to emphasize capturing ER in context as the next crucial step for the field (Tang & Huang, 2019; Rottweiler, Taxer, & Nett, 2018; Dixon-Gordon et al., 2015; Aldao, 2013).

Experiential sampling and ecological momentary assessment approaches capture strategy choice directly within naturalistic settings (e.g., Haines et al., 2016; Heiy & Ceavens, 2014; Colombo et al., 2020), avoiding many ecological issues. However, capturing emotion in relatively mundane circumstances often leads to relatively average emotional responses, which might not necessarily reflect behaviors at emotional extremes; when ER usage is perhaps most consequential. A paradigm which reliably captures naturalistic ER responses in controlled but high-intensity environments has not yet been established within the field, and little is known about whether ER behavior observed within lab settings in response to high-intensity stimuli generalizes to situations outside of the lab. The goal of the present research is to examine whether these regulatory choice patterns emerge in a sample of untrained participants exposed to a controlled but naturalistic setting high in emotional variability: an immersive haunted house.

We focused on one particularly robust predictor of ER choice from the lab literature: emotional intensity (Sheppes, 2011). A recent meta-analysis found a very large effect size in the relationship

between strategy choice and emotional intensity (r+ = 0.46 – 0.61) (Matthews et al., 2021). To approximate the design of strategy choice paradigms, we focused on observations in which participants reported regulatory behaviors congruent with either distraction or reappraisal in response to negatively valenced emotional events. We focused on these strategies specifically because of their near-ubiquity in the literature (Heiy & Cheavens, 2014) and effectiveness in meeting pro-hedonic emotional goals (English et al., 2017).

We hypothesized that results from a relatively more naturalistic design may not replicate lab findings. Spontaneous or untrained ER, such as what we expected our participants to engage in, may rely more heavily on implicit, reflexive processes (Koole et al., 2015) wherein automatic cognitions like habits and schemas (Christou-Champi, Farrow, & Webb, 2014; Norem, 2007) guide regulatory behaviors more than characteristics of the context, such as emotional intensity. There's less room for deliberation and, as such, we predicted that we would *not* find significant associations between affective intensity and strategy choice when measuring ER in naturalistic unconstrained contexts. We also hypothesized that decontextualizing those same events, or presenting stimuli in a lab setting like lab studies do, would allow for more deliberative processes and produce a pattern of results more closely mirroring lab findings.

We used three experiments to explore these aims. In Experiment 1, we tested whether affective intensity failed to predict selection of low-effort (distraction) versus high-effort (reappraisal) regulatory strategies in naturalistic settings. To test this prediction, untrained participants navigated a haunted house and later reported emotional and spontaneous regulatory behaviors in a surprise recall task. Experiment 2 replicated and improved upon the design of Experiment 1 by doubling sample size and observing changes in affective and regulatory reporting over time. Experiment 3 built upon Experiments 1 and 2, using the events described by those who experienced the haunted house in Experiment 1 to test how a new set of participants might forecast regulatory behaviors.

EXPERIMENT 1 (0957 / 2000 Words): Experiment 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 naturalistic setting with an untrained sample. To assess emotional intensity, participants self-reported the emotional intensity of events from the haunted house one week after exposure during a surprise recall task. Participants also noted whether they wanted to reduce the intensity of these emotions and, if so, how they attempted to do so in their own words.

METHOD (1553 Words):

PARTICIPANTS: In October 2019, 54 participants ($\bar{x}_{age}$ = 24.22 yrs, range = 18 - 34 yrs, sd_{age} = 3.97 yrs, 26 female, 1 non-binary) were recruited from a large northeastern city via flyers for an IRB-approved fear and memory study. A priori power analyses using the WebPower (Zhang, Z., & Mai, Y., 2019) in R 3.6.1 (R Core Team, 2022) determined 18 participants would sufficiently power our main effect using the smallest effect size reported by Sheppes et al.'s 2011 examination of emotional intensity and regulatory choice ($\eta_p^2 = 0.43$). Participants were excluded for previously visiting the haunted house (n = 1), not completing the study (n = 1), identifying English as their second-language (n = 2), or not following instructions (n = 3). Participants received $70.00 in Visa debit cards for participating.

MATERIALS AND PROCEDURE: Our design consisted of an exposure session and a follow-up. Following consent, participants completed computerized questionnaires and were fitted with physiological monitors which are beyond the purview of this study (*See* [REMOVED FOR BLINDING]). Participants were then escorted by two research assistants to the remotely-located haunted house.

This specific haunted house experience was chosen because: A) it uses professional actors renowned for eliciting a range of affective responses, B) it contained six themed sections each with a unique aesthetic providing variability to the stimuli, C) it provides a remarkably consistent experience across sessions, and D) coordination with the facility allowed us to enter the haunted house before other patrons to guarantee consistency.

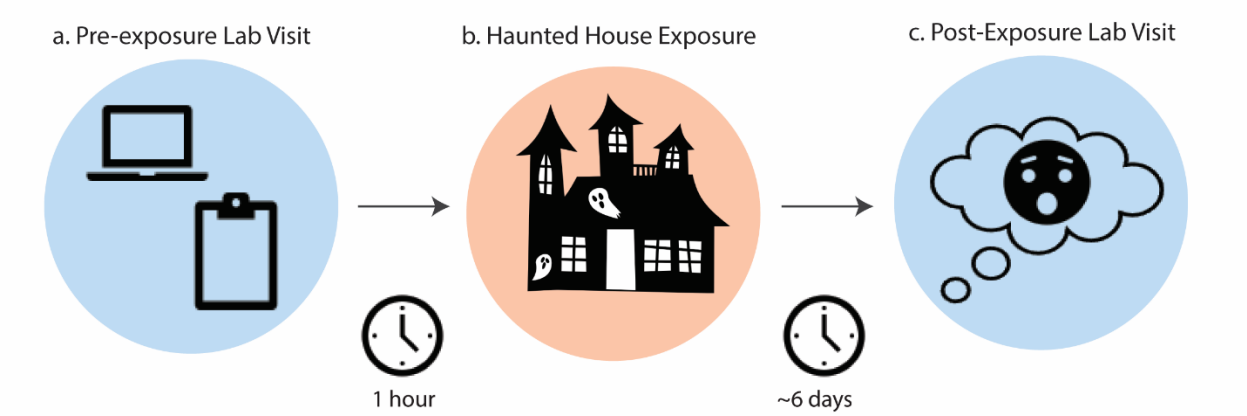


Fig 1. Study 1: Task Overview - Fifty-four (54) participants traversed a haunted house in small groups. a. Prior to the haunted house, participants completed baseline questionnaires. b. The haunted house lasted for ~1 hour. c) Participants returned to the laboratory ~6 days later for the emotion regulation recall task.

Session 1, Haunted House. Participants navigated the haunted house in twelve groups ($\bar{x}_{size}$ = 4.50 participants; sd_{size} = 0.79 participants) for approximately 55.40 minutes (*sd* = 5.05 minutes) and were provided with 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). However, they were encouraged to act and react as naturally as possible. Each participant was randomly assigned to lead the group through one section. The accompanying research assistant led the group through any sections without a participant-leader. Following exposure, participants were scheduled for an individual follow-up and were instructed to not discuss their experience with anyone.

