Emotional Granularity is Associated with Daily Experiential Diversity

Katie Hoemann[1] · Yeasle Lee[1] · Peter Kuppens[1] · Maria Gendron[2] · Ryan L. Boyd[3]

Abstract
Emotional granularity is the ability to create differentiated and nuanced emotional experiences and is associated with positive health outcomes. Individual differences in granularity are hypothesized to reflect differences in emotion concepts, which are informed by prior experience and impact current and future experience. Greater variation in experience, then, should be related to the rich and diverse emotion concepts that support higher granularity. Using natural language processing methods, we analyzed descriptions of everyday events to estimate the diversity of contexts and activities encountered by participants. Across three studies varying in language (English, Dutch) and modality (written, spoken), we found that participants who referred to a more varied and balanced set of contexts and activities reported more differentiated and nuanced negative emotions. Experiential diversity was not consistently associated with granularity for positive emotions. We discuss the contents of daily life as a potential source and outcome of individual differences in emotion.

Keywords Emotion differentiation · Concepts · Natural language · Experience sampling · Meaning extraction

Emotional granularity – also known as "emotion differentiation" – is the ability to create differentiated and nuanced emotional experiences (Barrett et al., 2001; Tugade et al., 2004). A growing literature attests to a link between higher emotional granularity and positive outcomes (for recent reviews, see O'Toole et al., 2020; Seah & Coifman, 2021; Thompson et al., 2021). Benefits include more successful emotion regulation (e.g., Kalokerinos et al., 2019) and healthier coping (e.g., Kashdan et al., 2010), as well as fewer symptoms of anxiety and depression (e.g., Seah et al., 2020; Willroth et al., 2019). Motivated by these findings, studies have begun to look at how granularity develops (e.g., Nook et al., 2017) and how it may manifest in a larger set of biological and interpersonal processes (e.g., Hoemann, Khan et al., 2021; Israelashvili et al., 2019). In this paper, we complement these efforts by examining the relationship

between emotional granularity and the diversity of everyday experiences.

Among other constructs for individual differences in emotional experience (e.g., emotional awareness), emotional granularity foregrounds the need for context-specificity (Hoemann, Barrett et al., 2021; Hoemann, Nielson et al., 2021). Granular emotional experiences are tailored to current or anticipated circumstances, facilitating adaptive patterns of behavior (e.g., O'Toole et al., 2020). As an example, imagine dealing with a daily stressor, such as accidentally breaking a friend's phone. Experiencing a finer-grained (i.e., more granular) emotion – 'guilty for one's clumsiness' – as opposed to a non-specific emotion – 'generally upset' – makes it more likely that feelings and actions are appropriate to the situation (e.g., making restitution rather than simply feeling bad). Consistently experiencing granular emotions means that feelings and actions are flexible to ever-changing circumstances.

Experiencing granular emotions is not the same as using specific emotion words. Labeling is not related to other measures of emotional granularity (Ottenstein & Lischetzke, 2019; Williams & Uliaszek, 2021) and can even interfere with the effective regulation that granularity is thought to facilitate (Nook et al., 2021; Vine et al., 2019). Instead, a constructionist account of emotion (e.g., Barrett, 2006, 2017) proposes that granular emotional experiences are a

---

✉ Katie Hoemann
khoemann@gmail.com

1 Department of Psychology, KU Leuven, Tiensestraat 102, Box 3727, 3000 Leuven, BE, Belgium

2 Department of Psychology, Yale University, New Haven, CT, USA

3 The Obelus Institute, Washington, DC, USA

product of how the brain uses emotion concepts. On this account, all concepts are understood as dynamic collections of prior experience that the brain uses to make meaning of the current context and issue predictions about what is likely to occur next. Predictions that are more context-specific are more efficient because they better anticipate probable actions and upcoming energy needs. When context-specific predictions are constructed using concepts for emotion, the result is higher emotional granularity.

Because concepts are dynamic, they are developed and constantly updated through engagement with the world. The brain accrues knowledge about appropriate actions and expected outcomes through interactions with other people, places, and things (Meteyard et al., 2012). From this perspective, greater variation in experience could result in more elaborated (i.e., rich, diverse) emotion concepts: exposure to a broader range of sensorimotor inputs and social dynamics – that is, the contexts and activities encountered over time – might allow the brain to construct predictions that are better tailored to the situation at hand, increasing emotional granularity. At the same time, a richer and more diverse set of emotion concepts could contribute to greater variation in the types and frequencies of everyday experiences (Lee et al., 2021). Concepts are known to shape how events are perceived, and this process may extend to the experience of emotion (e.g., Richmond & Zacks, 2017). Finer-grained emotion concepts may correspond with more specific patterns of attention to, and discrimination between, relevant situational features. From both perspectives, a relationship between emotional granularity and the contents of daily life can be expected.

To test this hypothesis, data are needed that capture not only how people experience their emotions across various situations, but also something about the nature of those situations. Although it is possible to assess the contents of daily life using self-reported location, activity, or social context, such data are limited in the insights they can provide into how people represent their experience. For example, someone could be at work but thinking about their plans for the weekend, reading but focusing on their grumbling stomach, with family but feeling alone. It is not feasible to anticipate beforehand all the aspects of experience people may attend to and in what combination. In contrast, natural language data provide a window onto how people see themselves and the world around them without constraining or priming attention to one aspect of experience or another.

Prior research has estimated the meaningful content of natural language data using both closed- and open-vocabulary approaches. Closed-vocabulary approaches, including the software Linguistic Inquiry and Word Count (LIWC; e.g., Pennebaker et al., 2015), use predefined word lists (i.e., dictionaries) to score texts for the presence of researcher-developed or -selected topics. By comparison,

open-vocabulary approaches produce clusters of semantically related words, or themes, in a data-driven manner. While closed-vocabulary approaches are useful for confirming and comparing themes, open-vocabulary approaches are critical for discovering new ones. Studies using an open-vocabulary approach have found themes to predict between-person differences in life satisfaction (Schwartz et al., 2016) as well as within-person fluctuations in self-reported emotion (Sun et al., 2019). Open-vocabulary approaches have also been used to describe self-schemas (Rodríguez-Arauz et al., 2017) and mental health concerns across cultures (Ramirez-Esparza et al., 2008).

The present paper builds on this prior work by employing an open-vocabulary approach to capturing the contents of daily life. In Study 1, we extracted themes from descriptions of everyday events, interpreting these themes as the contexts and activities experienced by participants in our sample. We then used the distribution of themes within participants as an estimate of experiential diversity and assessed the relationship of this measure with emotional granularity. In Studies 2 and 3, we examined the robustness of this relationship within additional data sets. In all three studies, we used repeated emotion intensity ratings to compute separate estimates of emotional granularity for negative versus positive emotions (O'Toole et al., 2020; Thompson et al., 2021). We predicted that experiential diversity would be associated with both estimates of granularity, such that participants who referred to a more varied and balanced set of contexts and activities in their daily lives would report more differentiated and nuanced emotions, regardless of valence.

Methods Overview

We tested the hypothesized relationship between experiential diversity and emotional granularity through a secondary analysis of three experience sampling data sets from our own labs. Each of these data sets included self-reported emotions as well as natural language descriptions of everyday events, allowing us to examine how participants conceptualized their experiences in typical settings (following e.g., Sun et al., 2019) and estimate the diversity of contexts and activities they encountered based on patterns of overall word use. In Study 1, we analyzed elements of the data set previously reported in Hoemann, Khan et al. (2020, see also Hoemann, Barrett et al., 2021; Hoemann, Khan et al., 2021). For each experience sampling event, English-speaking participants rated the intensity of their experience on a set of emotion adjectives and wrote about what was happening at the time they received the prompt. In Study 2 (Hoemann, Fan et al., 2020), English-speaking participants rated the intensity of their experience at each prompt, but only wrote about a small subset. In Study 3 (Carlier et al., 2021), Dutch-speaking

participants rated the intensity of their experience at each prompt and could optionally record a verbal description. To quantify results across studies, we performed both meta- and integrative data analyses of our findings.

