Supplementary Materials

PILOT STUDY METHODS

A pilot study 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 quasi-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. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

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, 18.51% Hispanic) were recruited from a large northeastern city via flyers for an IRB-approved fear and memory study. Participants were predominantly well-educated ($\bar{x}_{Years\ of\ Education}$ = 15.3 yrs, $sd_{Years\ of\ Education}$ = 2.4 yrs), with 92.6% reporting having completed some college (35.2%), a 4-year degree (29.6%), some post-graduate studies (16.7%), or a post-graduate degree (11.1%). Socioeconomic status was more normally distributed, with 12.7% of respondents reporting making less than $15,000 per year, 16.3% reporting between $15,001 and $25,000, 09.1% reporting $25,001 to $35,000, 23.6% reporting between $35,001 and $50,000, 12.7% reporting between $50,001 and $75,000, 16.4% reporting between $75,001 and $100,000, 09.1% reporting between $100,001 and $150,000, and no one reporting greater than $150,000. The racial identity of participants was not assessed.

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 (η_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* Stasiak et al., 2023). Participants were then escorted by two research assistants to the remotely-located haunted house which was the same as used within Experiment 1.

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 effort participants exerted trying 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 Experiment 1, we ask participants about their regulatory responses to each emotion they endorsed, which we refer to as capturing regulatory responses at the "emotion-level". For any one event, one or more emotions may have been endorsed by the participant.

Participants were not trained in emotion regulation strategies, nor were they primed to consider their emotion regulation strategies prior to these questionnaires. Though EMA approaches could also be applied in such a setting, the training and/or consideration of experience required would violate the immersive, naturalistic experience we aimed to model. Emotion categories were adapted from the Positive and Negative Affect Schedule (PANAS) (Watson et al., 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 Cronbac''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 effort 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.

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. All data and scripts used to produce this analysis are publicly available at OSF (*See* Open Practices). This study was not pre-registered.

PILOT STUDY RESULTS

As we primarily aimed to determine how analogous associations observed in ER strategy choice paradigms were to ER usage in quasi-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 and state anxiety predicts regulatory extent. Pro-hedonic trends in emotion regulation (i.e., minimizing aversive experiences) suggest that a positive linear relationship should exist between negative affective states and efforts to regulate those experiences. If we did not find associations between how intense an emotional experience was or how anxiety-provoking an experience was and how much effort someone devoted to regulating that event – regardless of *how* they regulate it – it may call into question whether our paradigm is well suited to elicit emotion regulation behaviors. For each event, we asked participants how intense the emotions that they endorse were as well as the extent to which they attempted to regulate that event (i.e., "To what extent did you attempt to change or regulate how you felt during this event?"), which may be comparable to effort. We also assessed state anxiety prior to entering the haunted house. To assess an association between event intensity and event regulation extent, 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 model as determined by AIC comparison found person-centered emotional intensity to be a significant predictor of regulation extent ($\beta = 0.31$, *95% CI* = [0.17, 0.46], $p < 0.001$). To assess an association between state anxiety and regulatory extent, we built a simple linear model regressing each participant's average self-reported regulatory extent across events onto State STAI scores and again found a significant positive association ($\beta = 0.34$, *95% CI* = [0.054, 0.626], $p = 0.021$). These results suggest our paradigm elicited regulatory behaviors from participants that follow a logical, predictable pattern.

Intensity did not predict regulatory strategy usage. Using multilevel binary logistic regression models specifying regulation strategy usage 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 usage 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 predicted strategy usage in that model (OR = 1.83, 95% CI = [0.65, 3.2], p = 0.079).

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 usage. 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 predicted strategy usage, 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.

STUDY 1 MULTIVERSE METHODS & RESULTS

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 choice 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). Expanding our groups yielded a subset of 360 observations in which a negative emotion was

downregulated by either disengagement or engagement strategy. These observations were reported by 89 participants. Of the total observations, 237 (65.80%) reported using distraction or suppression to regulate their emotions. The average emotional intensity of observations was 2.40. However, our best performing non-null model, including only intensity as a fixed effect ($p = 0.32$ when compared to null), again did not find an association between emotional intensity and strategy usage ($OR = 1.18$, 95% $CI = [0.85, 1.63]$, $p = 0.32$). Again, though not significant, this statistic suggests that every one standard deviation unit increase in emotional intensity increases the odds of choosing a disengagement strategy by approximately 18%.