Session 2, Laboratory follow-up session. At follow-up (*time since exposure:* $\bar{x}_{delay}$ = 5.98 days; sd_{delay} = 0.79 days), participants completed a surprise free-recall memory task and questionnaires. Notably, participants identified, described, and chronologically ordered ten (10) discrete events from within the haunted house. For each event, participants identified which of 13 emotion categories they had felt, the intensity of each of those emotions, the extent to which they tried to regulate each of those emotions, and to describe how they attempted to regulate them (if at all). It must be noted that while participants were able to endorse multiple emotions of differing intensities for any one event, participants were only asked about their regulatory behavior once per event and not whether that regulatory behavior was directed towards any specific emotions endorsed during

that event. Thus, because we have greater granularity of emotion than we do the regulatory responses to those emotions, the association between these variables was assessed through multiple approaches to account for this discrepancy. To avoid confusion, we refer to this approach as capturing regulatory behaviors at the "event-level". During Study 2, we ask participants about their regulatory responses to each emotion they endorsed, which we refer to as capturing regulatory responses at the "emotion-level".

Participants were not trained in emotion regulation strategies, nor were they primed to consider their emotion regulation strategies prior to these questionnaires. Emotion categories were adapted from the Positive and Negative Affect Schedule (PANAS) (Watson, Clark, & Tellegen, 1988). Some noted additions relevant to a typical haunted house experience included "tense", and "disgusted". Ten of the 13 options were negatively-valenced emotions (i.e., *Disgusted/Grossed Out, Fearful/Afraid, Hostile/Aggressive, Irritable/Annoyed, Nervous/Jittery, Overwhelmed, Panicked, Shocked/Surprised, Tense, Upset/Distressed*). Applying Cronbach's alpha to the emotional intensity value of negative emotions yielded a value of $\alpha = 0.91$ (95% CI = [0.89, 0.92]), suggesting excellent internal consistency. Both emotional intensity and regulation extent were captured on a 7-point Likert scale, with 1 representing "Not at all" and 7 representing "A great deal". Regulation strategies were captured for each event using free-response to the prompt: *If you did attempt to change or regulate your emotions, how did you do so?* Participants were subsequently debriefed and paid for their participation.

Questionnaire response coding. Two hypotheses-blind raters classified strategy descriptions into one or more strategy categories: Reappraisal, Distraction, another Process Model strategy (e.g., Suppression, Situation Modification), a combination of the three, or none of the above (IRR Agreement = 0.88). Raters were not privy to the emotion intensity ratings participants endorsed. 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 the lab (Sheppes, 2011; Shafir, 2016). Lastly, raters independently completed classification exercises using examples of regulation strategy descriptions not included in the primary sample set. 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 researcher 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 was not involved in classifications and was muted and had their video off during these compromise exchanges to avoid undue influence.

Analysis. To explore our primary question, the effect of emotional intensity upon regulatory strategy usage, we specified multilevel binary logistic regressions accounting for the random effect of participants using the "lme4" package (Bates et al., 2015) in R (R Core Team, 2022). We followed an information theoretic approach via AIC comparison. Our model comparisons included a: A) a null model without fixed effects, B) a model containing only z-scored emotional intensity as a fixed effect, C) a model containing only person-centered, z-scored emotional intensity and person-mean, z-scored emotional intensity as fixed effects, and D) a model including person-centered and person-mean z-scored emotional intensity with covariates.

RESULTS: As we primarily aimed to determine the replicability of results from ER strategy choice paradigms in naturalistic contexts, observations that did not contain negative emotions regulated by either distraction or reappraisal were beyond the purview of our study. Our full dataset consisted of 469 unique events from 47 participants with an average of 2.43 emotions (SD) endorsed per event. Of the 1138 endorsed emotions, 603 (52.99%) were classified as being negatively valenced. Of the 603 negative emotions endorsed, there were 166 observations in which a negative emotion was downregulated by either distraction or reappraisal. These 166 observations came from 78 unique events reported by 32 participants. Of the 78 unique events, 57 events (74.36%), or 130 observations (78.31%) reported using distraction to regulate their emotions. The average emotional intensity of observations was 5.55 (range: 1 – 7, Likert scale).

Intensity predicts regulatory extent. To determine whether emotional intensity was associated with the extent to which individuals reported attempting to regulate affective responses, we ran multilevel linear models specifying regulatory extent as our criterion variable, participant as a random intercept, and building fixed effects from a null model (ICC = 0.35). Our best performing multilevel linear model as determined by model AIC comparison found person-centered emotional intensity to be a significant predictor of regulation extent ($\beta = 0.31$, 95% CI = [0.17, 0.46], $p < 0.001$). This suggests our paradigm elicited regulatory behaviors from participants in a predictable manner – that is, that more intense emotions led to increased attempts to regulate emotion.