Study 1

Method

Study 1 was approved by the Northeastern University Institutional Review Board (IRB# 16-01-13) and has been reported in detail in Hoemann, Khan et al. (2020, see also Hoemann, Barrett et al., 2021; Hoemann, Khan et al., 2021). Below we report aspects relevant for the present analyses.

Participants An initial 67 adults were recruited from Northeastern University classrooms and online portals as well as the greater Boston area through posted advertisements; all eligible participants were fluent English-speakers. Informed consent was obtained from all participants before beginning the study. Participants received $490 for completing all parts of the study, plus up to $55 in compliance and task incentives. Six participants withdrew, nine were dismissed due to poor compliance, and two were excluded from data analysis because they did not complete the full study protocol, for a final sample size of 50 (54% female; 40% White, 2% Black, 44% Asian, 14% other; $M = 22.5$ years, $SD = 4.4$ years). We conducted a sensitivity analysis for this data set in G*Power (Faul et al., 2009), assuming $\alpha < 0.05$, two-tailed and power $(1\text{-}\beta) > .80$ in linear multiple regression analyses with one tested and two control predictors. For comparison with our reported results, we converted the resulting effect size estimate (given as Cohen's f^2) to a standardized regression coefficient (β) using the formula given in Eq. 1:

$$\beta = \sqrt{f^2 / \left(1 + f^2\right)} \tag{1}$$

The sensitivity analysis indicated that this data set was adequately powered to detect moderate size effects, $f^2 \geq 0.16$, $\beta \geq 0.37$.

Procedure Participants completed approximately 14 days ($M = 14.4$, $SD = 0.6$) of experience sampling including peripheral physiological monitoring and end-of-day diaries. Only the end-of-day diary data are used in the present analysis. Participants also attended two in-lab sessions that are not reported here.

Each day of experience sampling lasted for 8 h and began when participants were outfitted with physiological sensors and a smartphone with an associated smartphone application. Most experience sampling prompts were physiologically triggered to enable more efficient sampling of

psychologically salient moments. These prompts occurred any time there was a substantial, sustained change in cardiac activity in the absence of movement. Participants also received two 'random' prompts each day that were not contingent on changes in cardiac activity. Altogether, participants responded to an average of 8.65 prompts ($SD = 1.09$) per day.

At each sampling prompt, participants responded to a series of questions presented on the smartphone app. These data were not analyzed in the present paper; however, several elements were presented to participants in the end-of-day diary and so are briefly reviewed. First, participants provided a brief free-text description of what was going on at the time they received the prompt. Participants also provided a brief free-text description of their social context (by writing "alone", listing interaction partners' initials, or writing "group") and selected their main activity from a drop-down list (e.g., "socializing", "eating", "working").

Immediately upon finishing each day of experience sampling, participants automatically received an online end-of-day diary. In this diary, they were presented with the event time, brief description, social context, and main activity for each completed sampling prompt, which served as a guide for participants to provide additional details about each event. Participants described in writing what was happening when they received each prompt. For every event, participants rated the intensity of their emotional experience on a set of 18 emotion adjectives using Likert-style scales from 0 ("not at all") to 6 ("very much"). The full list of emotion adjectives is reported below. These intensity ratings were requested in the end-of-day diary, rather than at each experience sampling prompt, to reduce participant burden in the moment.

Data Preparation

Experiential Diversity Estimates of experiential diversity were computed from participants' event descriptions following the Meaning Extraction Method (MEM) topic modeling approach (e.g., Chung & Pennebaker, 2008). The initial set of event descriptions contained 6,399 entries. To prepare the data for analysis, we removed any empty or duplicate entries and entries from participants excluded from the final sample (8.86% of descriptions, new total 5,832; $M = 117$, $SD = 23$, min $= 61$, max $= 172$ per participant). We also removed short entries, as these texts did not contain enough content words to contribute to stable themes and so can be considered partial or missing data. A 25–50-word minimum word count is recommended for texts submitted to these types of automated analyses, especially when the total number of texts is relatively small (as ours is; for discussion, see Boyd, 2017). Applying a 25-word minimum resulted in the

removal of an additional 32.10% of descriptions, for a new total of 3,960 event descriptions ($M = 79$, $SD = 31$, min $= 38$, max $= 164$ per participant), ranging in length between 25 and 509 words ($M = 131$, $SD = 86$). The percentage of data removed varied by participant around an average of 31.22% ($SD = 23.85\%$, min $= 0\%$, max $= 65.32\%$). The design of Study 1 provides some insight into why there are so many short texts: participants were required to write something about every event they documented during the day ($\sim$ 8–9) across 14 days of experience sampling. To balance the need for substantive texts against the need to maximize data usage, we also implemented analyses with 20- and 30-word minimum text lengths (for descriptive statistics, see supplemental Table S1). Of note, a typical English sentence is 15–20 words long (Cutts, 2009), so these minimum lengths correspond to texts of approximately 1 or 2 sentences.

The first step in performing the MEM was identifying the most frequently used content words and their locations in the texts, which we completed using the Meaning Extraction Helper (MEH) software (Boyd, 2018). The MEH prepares texts for analysis by converting words to their root form (e.g., "ran" becomes "run"; "friends" becomes "friend") and by removing common closed-class or 'stop' words (e.g., articles, auxiliary verbs, prepositions, pronouns). The MEH allows users to add data-specific conversions and stop words. In the present analysis, we removed all content words that appeared in the end-of-day diary instructions, as participants often echoed these instructions in their writing (e.g., "at the time I received the prompt…"). We also supplemented the conversion list with abbreviations commonly used by our sample (e.g., "prof" for "professor"). The MEH counts the content words in the prepared texts and produces co-occurrence matrices in which each text is scored for the most frequent words. As a robustness check to assess the impact of this parameter on our results, we generated separate co-occurrence matrices for the 100, 150, and 200 most frequent single words (i.e., unigrams). These values correspond roughly to average text length ± 50. Retaining more words would have gone beyond the co-occurrences available in a typical text; retaining fewer would have provided insufficient data for theme extraction. The number of retained words also corresponds with that used in prior analyses of texts of comparable length (e.g., Rodríguez-Arauz et al., 2017). We computed two types of co-occurrence matrices: a 'relative frequency' matrix, in which the raw count of each word was presented as a proportion of each text's length, and a binary (i.e., 'one-hot' encoded) matrix in which each word was coded as present (1) or absent (0) in each text.

The next step in performing the MEM was extracting themes, or words that consistently grouped together across all texts. We conducted a separate principal components analysis (PCA) with varimax rotation on each set of unigrams (100, 150, and 200) using the binary co-occurrence matrices.

Diagnostic tests indicated that the model was appropriate for the data (see supplemental Table S2). Inspection of the scree plot for the PCA on each set of unigrams suggested approximately 10 components at the 'elbow' with eigenvalues above 1 (Cattell, 1966). The scree test is an established method of selecting the number of themes to extract in the MEM (Rodríguez-Arauz et al., 2017), where the goal is to retain the smallest number that are interpretable or internally coherent (Boyd, 2017). We extracted the elbow ± 5 (i.e., 5, 10, and 15 components) so that we could also assess the impact of this parameter on our results. PCAs were performed in R (R Core Team, 2022) using the psych (Revelle, 2020) and factoextra (Kassambara & Mundt, 2020) packages.

We repeated these steps for each combination of the above-mentioned parameters (i.e., specification), producing a final set of 27 (3 minimum text length $\times$ 3 number of unigrams $\times$ 3 number of components) matrices. For each parameter, the middle value is the most appropriate for the data, such that the combination of ≥ 25-word texts, 150 unigrams, and 10 components is the a priori ideal. Step-by-step instructions are provided via our OSF repository (https://osf.io/gn8ca/).