Because it could be argued that a haunted house setting could elicit greater contra-hedonic regulation activity (i.e., downregulating positive emotion), we then constructed a series of additional models which were beyond the purview of our initial aims and hypotheses to determine whether any statistically significant relationship could be observed between affective intensity and regulation strategy usage in this context. We iteratively modified predictors and outcome variables across 14 additional models, including covariate models adjusting for sex, cognitive load, and ERQ reappraisal subscale scores, which despite failing to demonstrate significance in this dataset often predict regulation choice in lab settings. 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 ($p = 0.32$). This model included a random intercept for participant and a single predictor, affective intensity of positive and negative emotions, regressed upon distraction versus reappraisal strategy usage with data from unique events reported both immediately after and one-week after exposure ($OR = 1.42$, 95% $CI = [1.03, 1.98]$, $p = 0.04$). The model composition, comparison and results of all of these models can be found in **Table 1**.

Model Details

Outcome	Emotions Included	Data Collection Time	Covariates †	Participants (n)	Observations	Null ICC	Model Comparison (x^2)	Model Significance	Bonferroni Adjustment	Odds Ratio	Lower Bound ^	Upper Bound ^
Disengage. V. Engage.	Positive & Negative	Immediate & Delayed	Included	90	397	0.34	-	-	-	1.27	0.93	1.73
			Not Included				2.64	0.10	0.80	1.28	0.95	1.71
Disengage. V. Engage.	Negative	Immediate & Delayed	Included	89	360	0.36	-	-	-	1.17	0.83	1.64
			Not Included				0.98	0.32	1.00	1.18	0.85	1.63
Distract. V. Reappraisal	Positive & Negative	Immediate & Delayed	Included	78	328	0.37	-	-	-	1.45	1.03	2.05
			Not Included				4.45	0.04 *	0.32	1.42	1.03	1.98
Distract. V. Reappraisal	Negative	Immediate & Delayed	Included	77	298	0.40	-	-	-	1.38	0.95	1.99
			Not Included				2.70	0.10	0.80	1.36	0.95	1.95
Disengage. V. Engage.	Positive & Negative	Immediate	Included	79	213	0.42	-	-	-	1.02	0.65	1.58
			Not Included				0.09	0.76	1.00	1.07	0.71	1.61
Disengage. V. Engage.	Negative	Immediate	Included	77	194	0.39	-	-	-	1.05	0.65	1.68
			Not Included				0.14	0.71	1.00	1.09	0.70	1.68
Distract. V. Reappraisal	Positive & Negative	Immediate	Included	64	171	0.45	-	-	-	1.28	0.78	2.11
			Not Included				0.98	0.32	1.00	1.26	0.79	2.01
Distract. V. Reappraisal	Negative	Immediate	Included	63	155	4.01	-	-	-	1.33	0.77	2.28
			Not Included				1.17	0.28	1.00	1.30	0.81	2.09

* = p < 0.05   *† Covariates included Sex, Cognitive Load, and ERQ reappraisal subscale*   *^ Bounds represent 95% confidence intervals*

Table 1. *Results of Study 1 Exploratory Multiverse Models* – Hierarchical binary logistic regression models were constructed to explore how predictive affective intensity is of strategy usage and compared against null models allowing each participants' intercept to vary randomly. The outcome variable used, data inclusion criteria, and variables included are listed under the first four columns. Data size and model comparison results are listed under the subsequent five columns. Model comparisons against the null for not conducted for covariate models because comparing models with such a disproportionate number of terms can be misleading and is not statistically recommended. The odds ratio of affective intensity within each model is listed in the latter three columns. Only a single model performed better than its null and found affective intensity to predict regulation strategy usage, but did not survive family-wise adjustment.