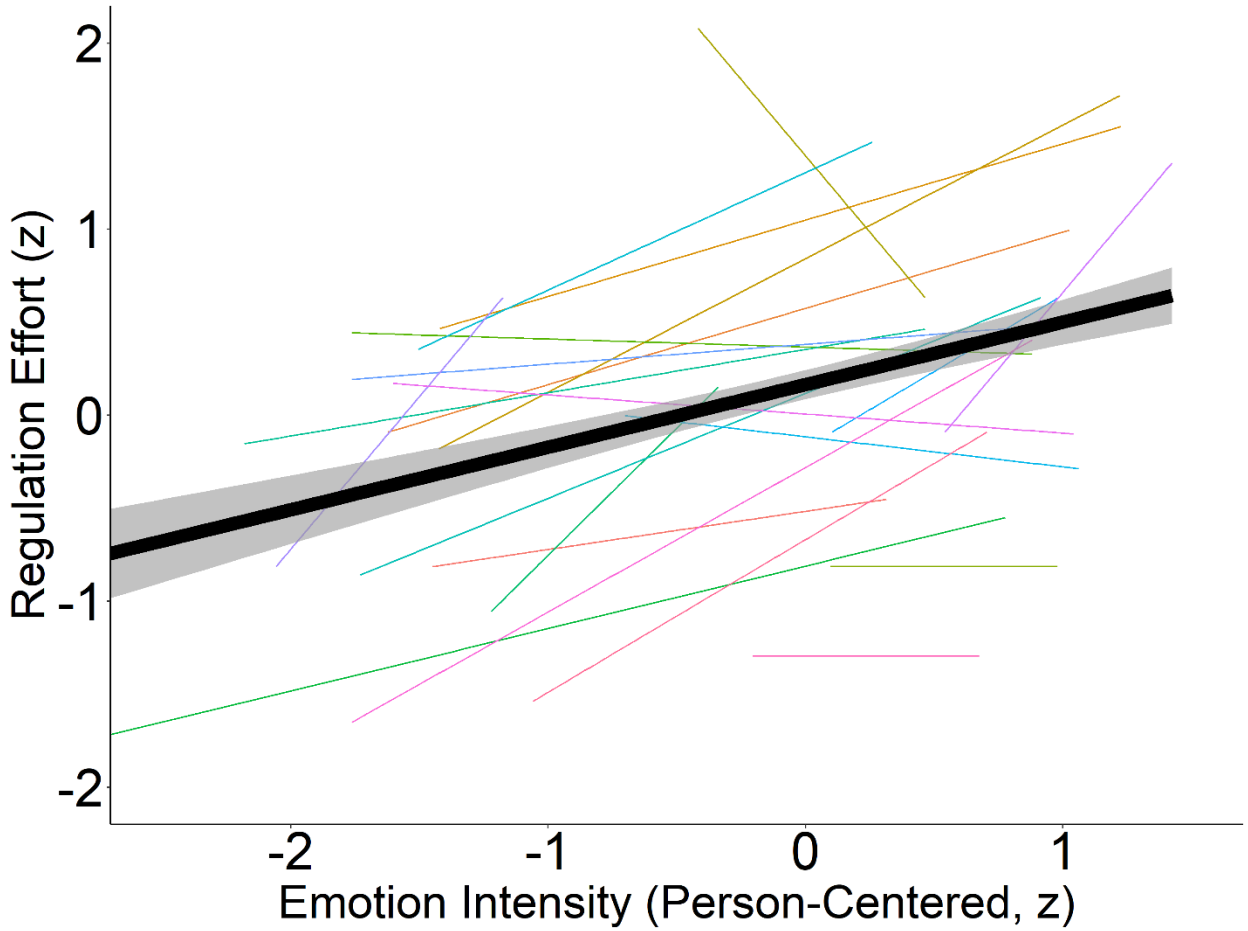


Fig 2. Emotional intensity predicts regulation extent (β = 0.31, p < 0.001) using a person-centered mixed effects linear model. The thick black line represents the sample trend while thinner lines represent the trajectories of individual participants. Regression ribbon represents standard error.

Intensity does not predict regulatory strategy choice. Using multilevel binary logistic regression models specifying regulation strategy choice as our criterion variable, participant as a random intercept, and building fixed effects from a null model (ICC = 0.70), we failed to find any model that performed better than our null model by AIC comparison. Our model using only emotional intensity to predict regulation strategy choice trended significant in the model comparison (p = 0.063). However, even if traditional statistical thresholds were loosened and the model was deemed superior to our null, within that model, we did not find that emotional intensity demonstrated predictive utility towards strategy usage in that model (OR = 1.83, 95% CI = [0.65, 3.2], p = 0.079).

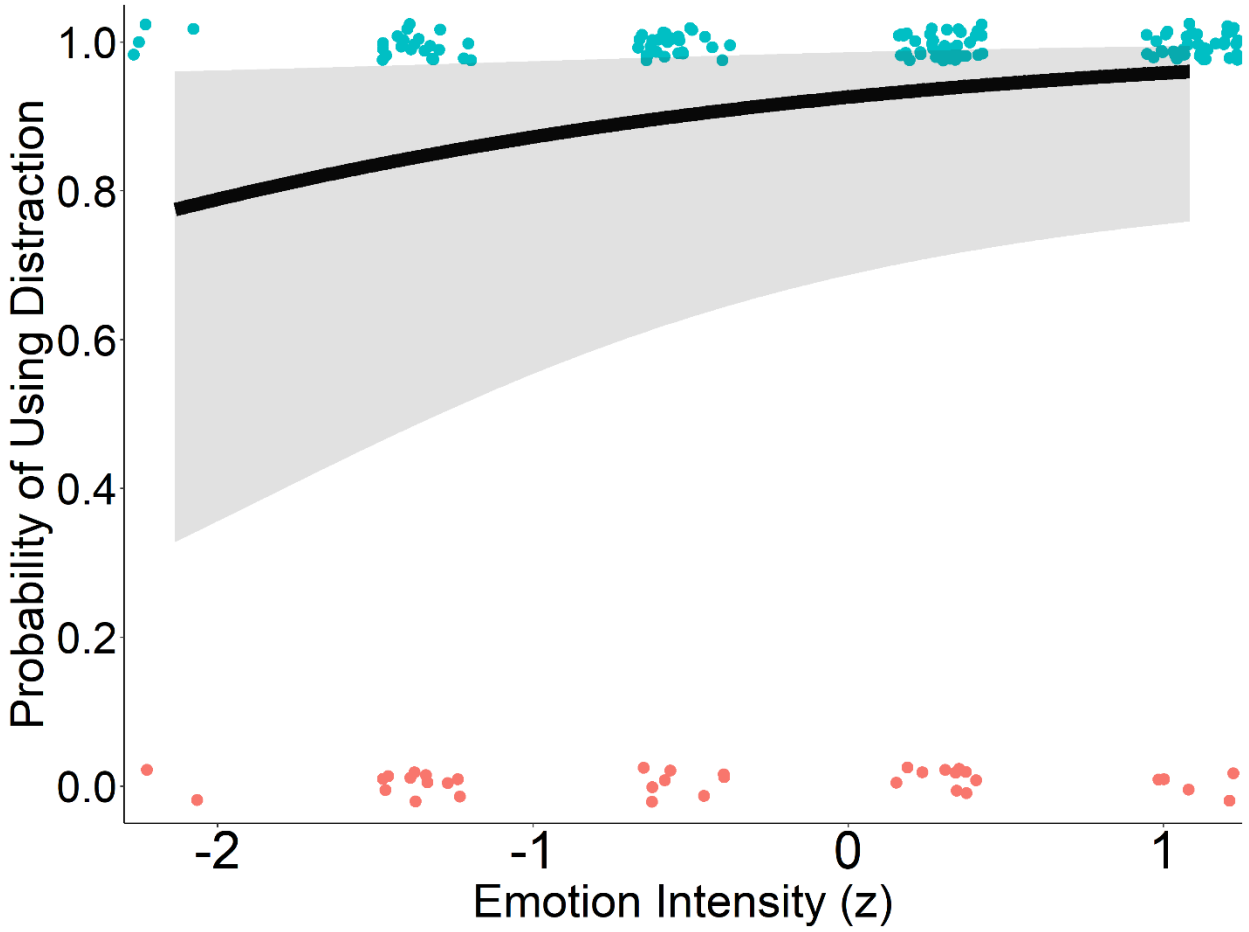


Fig 3. Across all tested mixed effects binary logistic regression models, emotional intensity failed to predict strategy usage. Visualized is our model using only emotional intensity to predict regulation strategy choice. Regression line represents likelihood of selecting distraction as opposed to reappraisal at any given emotional intensity value. Points represent individual observations. Regression ribbon represents standard error.