We performed a cursory review of each set of extracted themes by comparing them to a randomly sampled subset of the event descriptions. Although the exact number and identity of words within each theme varied according to parameter settings, some clear trends emerged. For example, prevalent themes often described being in class (e.g., "professor", "lecture", "note[s]"), mealtimes (e.g., "eat", "food", "hungry"), and socializing (e.g., "weekend", "friend", "happy"). These themes matched what we read in the event descriptions and had face validity given that Study 1 participants were predominantly university students and early career researchers. An example set of themes is presented in Table 1.

To see how the diversity of themes varied by participant, we first examined the extent to which themes were present by scoring the texts using the rotated component matrix from each PCA and the corresponding relative frequency co-occurrence matrix generated by the MEH. We then classified each event description as either containing (1) or not containing (0) each theme by submitting the scored texts for all participants to an At Most One Change (AMOC) changepoints analysis (Killick & Eckley, 2014). This analysis ensured that only texts with scores above threshold were considered to meaningfully evidence each theme (following Entwistle et al., 2021). Based on the themes in Table 1, for example, if a given text included the words "lecture" and "note" but did not include the words "weekend" and "together", then it would receive a 1 for the 'in class' theme, and a 0 for the 'socializing' theme. We then computed a Gini coefficient over each participant's scored texts. This coefficient captures how broadly and evenly a phenomenon is observed across various types or categories; here, it captured

Table 1 Example themes from Study 1 event descriptions

	Description	Words (Loadings)
1	'exam'	exam (.55), study (.41), grade (.40), material (.39), class (.38), nervous (.37), test (.37), quiz (.37), prepare (.33), stress (.33), final (.32), anxious (.29), professor (.28), question (.26), note (.26), start (.20)
2	'in class'	attention (.68), pay (.66), professor (.54), lecture (.54), class (.46), listen (.30), material (.29), note (.28), interest (.24), topic (.23), bore (.23), learn (.21)
3	'mealtimes'	eat (.69), hungry (.62), food (.58), lunch (.48), breakfast (.46), finish (.25), hallway (.22), enjoy (.21), morning (.20)
4	'socializing'	friend (.49), plan (.39), talk (.39), meet (.34), happy (.34), excite (.34), together (.34), discuss (.32), semester (.31), week-end (.27), conversation (.25), year (.25), group (.25), told (.22), week (.21), nice (.21)
5	'at work'	train (.35), answer (.33), office (.33), lab (.33), annoy (.29), day (.27), wait (.27), minute (.27), today (.26), morning (.26), early (.26), email (.25), walk (.25), late (.24), question (.23), told (.23), work (.22), stop (.22), move (.21), left (.21), meet (.21), frustrate (.20), leave (.20)
6	'schedule'	sleep (.55), tired (.47), night (.42), nap (.42), bed (.37), morning (.29), wear (.28), early (.25), late (.25), rest (.24), day (.21)
7	'coursework'	complete (.47), alone (.46), library (.46), mind (.44), assignment (.39), work (.37), project (.37), task (.34), emotion (.30), due (.24), focus (.23)
8	'at home'	apartment (.38), live (.33), roommate (.33), read (.31), thought (.27), phone (.25), move (.23), stop (.22), clean (.22), talk (.21), told (.21), call (.21), room (.21), project (-.21)
9	'downtime'	watch (.57), play (.54), game (.51), video (.47), relax (.39), show (.33), finish (.24), enjoy (.23), music (.23), happy (.22)
10	'relationships'	people (.39), understand (.30), expect (.27), homework (.27), frustrate (.26), sad (.25), study (.24), talk (.23), annoy (.23), problem (.23), learn (.22), question (.21), life (.21), attention (-.24), pay (-.26)

Note: Components (10) extracted from a binary co-occurrence matrix of the 150 most frequent words (i.e., unigrams) in Study 1 texts with a minimum word count of 25. Words with |factor loadings| <.20 are omitted for ease of presentation

the relative spread of themes across an individual's event descriptions. We used the formula from Benson et al. (2018), given in Eq. 2,

$$G = 1 - \left(\left(\frac{2 \sum_{j=1}^{m} j c_{ij}}{m \sum_{j=1}^{m} c_{ij}} \right) - \frac{m+1}{m} \right) \quad (2)$$

where c_{ij} is the count of individual i's experiences within $j = 1$ to m categories (i.e., themes) indexed in a non-decreasing order ($c_{ij} \leq c_{ij+1}$). The main term in Eq. 2 calculates the weighted sum of the frequencies of a set of themes, divided by the product of the total frequency of all themes and the total number of themes. This is then subtracted from 1 to provide an estimate that scales intuitively from 0 (no diversity) to 1 (high diversity). We calculated a Gini coefficient for each combination of parameters (i.e., specification), producing a set of 27 (3 minimum text length × 3 number of unigrams × 3 number of components) estimates of experiential diversity. Text scoring and changepoints analyses were performed in R using the psych, GPArotation (Bernaards & Jennrich, 2005), MASS (Venables & Ripley, 2002), and changepoint (Killick & Eckley, 2014) packages. Gini coefficients were calculated using custom functions in MATLAB (MATLAB, 2018).

Emotional Granularity Estimates of emotional granularity were computed from the intensity ratings for the 18 emotion adjectives sampled in the end-of-day diary. The final sample of 50 participants completed a total of 721 sampling days, and 6,307 prompts, across the study. End-of-day diaries were

missing for 3 days (0.42% of diaries), affecting 3 participants (one diary each) and resulting in the loss of emotion intensity ratings for 19 prompts (0.30% of prompts). An additional 22 end-of-day diaries were completed late (i.e., the following day; 3.05% of diaries), affecting 17 participants (12 with one late entry; 5 with two) and 210 prompts (3.33% of prompts) but with no data loss. These high completion rates can be attributed to the check-ins participants had with experimenters at the start of each experience sampling day; continued compliance issues with end-of-day diaries were considered grounds for dismissal from the study.

Following recent literature (e.g., Kalokerinos et al., 2019), we used an intraclass correlation (ICC) for consistency with averaged raters (i.e., 'C-k' method).[1] Higher ICC values reflected lower emotional granularity (i.e., greater shared variance among adjectives' ratings). Negative values would have been recoded as 0 because they are outside the theoretical range for an ICC; however, none were observed in this data set. We computed separate estimates of granularity for 8 negative ("afraid", "angry", "bored", "disgusted", "embarrassed", "frustrated", "sad", "worn out") and 10 positive emotions ("amused", "calm", "excited", "grateful", "happy", "neutral",

---

[1] Emotional granularity can also be estimated using an intraclass correlation (ICC) for agreement with averaged raters (i.e., 'A-k' method). There is no apparent consensus in the literature as to which approach is preferred (Thompson et al., 2021) and in general consistency and agreement estimates are highly correlated (rs=.95-.99; Erbas et al., 2014). In the present studies, changing the ICC type did not affect the overall pattern of results.

"proud", "relieved", "serene", "surprised"), with these assignments based a median split of normative ratings (Warriner et al., 2013). ICCs were Fisher *r*-to-*z* transformed to fit the variable to a normal probability distribution. These transformed values were multiplied by -1 to yield estimates of granularity that scaled intuitively, such that lower (more negative) values reflected lower granularity, and higher (less negative) values reflected higher granularity. Granularity was calculated using the ICC (Salarian, 2016) and custom functions in MATLAB.

Analysis To test our hypothesis that experiential diversity would be positively associated with emotional granularity, we examined the relationship between participants' Gini coefficients and their inverse ICCs for negative and positive emotions, respectively, in the context of two control variables: mean affect and number of event descriptions. It could be that differences in experiential diversity were due in part to differences in the number of event descriptions included for each participant (either due to the number of prompts completed during experience sampling or to data exclusion). It could also be that differences in emotional granularity were more closely related to differences in overall mood rather than experiential diversity (e.g., Dejonckheere et al., 2019). To control for these possibilities, we conducted separate multiple regressions predicting negative and positive emotional granularity from experiential diversity, number of event descriptions, and mean (negative or positive) affect. Estimates of mean affect were derived by calculating the average intensity of negative or positive emotions across all end-of-day diary entries. We used a two-tailed test of significance at $\alpha = .05$ and considered observations with Cook's distance (D) values $> 3*M_D$ to be multivariate outliers. These outliers are identified in the figures below. We report results with these points removed. Results are reported using standardized coefficients for comparability across studies. Regression analyses were performed in MATLAB using fitlm (Statistics and Machine Learning Toolbox, 2019). All analytic code is available via our OSF repository.