Because of the previously noted discrepancy in granularity of capturing emotion and regulation, analyzing the relationship between individual emotions and event-level regulatory strategies may be missing stronger relationships between emotions and regulation that exist when analyzing data at the event-level only. As such, additional analyses were conducted using the average emotional intensity of each event to predict strategy choice as well as the sum of emotional intensity for each event to predict strategy choice. In both cases, multilevel binary logistic regressions failed to perform better than the null model (ICC = 0.28; Emotion Sum: $p = 0.130$; Emotion Average: $p = 0.430$) and none of the affective variables demonstrated predictive utility towards strategy choice, reinforcing the results of our primary analysis. Thus, regardless of whether emotions are considered individually or concurrently, we do not find evidence to support an association between affective intensity and ER strategy usage in this context.

EXPERIMENT 2 (1031 / 2000 Words): Experiment 1 found that emotional intensity failed to predict which strategies people actually used. In Experiment 2, we aimed to replicate and address limitations from Experiment 1 by A) doubling our sample size, B) capturing experiences

immediately after exposure and at one week post-exposure to assess recall biases, C) assessing regulatory behaviors at the emotion-level D) capturing emotion in free response, and E) measuring notable covariates such as cognitive load, affective expectations, and motivations.

METHOD (1643):

PARTICIPANTS: In October, 2021, 118 participants ($\bar{x}_{age}$ = 20.80 yrs, range = 18 – 34 yrs, sd_{age} = 2.87 yrs, 79 female) were recruited from a large northeastern city via flyers for an IRB-approved fear and memory study. Participants were compensated $60.00 in Visa debit cards.

MATERIALS AND PROCEDURE: Study 2 primarily mirrored Study 1 in design; therefore, we focus on deviations from Study 1.

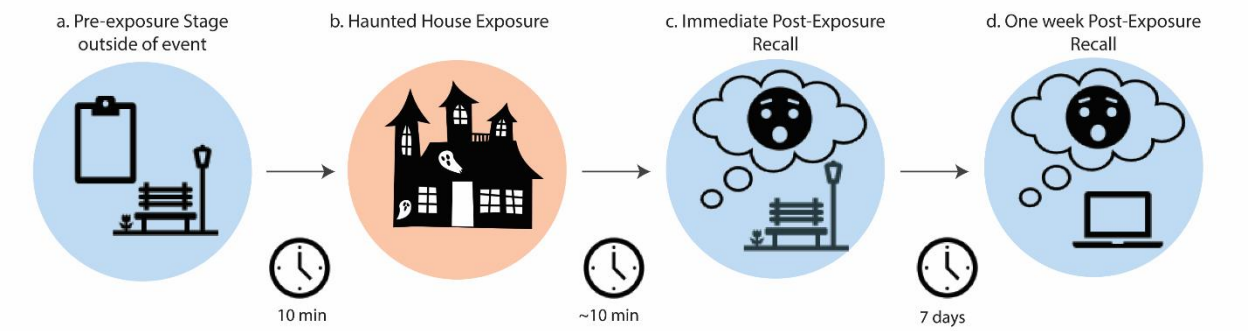


Fig 4. Study 2: Task Overview - One hundred and eighteen (118) participants traversed a haunted house in small groups. a. Prior to the haunted house, participants completed baseline questionnaires outside of the event at a local park. b. The haunted house lasted for ~37 minutes. c) Participants then 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.

Participants reported to a remote site located on 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. Cognitive load was assessed prior to exposure, immediately after exposure and at 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 et al., 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.

Session 1, Haunted House. Due to COVID-related safety measures, the haunted house restructured to a non-linear format and reduced the number of themed sections. Our participants no longer entered before other patrons, but the format allowed the actors to reset before each new group entered the haunted house. Participants were binned into 31 groups across 11 nights ($\bar{x}_{size}$ = 3.81 participants; sd_{size} = 1.12 participants). Participants were briefed with the same instructions as Study 1 before entry. The accompanying research assistant led the group through each section. The approximate exposure time was 37.40 minutes.

Session 1, Immediate follow-up session. Following exposure, participants completed immediate post-exposure assessments at the remote site. This included a surprise free recall and questionnaire for a randomly selected haunted house section. Participants were also tasked with identifying three emotionally salient events that occurred within that section and reporting affective and regulatory details of that event. Like Study 1, participants described the events, noted which emotions they felt, how intense those emotions were, and described how they tried to regulate those emotions, if at all. However, free response replaced emotion categories to better reflect the natural, idiosyncratic affective experiences of participants. 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. We refer to data captured at this time point as being "immediately reported". These changes allowed us to examine regulation with greater precision than in Study 1, during which regulation was captured at the event-level and success and direction could only be inferred. Following completion of immediate post-exposure measures, participants were dismissed, instructed to not discuss their experiences, and to remain in contact with researchers for the one-week follow-up.

Session 2, Online follow-up session. Follow-up sessions were conducted remotely (*time since exposure:* $\bar{x}_{delay}$ = 7.01 days; sd_{delay} = 0.91 days), though staff were readily available to respond to participant issues and questions. At follow-up, participants were reminded of the events they identified at immediately post-exposure with a keyword they provided to summarize the event. Participants were then tasked with recalling the affective and regulatory details of those previously recorded events. We refer to this data as "delayed recall". Participants also identified 6 new events that they would be reporting on for the first time, which we refer to as "delayed reporting". This allowed us to assess how the report of affective and regulatory event details may have changed between exposure and follow-up.

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.

Two hypotheses-blind raters classified each observation's strategy description into one or more strategy categories: Reappraisal, Distraction, a combination of the two, or none of the above (IRR Agreement = 0.877). Raters were undergraduate research assistants who were trained using the same methodology described in Study 1.

To explore our primary question, the effect of emotional intensity upon regulatory strategy usage, we again 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. Preregistration for Experiment 2 methods and hypotheses is publicly available at As Predicted (https://aspredicted.org/DP1_453).

RESULTS: A subset of 436 observations in which a negative emotion was downregulated by either distraction or reappraisal were used for these analyses. These observations came from 302 unique events reported by 84 participants across both timepoints. Of the 436 total observations, 254 (58.30%) reported using distraction to regulate their emotions. The average emotional intensity of observations was 2.34 (range: 0 – 4, Likert scale).

There were no differences in reported emotional intensity between immediate and delayed recall. It may be the case that emotions and related behavior experienced at one point do not correlate with how they are described at a later timepoint, which would be a significant issue for the findings of Study 1. As such, our first analyses aimed to retroactively determine the viability of examining emotion and regulation after a delay, as we had in Study 1, by comparing changes in affective and regulatory detail reporting over time. The z-scored emotional intensity of events immediately reported after exposure and reported after a delay (one week post-exposure) were not statistically different, as determined by a two-sample t-test (x_{diff} = 0.10, 95% CI = [-0.12, 0.31], t(297) = 0.90, p = 0.4) and the frequency of strategy reporting between the two time points also failed to demonstrate statistically significant differences, as determined by a chi square test (x^2= 0.4, p = 0.8).