We conducted 27 sets of analyses, one for each specification or combination of minimum text length (20, 25, and 30 words), number of unigrams (100, 150, 200), and number of components (5, 10, 15) extracted using the MEM. This approach follows from multiverse analyses in which model outcomes are explored across a set of reasonable parameter values (Steegen et al., 2016; see also Orben & Przybylski, 2019; Simonsohn et al., 2020). We summarize the results of our multiverse analysis descriptively and visualize them using specification curves.

Results

We first examined the relationship between the 27 estimates of experiential diversity and negative emotional granularity,

controlling for the number of event descriptions and mean negative affect. As can be seen in Fig. 1, the effect of experiential diversity ranged between $\beta = .17$ and $\beta = .45$, with a median of $\beta = .34$, 95% CI [0.09, 0.59], $t(44) = 2.69$, $p = .01$, two-tailed, 2 outliers excluded. In the 'middle' specification (≥ 25-word texts, 150 unigrams, 10 components), our a priori ideal, the effect of experiential diversity was $\beta = .44$, 95% CI [0.19, 0.69], $t(43) = 3.55$, $p = .001$, two-tailed, 3 outliers excluded. The effect of experiential diversity was not significant in two of the 27 regression models, and in two models it did not reach a conventional level of significance ($.06 \leq p \leq .08$, two-tailed). See supplemental Table S3 for the full results of all 27 models. Scatter plots of the relationship between experiential diversity and negative emotional granularity (residuals) are presented in supplemental Figures S1a-c. Overall, we interpret these results as consistent with our prediction: Experiential diversity was positively associated with negative emotional granularity, such that participants who referred to a more varied and balanced set of contexts and activities in their daily lives also reported more differentiated and nuanced negative emotions.
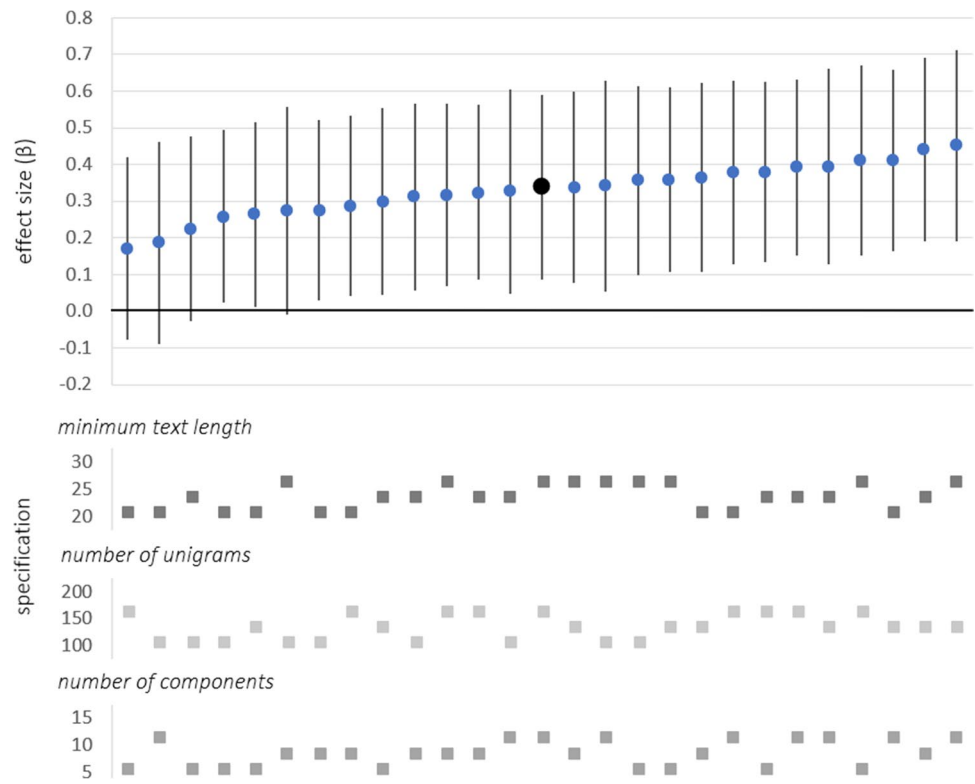
We next examined the relationship between the 27 estimates of experiential diversity and positive emotional granularity, controlling for the number of event descriptions and mean positive affect. As can be seen in Fig. 2, the effect of experiential diversity ranged between $\beta = 0.16$ and $\beta = .42$, with a median of $\beta = .28$, 95% CI [0.05, 0.51], $t(42) = 2.42$, $p = .02$, two-tailed, 4 outliers excluded. The effect of experiential diversity was also significant in our a priori ideal model (≥ 25-word texts, 150 unigrams, 10 components): $\beta = .42$, 95% CI [0.15, 0.69], $t(42) = 3.16$, $p = .003$, two-tailed, 4 outliers excluded. The effect of experiential diversity was not significant in four of the 27 regression models, and in three models it did not reach a conventional level of significance ($.06 \leq p \leq .07$, two-tailed). Full regression results are provided in supplemental Table S4, with scatter plots in supplemental Figures S2a-c. These results are again consistent with our prediction: Experiential diversity was also positively associated with positive emotional granularity.

To further illustrate our findings, Fig. 3 provides distributions of themes and emotion rating correlation matrices for example participants with high (left panel) and low (right panel) experiential diversity and emotional granularity.

Discussion

In Study 1, English-speaking participants provided written descriptions of and emotion intensity ratings for everyday events via end-of-day diaries. Secondary analysis of this data set revealed that experiential diversity was positively associated with both negative and positive emotional granularity. To examine whether these effects were specific to this data set and the methods used to collect it, in Study 2 we sought

Fig. 1 Specification curve for the negative emotional granularity multiverse analysis in Study 1. The upper panel depicts the estimated effect size, β, and its 95% CI for experiential diversity when predicting negative emotional granularity, controlling for number of event descriptions and mean negative affect, with outliers removed. Effect size estimates are arranged in ascending order with the median value identified in bold black. The lower panel shows the specification corresponding to each effect size, with the parameter values for minimum text length, number of unigrams, and number of components implemented in the analysis identified in gray squares

Fig. 2 Specification curve for the positive emotional granularity multiverse analysis in Study 1. The upper panel depicts the estimated effect size, β, and its 95% CI for experiential diversity when predicting positive emotional granularity, controlling for number of event descriptions and mean positive affect, with outliers removed. Effect size estimates are arranged in ascending order with the median value identified in bold black. The lower panel shows the specification corresponding to each effect size, with the parameter values for minimum text length, number of unigrams, and number of components implemented in the analysis indicated in gray squares
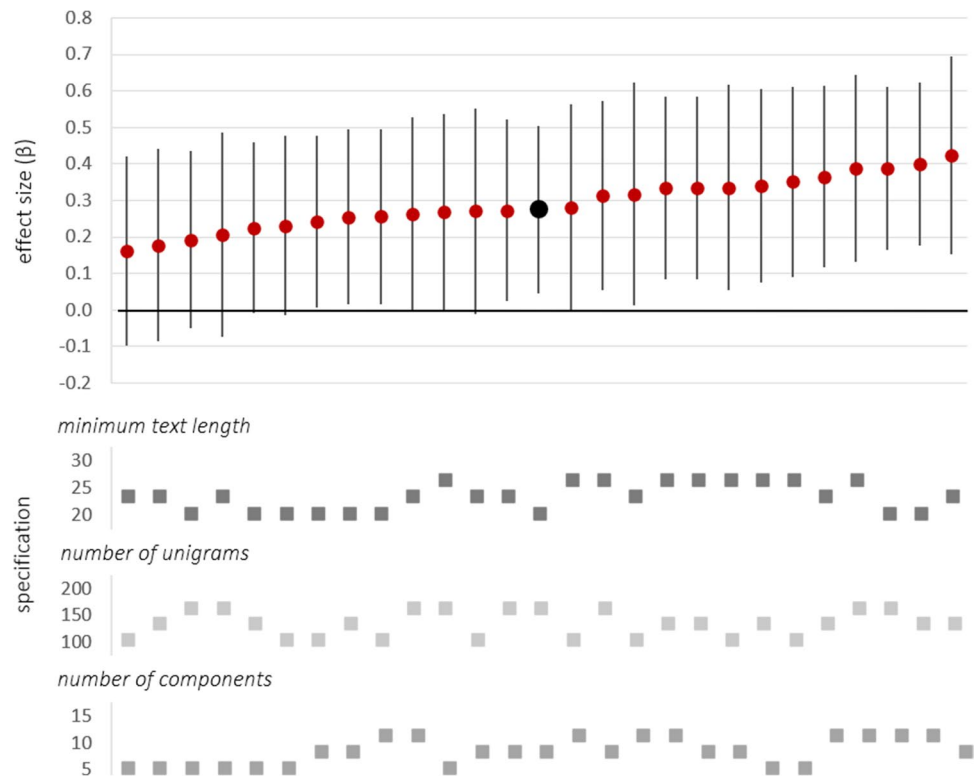
Fig. 3 Example participants with high (left panel) and low (right panel) experiential diversity and emotional granularity in Study 1. The upper part of each panel shows a proportional count of how many times each theme appeared in those texts, with these counts based on one of 27 possible specifications (i.e., $\geq$ 25-word texts, 150 unigrams, 10 components). Higher diversity is indicated by a broader and more even distribution across a given set of themes (i.e., more bars of roughly similar height). The lower part of each panel shows a heatmap of pairwise correlations between intensity ratings for the sampled emotions. Lighter values indicate negative or weaker correlations; darker values indicate stronger positive correlations. Emotions are arranged in alphabetical order by negative valence (top/left) and then positive valence (bottom/right). Higher granularity is indicated by fewer strong intra-correlations (i.e., fewer darker boxes forming around the diagonal)
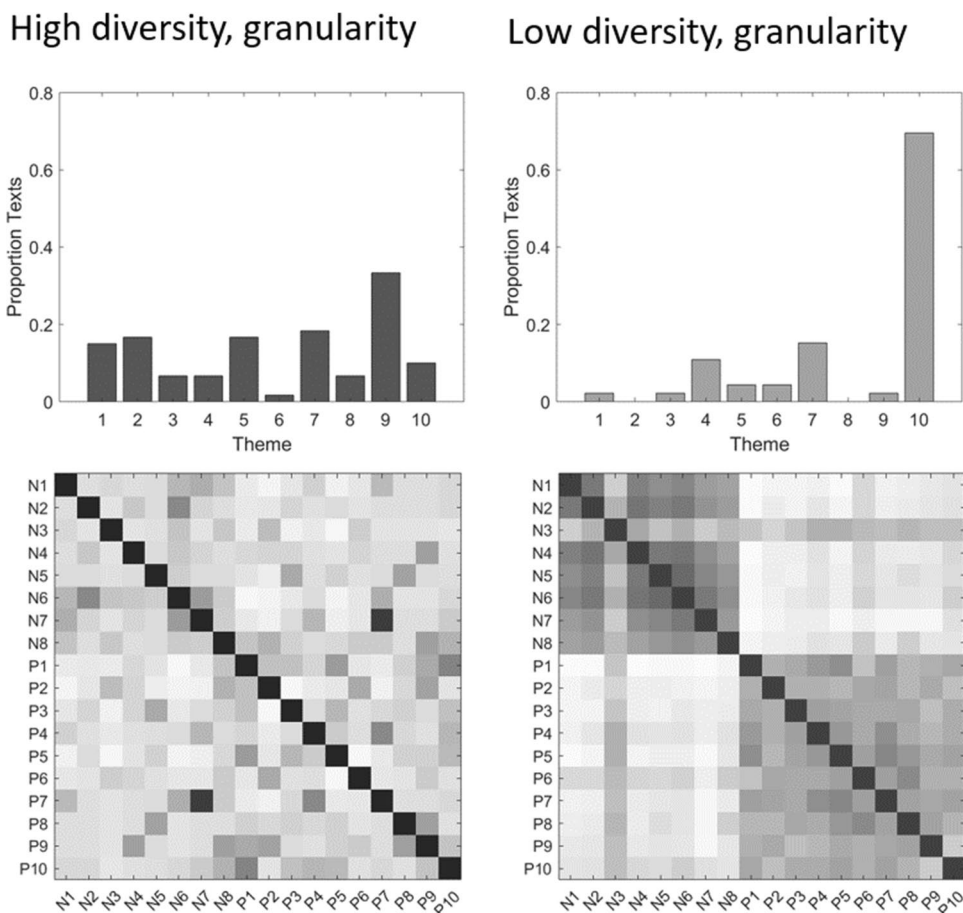
to replicate these findings in another archival data set including both self-reported emotions and written descriptions of everyday events from English-speaking participants.

Study 2

Method

Study 2 was approved by the Northeastern University Institutional Review Board (IRB# 11-04-07) and has been reported in detail in Hoemann, Fan et al. (2020).

Participants An initial 82 adults were recruited from Northeastern University classrooms and online portals. Eligible participants were native English speakers enrolled in years 1–3 of their undergraduate course of study. Informed consent was obtained from participants before beginning the study. Participants received $200 for completing all parts of the study, and a pro-rated amount for partial completion. Participants with a response rate over 90% were entered into a gift card raffle. Six participants were excluded from data analysis due to compliance issues, resulting in a final sample

size of 76 (61% female; 63% White, 11% Black, 13% Asian, 13% other; $M = 19.29$ years, $SD = 1.28$ years). We again conducted a sensitivity analysis to formalize the effect sizes we were powered to detect. Because we had directional predictions based on our results from Study 1, we conducted the Study 2 sensitivity analysis using a one-tailed test of the regression coefficient for experiential diversity (maintaining $\alpha < .05$ and power $(1-\beta) > .80$). This analysis indicated that the Study 2 data were adequately powered to detect small to moderate size effects, $f^2 \geq 0.08$, $\beta \geq .27$.

Procedure Participants completed between 4 and 16 days of experience sampling including end-of-day diaries. Participants also attended three in-lab sessions, in which they completed tasks and questionnaires not reported here.

Participants received a palm pilot programmed for experience sampling, on which they received 10 randomly generated prompts per day between the hours of 8 am and 11 pm. Participants were dismissed from the study if they did not respond to at least 75% of prompts. Altogether, participants completed an average of 8.02 prompts per day ($SD = 2.45$) over an average of 9.47 days ($SD = 3.21$). At each sampling prompt, participants rated the intensity of their emotional

experience on a set of 39 emotion adjectives using Likert-style scales from 1 ("not at all") to 5 ("very much").

Approximately three times during experience sampling (near the beginning, middle, and end), participants completed online end-of-day diaries. They first 'mapped out' their day into a series of episodes, lasting between 15 min and 2 h, and provided a brief name (e.g., "at lunch with friend") as well as the approximate start and end time. Participants then wrote detailed descriptions of three events of their choice. These event descriptions were not directly associated with experience sampling prompts, and indeed may have represented experiences that were not captured by prompts.