Participants accurately remembered their reported strategy choice, but not the emotion they reported experiencing. When comparing events reported immediately after exposure to what they recalled reporting one week later, we found that 84.62% of matched observations were classified with the same regulation strategy. However, we also found participants were only able to recall the same emotion they initially reported 26.21% of the time and a paired t-test revealed that the emotional intensity between these matched events decreased by an average of 0.29 standard deviation units over time (95% CI = [-0.53, -0.062, t(57) = -3, p = 0.01). Though, in comparing one-week post-exposure recall to events reported for the first time one-week post-exposure, we only find a trending statistically significant difference in affective intensity (x_{diff} = 0.20, 95% CI = [-0.40, 0.02], t(278) = -2, p = 0.06) and frequency of strategy usage. These results finding differences in average intensity between immediate report and delayed recall, but not immediate report and delayed report or delayed recall and delayed report may be indicative of affective labeling as a regulatory mechanism. Regardless, because of the lack of consistency in recalling the same emotion labels, we can only offer mixed support for the aim of our initial analyses.

Intensity does not predict regulatory strategy choice. To test our primary hypotheses, models using either z-scored emotional intensity or person-centered emotional intensity as the primary predictor were constructed, but across all model comparisons, including those adjusting for gender, age, motivations for participating, the extent to which participants anticipated feeling positive or negative emotions, cognitive load, an individual's general enjoyment of fear, State-Trait Anxiety Inventory scores, or Emotion Regulation Questionnaire scores , no model performed better than our null (ICC = 0.488). Our best performing non-null model, including only intensity as a fixed effect (p = 0.16 when compared to null), failed to demonstrate that emotional intensity had predictive utility towards usage (OR = 1.27, 95% CI = [0.92, 1.75], p = 0.15). This lack of effect

was observed across time points as well. As such, we did not find evidence to support that emotional intensity predicts strategy usage in naturalistic circumstances.

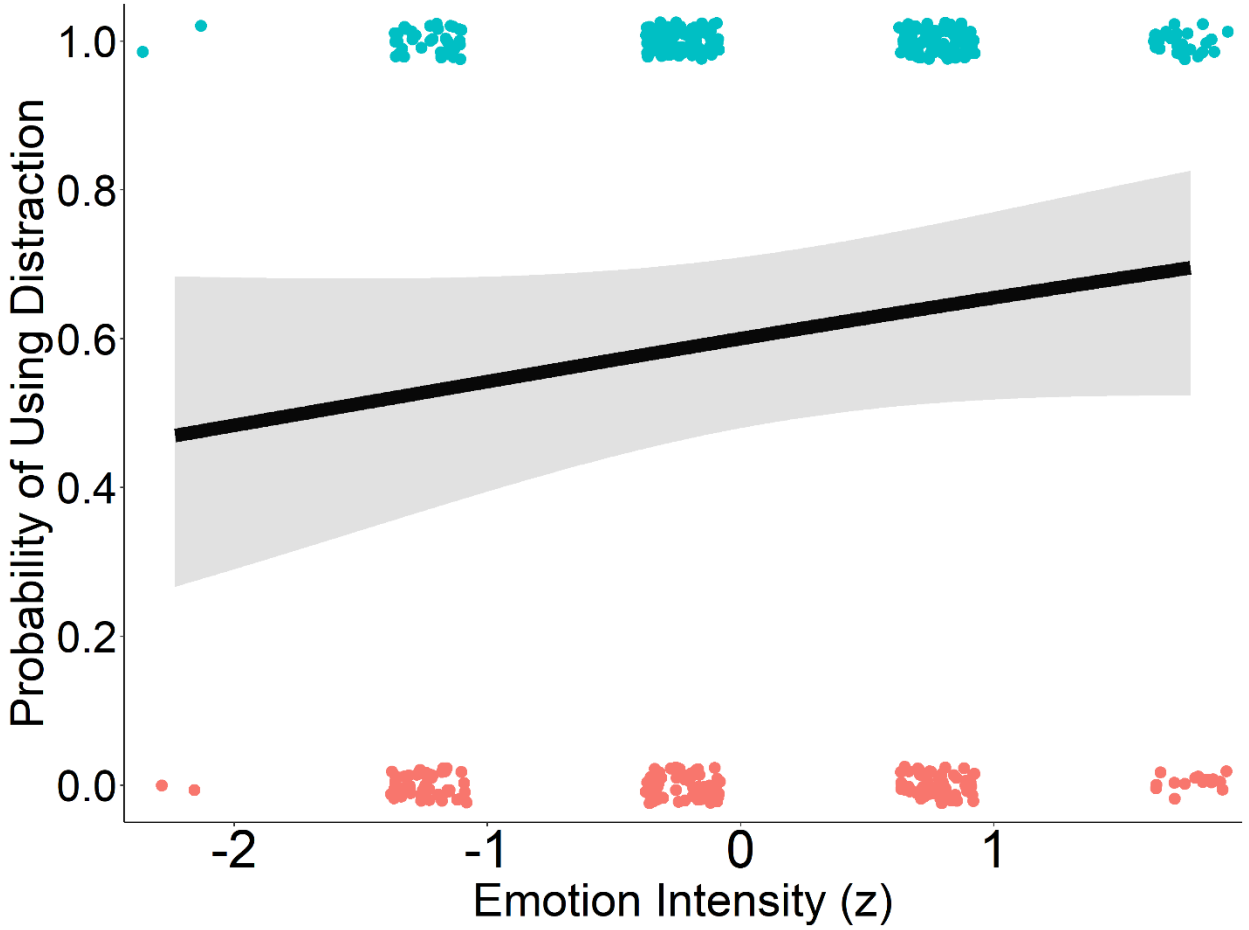


Fig 5. Across all tested mixed effects binary logistic regression models, emotional intensity again failed to predict strategy usage. Visualized is our model using only emotional intensity to predict regulation strategy choice. Regression line represents likelihood of selecting distraction as opposed to reappraisal at any given emotional intensity value. Points represent individual observations. Regression ribbon represents standard error.

Regulatory strategy choice and intensity interact to predict regulatory success. Following our emotional intensity analyses, we pivoted towards examining the association among regulatory choice, regulatory success and affective intensity as another means of potentially supporting lab hypotheses. Prior studies (Sheppes, 2011) suggest that high-intensity events using distraction should more successfully regulate emotions than high-intensity events using reappraisal. A preliminary t-test using z-scored regulatory success suggested that, in general, attempts to regulate using reappraisal were reportedly more successful than those using distraction (x_{diff} = 0.53, t(394) = 6, $p < 0.001$). 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 (β = 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 (β = -0.03, 95% CI = [-0.16, 0.10], p

379 = 0.70), but regulatory success was negatively associated with emotional intensity for distraction-
380 regulated events (β = -0.28, 95% CI = [-0.40, -0.16], p < 0.001). As such, our data suggests that
381 the efficacy of using distraction in high-intensity, naturalistic settings does not match that which
382 has been established with lab paradigms.
383

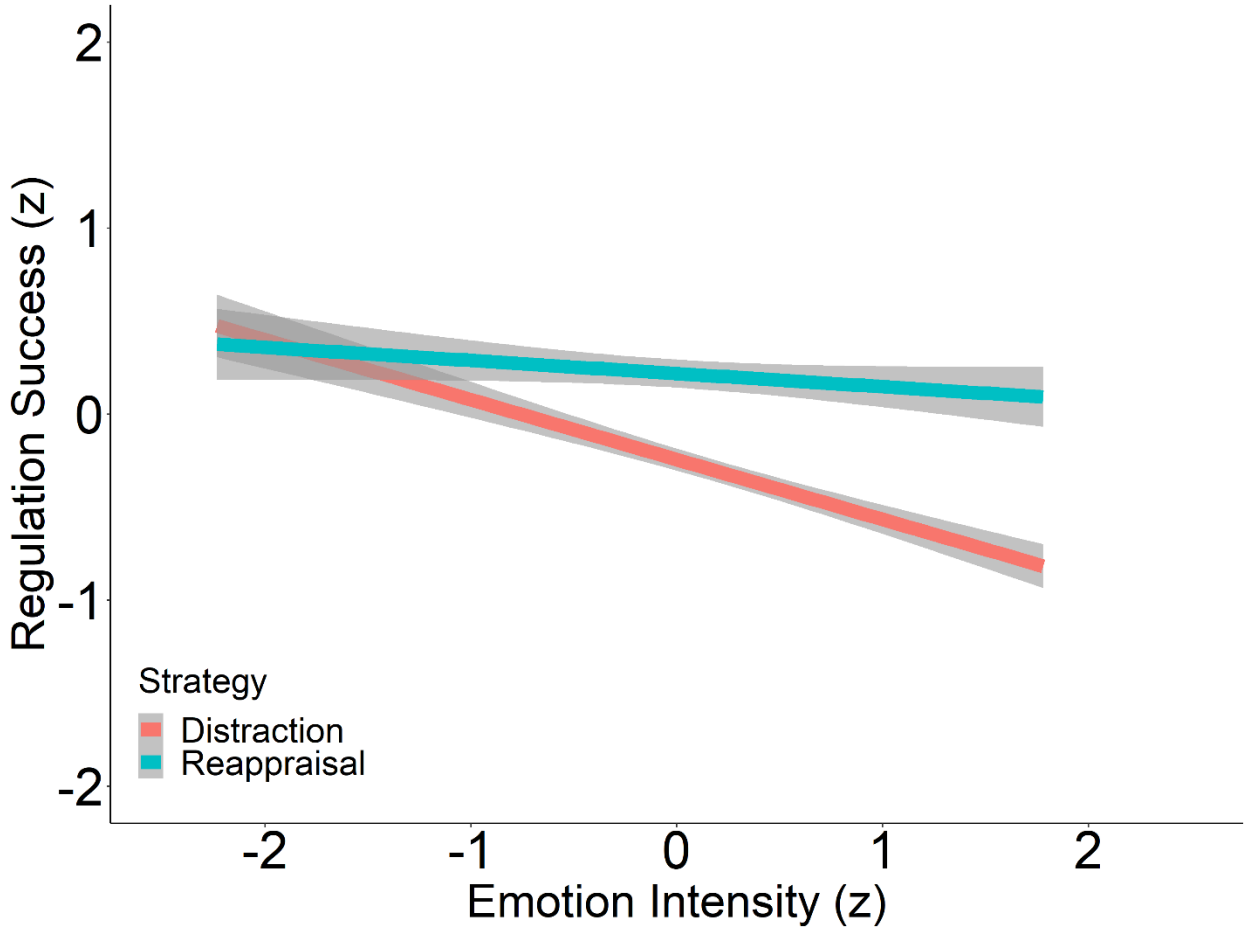


Fig 6. Strategy moderated the relationship between emotional intensity and regulatory success (β = 0.25, p = 0.003). While the success of reappraisal was relatively unrelated to emotional intensity, distraction demonstrated a negative association with emotional intensity, contrary to what extant literature 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.

384

EXPERIMENT 3 (1158 / 2000 Words): Experiments 1 and 2 found emotional intensity predicted regulation effort, but not strategy usage. We theorized that decontextualizing the stimuli and allowing participants to make a deliberative choice, rather than a reflexive response (as made in the haunted house), might reintroduce the association between ER strategy and emotional intensity observed in lab studies. We tested this in Experiment 3 by presenting the events recalled by experiencers (haunted house participants) to forecasters (people who had not been to the haunted house tasked with predicting how they would have responded to the events) (Wilson & Gilbert, 2003). Such a pattern of results would support the notion that associations between strategy choice and affective intensity, while strong in decontextualized situations, may be less predictive and more complex in naturalistic settings.

METHOD (516 Words):

PARTICIPANTS: In July 2021, 170 participants ($\bar{x}_{age}$ = 34.34 yrs, range = 18 -75 yrs, sd_{age} = 14.31 yrs, 100 female, 2 non-binary) consented to an IRB-approved online study described as measuring individual differences in choice predictions. Sample size was determined assuming a 10% attrition rate, an r-squared value of up to 0.10 for covariates, and a small effect size (OR = 1.68, 1-β = 0.80, α = 0.05, two-tailed). Participants were screened for English fluency, literacy difficulties, cognitive impairments, vision impairments and a United States based-location. 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.