Data Preparation

Experiential Diversity As in Study 1, estimates of experiential diversity were computed from the natural language used in the event descriptions (819 total entries). We removed entries that were empty or from participants excluded from the final sample (15.87% of descriptions, new total 689; $M = 9$, $SD = 1$, min $= 6$, max $= 12$ per participant). As in Study 1, we considered descriptions with fewer than 25 words as missing data. This resulted in the removal of an additional 2.47% of event descriptions, for a new total of 672. On average, 2.46% of each participant's texts were removed ($SD = 9.85\%$, min $= 0\%$, max $= 66.67\%$).

In contrast to Study 1, where we had an average of 79 texts per participant at the 25-word minimum (M length $= 131$), Study 2 participants ultimately produced between 3 and 12 event descriptions ($M = 9$, $SD = 1$) that ranged in length between 25 and 576 words ($M = 140$, $SD = 72$). Thus, although these texts were roughly the same length as those in Study 1, they numbered fewer per participant and in total. This meant that an open-vocabulary scoring of texts following the MEM was unlikely to provide stable themes. Instead, we scored the texts in a closed-vocabulary manner using the themes extracted from the Study 1 data (following e.g., Pulverman et al., 2017). Both data sets contain event descriptions in written English, provided in end-of-day diaries by university students. Visual inspection of the Study 2 event descriptions confirmed that they covered highly similar content to the descriptions from Study 1.

To score the Study 2 texts, we created a custom dictionary file for the LIWC2015 software (Pennebaker et al., 2015). LIWC counts the words that belong to a particular category of words within text files and produces a matrix in which the raw count of words within a given category is presented as a proportion of each text's length. Here we defined categories of words based on the themes from Study 1's 'middle' specification (≥ 25 words, 150 unigrams, 10 components). We selected this specification because it represented the ideal

for the Study 1 texts, which were of a similar length (~ 150 words) and elicited in the same manner and language as in Study 2. Using the results of this Study 1 PCA, we removed words with |factor loadings| < 0.20 and included wildcards (e.g., "class*" to capture both the root form and its plural) and other word forms necessary to account for the conversion procedures previously applied in the MEM.

We used the resulting LIWC dictionary to score the Study 2 texts and used these scores to classify each text as either containing or not containing each theme. Specifically, a text was considered to contain a theme if its score on that theme was greater than $M_{theme} + 1.25*SD_{theme}$ (e.g., assuming a standard normal distribution, a text with a score of 1.5 would be coded as '1' for that theme, whereas a text with a score of 0.5 would be coded as '0'). We found that a changepoints analysis, as conducted in Study 1, was too restrictive and resulted an overly sparse data matrix, with themes coded as present in only 5–6% of texts. By contrast, the Study 1 changepoints analysis coded themes as present in approximately 10% of texts. Our selected binarization threshold ($M_{theme} + 1.25*SD_{theme}$) resulted in an equivalent portion of texts per theme as Study 1, which we then used to derive an estimate of experiential diversity for each participant following the procedure described for Study 1. To ensure our results were not an artifact of our specific binarization threshold, we also ran analyses using cutoffs of $M_{theme} + 1*SD_{theme}$ (themes present in about 12% of texts) and $M_{theme} + 1.5*SD_{theme}$ (themes present in about 8% of texts). See the OSF repository for the Study 2 LIWC dictionary and data matrix (https://osf.io/gn8ca/).

Emotional Granularity Estimates of negative and positive emotional granularity were computed from the intensity ratings for the 39 emotion adjectives as described for Study 1. The estimate of negative granularity was based on the ratings of 20 emotions ("angry", "bored", "contemptuous", "depressed", "disgusted", "dislike", "down", "fearful", "furious", "guilty", "hateful", "irritated", "nervous", "remorseful", "repulsed", "sad", "scornful", "shocked", "sorry", "terrified"); the estimate of positive granularity was based on the ratings of 19 emotions ("admiring", "amazed", "amused", "appreciative", "calm", "content", "elated", "enthusiastic", "excited", "grateful", "happy", "joyous", "peaceful", "prideful", "relaxed", "restful", "successful", "superior", "surprised"). No negative ICC values were observed in this data set.

Analysis Regression analyses proceeded as described for Study 1. We had directional predictions based on the prior effects and so used one-tailed tests of significance at $\alpha = 0.05$.

Results

Experiential diversity was positively associated with negative emotional granularity when controlling for the number

of event descriptions and mean negative affect: β=0.31, 95% CI [0.10, 0.51], $t(67)$=3.00, p=.002, one-tailed, 5 outliers excluded. Experiential diversity was not associated with positive emotional granularity: β=0.04, 95% CI [-0.18, 0.25], $t(66)$=0.34, p=.37, one-tailed, 6 outliers excluded. These results held when using the alternative binarization thresholds noted above. Full results for all analyses are provided in supplemental Tables S5a-b; scatter plots are presented in supplemental Figure S3.

Discussion

In Study 2, English-speaking participants provided emotion intensity ratings for everyday events via in-the-moment prompts and written descriptions of a subset of these events via end-of-day diaries. Replicating Study 1, experiential diversity was positively associated with negative emotional granularity; however, experiential diversity was not associated with positive emotional granularity. To further examine the robustness of these effects, in Study 3 we extended our secondary analyses to a data set of self-reported emotions and descriptions of everyday events in spoken Dutch.

Study 3

Method

Study 3 was approved by the KU Leuven Social and Societal Ethics Committee (protocol G-2018-01-1095) and has been reported in detail in Carlier et al. (2021).

Participants An initial 66 adults were recruited through announcements posted to social media groups and other places frequently visited by KU Leuven students. Eligible participants were native Dutch speakers between 18 and 30 years old with a compatible Android smartphone. Informed consent was obtained from participants before beginning the study. Participants received 50€ for completing all parts of the study, and a pro-rated amount for partial completion. Three participants did not complete experience sampling, 10 participants did not produce any event descriptions, and an additional four were removed during data preparation (see below for details), resulting in a final sample size of 49 (72% female; M=21.53 years, SD=1.84 years). A sensitivity analysis conducted as in Study 2, assuming a one-tailed test of significance, indicated that this data set was adequately powered to detect moderate size effects, $f^2 \geq 0.13$, β≥0.34.

Procedure Participants completed 14 days of experience sampling. Participants also completed questionnaires that are not reported here. During experience sampling,

participants received 10 randomly generated prompts per day during waking hours. Participants received full compensation if they responded to at least 80% of prompts. Altogether, participants completed an average of 7.48 prompts per day (SD=2.15).

At each sampling prompt, participants rated the intensity of their emotional experience on a set of 7 emotion adjectives using a 100-point continuous slider bar. Participants were also asked to verbally describe what was happening and how they were feeling by recording 1–2 min of speech into the smartphone app. This was not required; as such, not all experience sampling prompts had a corresponding event description. The 10 participants who chose not to record any event descriptions were excluded from analysis. Participants also responded to other questions at each prompt and mobile sensing data (e.g., movement, phone use) were collected throughout the study; these variables are not analyzed in the present report.

Data Preparation

Experiential Diversity Estimates of experiential diversity were once again computed from the natural language used in the recorded event descriptions (1,039 total recordings; M=20, SD=23, min=1, max=98 per participant), which were manually transcribed by a separate team of research assistants. By design, there were no empty descriptions; there were also no descriptions from excluded participants, because recordings from these participants were not transcribed. We once again considered descriptions with fewer than 25 words as missing data (10.30% of descriptions, new total 932), removing 19.26% of each participant's descriptions on average (SD=30.12%, min=0%, max=100%). This minimum text length meant that four additional participants, who had originally produced between 2 and 5 event descriptions each, no longer had any usable data. Study 3 participants ultimately produced between 1 and 83 event descriptions (M=19, SD=22). These descriptions were shorter than those in Studies 1 and 2, ranging in length between 25 and 230 words (M=83, SD=42). Additionally, these texts represented spoken (as opposed to written) language (in this case, Dutch) and so contained hesitations, incomplete sentences, and fewer content words in general.