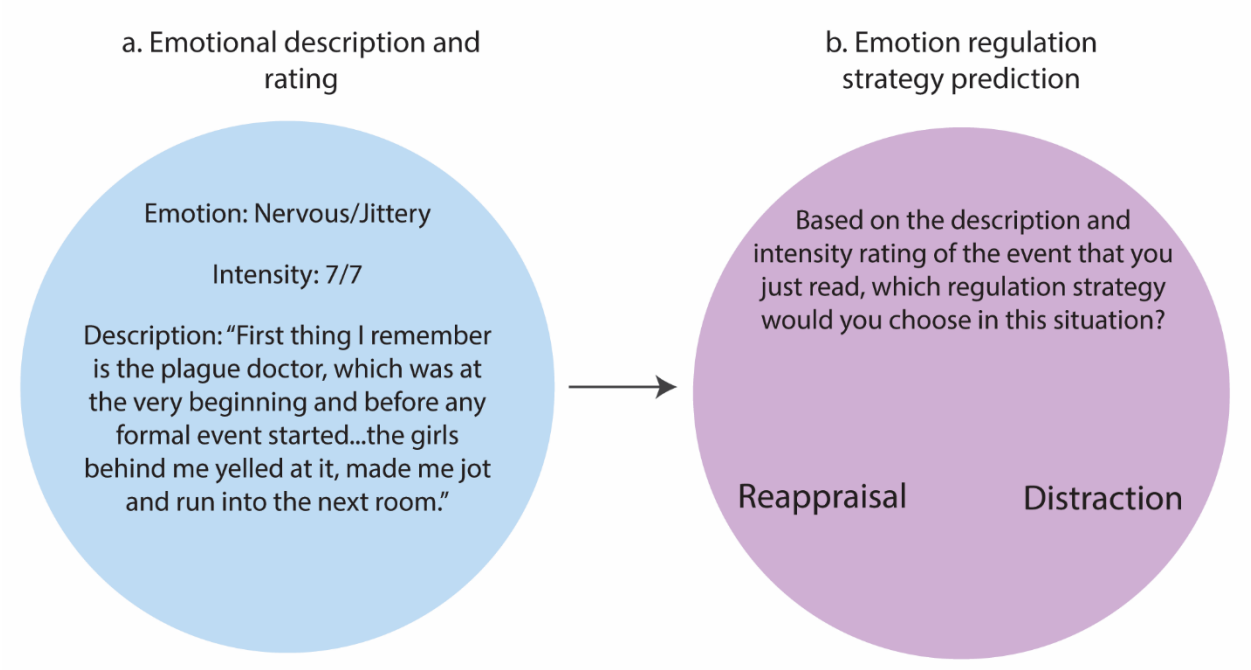
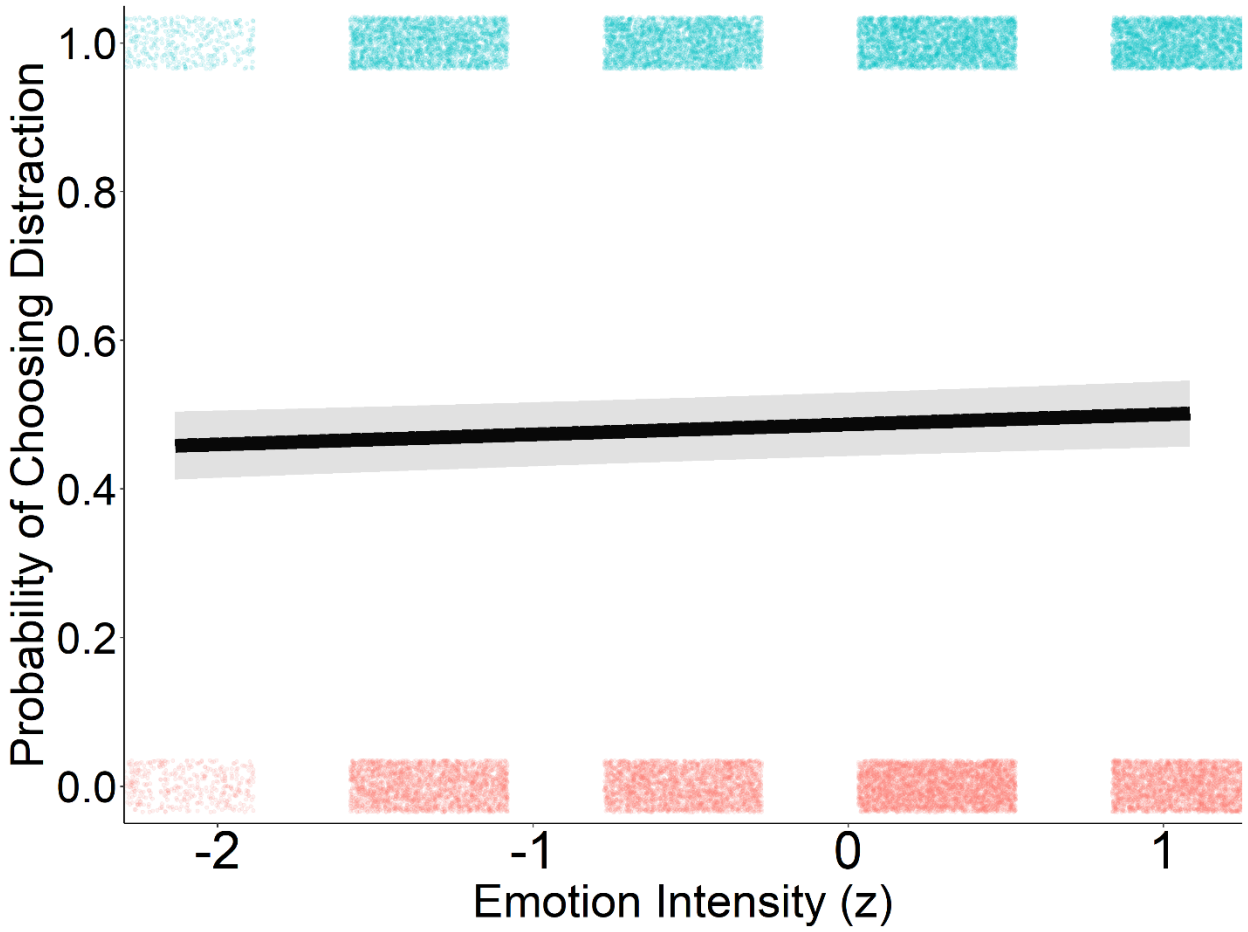


Fig 7. Study 3: Task Overview - One hundred and seventy (170) participants (forecasters) read the descriptions that Study 1 participants (experiencers) wrote about their emotional experience in the haunted house. a. Forecasters read the experiencers' emotional descriptions and intensity rating . b. Forecasters indicated what regulation strategy (distraction or reappraisal) they would use to regulate their emotions in the described event.

MATERIALS AND PROCEDURE: Details from the seventy-eight negatively-valenced Study 1 events regulated through either reappraisal or distraction were presented to participants who had not been to the haunted house. 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. Participants performed a brief practice task before the primary task began. All 78 events from the primary task were randomized and serially presented. For each event, the emotions experienced, the intensity of each emotion, how the experiencer 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. Following the primary task, participants completed individual difference measures. Study 3 was preregistered with AsPredicted (https://aspredicted.org/XXH_W1V)

RESULTS:

Intensity predicts regulatory strategy choice for forecasters, but not for experiencers. We had built multilevel binary logistic regression models specifying regulation strategy choice as our criterion variable and both the forecaster and experiencer as random intercepts. This was to account for both the multiple responses forecasters provided and the multiple events each forecaster provided as well, both likely contributing underlying random effects toward the data. Fixed effects models were built from a null model (ICC = 0.14). We found that our model using only emotional intensity to predict strategy choice performed better than our null model ($p = 0.004$) and comparable, more complex models via AIC comparison. In this model, we observed a small association between choice and emotional intensity (OR = 1.06, 95% CI = [1.02, 1.10], $p = 0.004$). Like Study 1, we also explored this relationship when emotional intensity was summed and averaged at the event level. We found that analyzing using event-average emotional intensity mirrored our primary analysis in model performance and criterion-predictor association (OR = 1.12, 95% CI = [1.06, 1.19], $p < 0.001$) as did our summed emotional intensity approach (OR = 1.03, 95% CI = [1.01, 1.05], $p = 0.032$). Though we remain agnostic as to which approach might be most appropriate, the pattern of significance across all three approaches is clear: the relationship between emotion intensity and strategy choice resembled that observed within emotion-regulation choice lab paradigms.



439 **Fig 8.** Emotional intensity did predict the strategies individuals chose once events were decontextualized (OR = 1.056, p = 0.004). 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.

GENERAL DISCUSSION (1957 / 2000 Words): Three experiments examined the association between emotional intensity and regulation strategy usage in a naturalistic context. Experiment 1 tasked untrained participants to recall emotional and spontaneous regulatory behaviors in a surprise recall task after exposure. As predicted, emotional intensity predicted regulation extent, but not strategy usage. Experiment 2 again failed to find an intensity-strategy use relationship with increased sample size and additional covariates. Experiment 3 demonstrated that decontextualizing the events reported in Study 1 introduced associations between affective intensity and strategy choice akin to lab studies, albeit, with a much smaller effect size. The present findings highlight challenges in translating emotion regulation theory to real-world application, as decontextualized high-intensity paradigms may not accurately reflect regulatory behaviors in everyday life.

A line of research that may shed light upon the discrepancy between theory and practice comes from cold-to-hot empathy gaps, or forecasting discrepancies between how people feel in relatively removed, decontextualized circumstances and how they think they would feel in more intense circumstances (Loewenstein, 1996). Individuals in "cold states" consistently under predict the challenges associated with meeting affectively-relevant goals during "hot states" (Fisher et al., 2014; Sayette et al., 2008; Van Boven et al., 2003). Such a pattern mirrors the differences observed between Experiments 1 & 2 and Experiment 3, wherein decontextualizing events (i.e., shifting them from a hot state to a cold state) yielded a predictable pattern in strategy choice not observed during hot state ER usage. This pattern might also be extended to ER choice paradigms and usage more generally, wherein the reliability of predicting ER choice is limited by how well the paradigm mirrors real-world circumstances. Though this study is the first to our knowledge that has utilized naturalistic paradigms to demonstrate this in the domain of ER, similar approaches have demonstrated similar discrepancies in moral domains (FeldmanHall et al., 2012).

We also observed in Study 2 that recalling previously reported emotion reduced intensity but differences were not observed between the emotional intensity of events reported immediately after exposure and one week after exposure. Such patterns might parallel similar trajectories outlined within the affective labeling literature, in which the act of semantically reviewing an emotional experience can itself be a form of ER and reduce the emotional intensity of an experience (Torre & Lieberman, 2018). The process of evaluating experiences within our surprise self-report task may itself be a form of regulation and introduce additional variability into the association between emotional intensity and ER strategy usage.

There are several limitations in our experimental approach that should be noted. First, our narrow aims resulted in excluding many observations, though the volume of observations not meeting lab-like standards further highlights the discrepancy between ER choice paradigms and actual ER usage. Additionally, haunted houses may not generalize to other high-intensity naturalistic settings, though, extant literature using similar settings suggest that this dynamic, multimodal approach may better model aversive stimuli than traditional paradigms (Clasen, Andersen, and Schjoedt, 2019 ; Tashjian et al., 2022). The use of a haunted house as our setting also may have resulted in self-selection biases. When asked about motivations for participating, thrill seeking ($x = 65.7$, 0 – 100 scale) was slightly above the average of all motivations ($x = 52.0$, $sd = 28.5$) and enjoyment of fear was just above the scale midpoint ($x = 3.88$, 0 – 6 scale). Lastly, the mixed results of time in Experiment 2 tempers the reliability of the findings of Experiment 1, as the extent

to which participants can reliably report the intensity of emotional experiences at a delay may be questionable. Future research should limit the delay between experience and report as much as possible without interfering with the emotional event. Despite 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.

Taken together, the present experiments represent what we believe to be the first attempt to test emotion regulation relationships established in the lab using a high-intensity naturalistic setting and untrained participants. This approach offers an alternative means of exploring emotion regulation usage in an ecologically-valid fashion. Our data and results may be of particular interest to other emotion, self-regulation, and cognitive control researchers interested in naturalistic design. In failing to replicate lab results with Studies 1 and 2 but finding a modest association in Study 3, we may have support for theories postulating a divergence in the mechanisms underlying ER choice and usage, though the limitations inherent to this study leave room for other possibilities. As such, if we aim to more accurately model human regulatory behavior, it may be necessary for researchers to capture phenomena beyond the labs; to observe the awards that left us unfulfilled, the feedback that we couldn't handle, and the Fielder-esque awkward moments that elicit strong emotional reactions.

Acknowledgments (1987 / 2000 Words):
We thank [REMOVED FOR BLINDING] 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 [REMOVED FOR BLINDING] and by the National Science Foundation (Grant No. [REMOVED FOR BLINDING]).

Open Practices:
Experiment 1 was not preregistered; the preregistration for Experiments 2 and 3 can be found at https://aspredicted.org/DP1_453 and https://aspredicted.org/XXH_W1V, respectively. Deidentified data, code, and questionnaires to replicate the findings have been made publicly available via OSF and can be accessed at https://osf.io/j5sku/?view_only=94cc65498cab4c09b7610462df82c662 [PLEASE NOTE: THIS LINK IS TO A BLINDED OSF REPOSITORY AND SHOULD BE REPLACED BEFORE PUBLISHING].

REFERENCES:

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. (2013). The Future of Emotion Regulation Research: Capturing Context. *Perspectives on Psychological Science*, *8*(2), 155–172. https://doi.org/10.1177/1745691612459518

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

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

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

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

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

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

Feldman, J. L., & Freitas, A. L. (2021). The generality of effects of emotional experience on emotion-regulation choice. *Emotion*, *21*(1), 211–219. psyh. https://doi.org/10.1037/emo0000611

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

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

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. pdh. 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

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. psyh. https://doi.org/10.1177/0956797616669086

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. psyh. 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

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

Loewenstein, G. (1996). Out of Control: Visceral Influences on Behavior. *Organizational Behavior and Human Decision Processes*, *65*(3), 272–292.

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

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. (2007). 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. psyh. https://doi.org/10.1016/j.paid.2015.06.048

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 and Emotion*, *33*(8), 1709–1717. psyh. https://doi.org/10.1080/02699931.2019.1592117

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

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

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., & Sheppes, G. (2018). When knowledge is (not) power- the influence of anticipatory information on subsequent emotion regulation: Neural and behavioral evidence. *Journal of Experimental Psychology: General*, *147*(8), 1225–1240. pdh. https://doi.org/10.1037/xge0000452

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

Sheppes, G. (2020). Transcending the "good & bad" and "here & now" in emotion regulation: Costs and benefits of strategies across regulatory stages. In *Advances in experimental social*

psychology (Vol. 61, pp. 185-236). Academic Press. https://doi.org/10.1016/bs.aesp.2019.09.003

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*. psyh. https://doi.org/10.3389/fpsyg.2014.00346

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

Szasz, P. L., Coman, M., 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

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

Torre, J. B., & Lieberman, M. D. (2018). Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation. *Emotion Review*, *10*(2), 116–124. https://doi.org/10.1177/1754073917742706

Wilson, T. D., & Gilbert, D. T. (2003). Affective Forecasting. In *Advances in Experimental Social Psychology* (Vol. 35, pp. 345–411). Academic Press. https://doi.org/10.1016/S0065-2601(03)01006-2

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

Young, G., & Suri, G. (2020). Emotion regulation choice: A broad examination of external factors. *Cognition and Emotion*, *34*(2), 242–261. psyh. 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