Due to the reduced length and number of the Study 3 texts, as well as the relative sparsity of content words they contained, we opted to score them using the closed-vocabulary approach described in Study 2. Before doing so, we inspected a random subset of the texts and confirmed that they again, coming from a predominantly student sample, covered similar content as the other two data sets. Next, we translated our custom LIWC dictionary from Study 2 into Dutch. Using the same set of themes

increased the comparability of results across all three studies. To ensure that the translations were appropriate to our data, we extracted the frequencies of all content words in the Study 3 event descriptions and selected translations from the Dutch words that appeared in this list. We then added any wildcards, alternative spellings, and other word forms necessary to account for the conversion procedures applied in the MEH analysis. We used our custom Dutch LIWC dictionary to score the Study 3 texts and used these scores to classify each text as containing or not containing each theme – as in Study 2, considering texts with scores $> M_{theme} + 1.25 * SD_{theme}$ to evidence a given theme and confirming our results at two other binarization thresholds. We then derived an estimate of experiential diversity for each participant as described for Studies 1 and 2. See the OSF repository for the Study 3 LIWC dictionary, supporting files, and resulting data matrix (https://osf.io/gn8ca/).

Emotional Granularity Estimates of negative and positive emotional granularity were computed from the intensity ratings for the 7 emotion adjectives as described for Studies 1 and 2. The estimate of negative granularity was based on the ratings of 5 emotions ("angstig" [anxious], "droevig" [sad], "gestresseerd" [stressed], "kwaad" [angry], "moe" [tired]). Three participants returned ICCs for negative granularity with a negative value that were recoded as 0. The estimate of positive granularity was based on the ratings of 2 emotions ("blij" [happy], "relaxed" [relaxed]). Guidelines for ICC interpretation have suggested a minimum of 3 raters (here, emotions) for reliable estimation (Koo & Li, 2017), although others have clarified that the number of required raters will depend on the variance between them (Saito et al., 2006). With this in mind, we interpret the results of our positive granularity analyses with caution.

Analysis Regression analyses proceeded as described for Study 2.

Results

Experiential diversity was positively associated with negative emotional granularity when controlling for the number of event descriptions and mean negative affect: $\beta = .34$, 95% CI [0.02, 0.67], $t(40) = 2.16$, $p = .02$, one-tailed, 5 outliers excluded. Experiential diversity was not associated with positive emotional granularity: $\beta = .02$, 95% CI [-0.34, 0.37], $t(42) = 0.09$, $p = .47$, one-tailed, 3 outliers excluded. These results also held when using alternative binarization thresholds. Full results for all analyses are provided in supplemental Tables S6a-b; scatter plots are presented in supplemental Figure S4.

Discussion

In Study 3, Dutch-speaking participants provided emotion intensity ratings and spoken descriptions of everyday events via in-the-moment prompts. Replicating Studies 1 and 2, experiential diversity was positively associated with negative emotional granularity, suggesting that this effect is robust to differences in language (English versus Dutch), modality (written versus spoken), and study protocol. Experiential diversity was not associated with positive emotional granularity. Although this finding should be viewed with caution due to the small number of positive emotions sampled, it joins Study 2 in suggesting that the effect of experiential diversity on positive emotional granularity is not as stable as on negative granularity.

Meta- and Integrative Data Analyses

The sensitivity analyses from Studies 1–3 suggest that these data sets were generally powered to detect the significant effects we observed, although there were minor deviations (e.g., Study 1 was powered for $\beta \geq .37$ but the median effect of experiential diversity on negative emotional granularity was $\beta = 0.34$). To solidify our findings and increase statistical power, we quantified the results of Studies 1–3 using a meta-analysis (Goh et al., 2016). A sensitivity analysis for three studies (i.e., $k = 3$), assuming an average within-study sample size of 55, $\alpha < .05$, and power $(1-\beta) > .80$ (Hedges & Pigott, 2001), indicated that we were powered to detect small-to-moderate size meta-analytic effects ($M_r \geq 0.22$).

We represented the unique variance captured by experiential diversity using the corresponding t values, and converted these to standard effect sizes (r) following established formulae (Borenstein et al., 2011). For Study 1, we used the median effect size across all 27 possible analyses. We then meta-analyzed these r values using a fixed effects model, in which the mean effect size was weighted by sample size (with outliers removed). Across all three studies, experiential diversity was significantly and positively associated with negative emotional granularity (controlling for number of event descriptions and mean negative affect): $M_r = 0.35$, 95% CI [0.20, 0.48], $Z = 4.51$, $p < .001$, two-tailed. These results held when using the effect size from Study 1's 'middle' specification (≥ 25-word texts, 150 unigrams, 10 components): $M_r = 0.38$, 95% CI [0.24, 0.51], $Z = 4.93$, $p < .001$, two-tailed. Experiential diversity was not associated with positive emotional granularity: $M_r = 0.12$, 95% CI [-0.03, 0.28], $Z = 1.54$, $p = .12$, two-tailed. These results also held when using the effect size from Study 1's 'middle' specification: $M_r = 0.15$, 95% CI [-0.004, 0.30], $Z = 1.91$, $p = .06$, two-tailed. Because the estimate of positive granularity from Study 3 was based on only two emotions, we examined the

impact of removing these results from the meta-analysis (see supplemental Table S7). Meta-analyses were performed in R using the meta package (Balduzzi et al., 2019).

The fact that each study sampled a different number of emotion adjectives (18 in Study 1, 39 in Study 2, 7 in Study 3) may have impacted the comparability of emotional granularity estimates across studies. It could be that the ability to choose from a larger number of emotion adjectives encouraged more granular patterns of rating; equally, the presence of more emotion adjectives could have resulted in terms being treated as synonyms for ease of rating. To address these possibilities, we simultaneously analyzed data from all three studies using integrative data analyses (IDAs; Curran & Hussong, 2009). We conducted fixed-effects IDAs in which we treated study membership as a property of each participant, which allowed us to account for differences in study design, including the number of emotion adjectives sampled. Study membership was coded for simple effects, with Study 1 as reference. All other variables were standardized within study and no outliers were removed. For Study 1, we used the estimates of experiential diversity that produced the median effect sizes. A sensitivity analysis for the regression included five tested (experiential diversity along with Study 2 and Study 3 effects and their respective interactions) and two control predictors (number of event descriptions, mean affect) indicated that the total sample size ($N = 175$) was powered to detect small effect sizes, $f^2 \geq 0.08$, $\beta \geq .27$. IDAs were performed in R using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages. Full results of the IDA models are presented in Table 2.

Consistent with the meta-analytic results, experiential diversity was significantly and positively associated with negative emotional granularity: $\beta = .27$, 95% CI [0.13, 0.42], $t(167) = 3.67$, $p = 0.001$, two-tailed. There were no effects of study or interactions between study and experiential diversity (all p's $\geq .50$), suggesting that differences in design were unlikely to have impacted the relationship between experiential diversity and emotional granularity. Mean negative affect was a significant predictor of negative emotional granularity ($\beta = -0.46$, $p < .001$), such that participants who reported higher intensity unpleasant affect also reported less differentiated and nuanced negative emotions. These results held when using the estimates of experiential diversity from Study 1's 'middle' specification (see supplemental Table S8). They also held when removing the control predictors from the model in case these were capturing variance attributable to differences in study design (see supplemental Table S9).

An IDA of the relationship between experiential diversity and positive emotional granularity likewise revealed a significant positive relationship: $\beta = .18$, 95% CI [0.01, 0.34], $t(167) = 2.12$, $p = .04$, two-tailed. Based on

the sensitivity analysis, however, we were underpowered to detect this effect and so it should be interpreted with caution. All other effects were non-significant, including study and interactions between study and experiential diversity (all p's $\geq .07$). These results held when using estimates from Study 1's 'middle' specification (supplemental Table S8) and when using a simplified regression model (supplemental Table S9). As with the meta-analysis, we examined the impact of removing the Study 3 data from the IDA (supplemental Table S10).

Our finding of a robust effect for negative, but not positive, emotional granularity might be due to differences in the distribution of these variables. To address this possibility, we conducted exploratory analyses in which we compared the mean and range of negative and positive granularity within each study. Our results are reported in supplemental Table S11. We found that granularity for negative emotions was significantly higher than for positive emotions in Studies 2 and 3. However, there was no systematic restriction of range in positive granularity; indeed, in Study 3 we observed a larger range for positive than negative granularity. This implies that a restriction of range was not responsible for the smaller, less stable effect for positive granularity.

Table 2 Integrative data analysis results

Regression Term	β	95% CI	t	df	p
Experiential diversity and negative emotional granularity					
Intercept	< .001	-.13, .13	0.00	167	1.00
Study 2	< .001	-.31, .31	0.00	167	1.00
Study 3	< .001	-.34, .34	0.00	167	1.00
Number of event descriptions	-.11	-.26, .03	-1.53	167	.13
Mean negative affect	-.46	-.59, -.33	-7.00	167	< .001
Experiential diversity	.27	.13, .42	3.67	167	< .001
Experiential diversity x Study 2	-.04	-.35, .26	-0.28	167	.78
Experiential diversity x Study 3	-.12	-.46, .23	-0.67	167	.50
Experiential diversity and positive emotional granularity					
Intercept	< .001	-.15, .15	0.00	167	1.00
Study 2	< .001	-.35, .35	0.00	167	1.00
Study 3	< .001	-.39, .39	0.00	167	1.00
Number of event descriptions	.08	-.09, .24	0.92	167	.36
Mean positive affect	-.04	-.19, .11	-0.48	167	.63
Experiential diversity	.18	.01, .34	2.12	167	.04
Experiential diversity x Study 2	-.33	-.69, .03	-1.81	167	.07
Experiential diversity x Study 3	-.29	-.70, .12	-1.41	167	.16

Note: All tests of significance are two-tailed

General Discussion

Across three studies varying in language (English versus Dutch) and modality (written versus spoken), we found that experiential diversity was positively associated with negative emotional granularity: Participants who referred to a more varied and balanced set of contexts and activities in their daily lives also reported more differentiated and nuanced negative emotions. However, experiential diversity was not consistently associated with positive emotional granularity, with a clear positive relationship emerging in only one study. These findings suggest a link between the content of daily life and how it is made meaningful as (negative) emotion, laying a path for future research that can investigate the nature and possible applications of this link.

Individual differences in emotional granularity are thought to reflect differences in emotion concepts that, as dynamic collections of prior experience, are used by the brain to construct current and future experience (e.g., Barrett, 2017). As such, greater variation in experience could be related to higher emotional granularity both as a source – because it allows the brain to issue more finely-tuned predictions – and as an outcome – because context-specific predictions result in nuanced, differentiated experiences. The present findings support this general hypothesis most clearly for negative emotion. In doing so, they join research showing that emotional granularity is positively related to diversity and specificity in patterns of peripheral physiological activity in everyday life (Hoemann, Khan et al., 2021), and more broadly to adaptive patterns of behavior (e.g., O'Toole et al., 2020).

The cross-sectional nature of our data means that we cannot use them to infer causality. However, there is already evidence of how life experiences and other forms of accrued knowledge may shape emotional granularity. For example, overall vocabulary has been found to mediate the development of more complex emotion concepts in children (Nook et al., 2017). Recent work has also shown that emotion concept learning interventions can increase emotional granularity in adults (Vedernikova et al., 2021). Future, longitudinal research is needed to directly test whether changes in contexts and activities, such as moves and job transitions, lead to shifts in granularity, and the mediating role of emotion concepts.

Emotional granularity may also be related to experiential diversity through domain-general processes that are not limited or specific to emotion. For instance, individuals with higher granularity may differentiate more between all types of experiences (Richmond & Zacks, 2017), as they do with emotions, because they segment events more finely (e.g., Kurby & Zacks, 2008). Individuals with higher granularity may also describe a greater diversity of contexts and activities because they are more mindful (e.g., Tong & Keng, 2017), which helps them to be more fluid or 'in the moment', letting go of past and possible future experiences. Additionally, individuals

with higher granularity may lead more psychologically rich lives, seeking a variety of interesting and perspective-changing experiences (e.g., travel; Oishi & Westgate, 2021) that, in turn, build more granular emotion concepts. These and other possible relationships should be tested by future research.

In the present analyses, we found that positive emotional granularity was not reliably associated with experiential diversity. Although contrary to our prediction, this finding is in keeping with the broader literature, in which more consistent relationships are observed for negative than for positive granularity (e.g., O'Toole et al., 2020). Differential associations for negative versus positive granularity are often interpreted with regard to the functions served by negative versus positive emotions (e.g., O'Toole et al., 2020; Thompson et al., 2021). These differences may also be anchored in well-known asymmetries in how negative versus positive experiences are processed (Alves et al., 2017), with negative information capturing more attention (e.g., Pratto & John, 1991) and perceived as more self-relevant (Taylor, 1991). Negative emotions are also associated with more analytic and elaborative processing (Schwarz & Bless, 1991), which may be related to how participants break down daily experiences into component themes and warrants further study.

It is important to note that the generalizability of the present findings is limited by the method used to estimate experiential diversity in Studies 2 and 3. Because we had fewer, and shorter, event descriptions for participants in these studies, we were unable to derive themes in an open-vocabulary manner, instead scoring event descriptions using the themes derived in Study 1. It is likely that this process only yielded meaningful results because all three samples were comprised predominantly of university students who described a similar range of contexts and activities in their daily lives. To assess whether emotional granularity is associated with experiential diversity outside of these circumstances, studies are needed that capture sufficient natural language descriptions of everyday events, as well as emotion intensity ratings, in larger and more diverse samples.

Another consideration is that the present studies (especially Study 1) required a notable amount of self-monitoring and diligence from participants. It is possible that continually attending to emotions and experiences influenced behaviors in and interpretations of everyday events. That is, our findings may have been shaped by measurement reactivity or a Hawthorne effect, in which the phenomenon measured changes because of the tools used to measure it. There is indeed evidence that intensive sampling influences experience and awareness of emotions over time (Eisele et al., 2022), although these effects may be comparable across participants and mitigated by study parameters such as randomized prompts (Myin-Germeys et al., 2018). At the same time, asking people to pay more regular or careful attention to their emotions may have translational benefits, with recent

studies finding increases in emotional granularity following experience sampling (Hoemann, Barrett et al., 2021; Widdershoven et al., 2019) and mindfulness-based interventions (Van der Gucht et al., 2019).

In future research, experiential diversity may also be leveraged as a means of increasing emotional granularity. Diversification of affective experience is hypothesized to facilitate improvement in emotional granularity (Barrett, 2017), whose associations with positive outcomes make it a compelling target for intervention (e.g., Thompson et al., 2021). As mentioned above, recent studies suggest that shifting *how* people attend to their everyday experiences makes a difference for emotional granularity. However, studies have yet to explore how emotional granularity may be impacted by shifting *what* people encounter when they go about their daily lives. Building on other recent research, for example, future research could examine how emotional granularity is impacted by increased variation in daily activities (Lee et al., 2021) or in spatial and sociodemographic environments (Heller et al., 2020). Such work has the potential to not only establish novel, behavior-based interventions, but to reveal causal paths linking experiential diversity to emotional granularity.

Additional Information

Availability of Data and Material Anonymized data and analytic code are available via a repository hosted by the Center for Open Science (OSF) at https://osf.io/gn8ca/. To protect participant privacy, the raw language data analyzed in the current studies are not publicly available.

Author Contributions K.H. assisted with data collection for Study 1. M.G. collected the data for Study 2. P.K. oversaw data collection for Study 3. K.H. and Y.L. analyzed the data. K.H. drafted the manuscript. All authors provided revisions and approved the final version of the manuscript.

Ethics Approval These studies were conducted in line with the principles of the Declaration of Helsinki. Approval for Studies 1 and 2 was granted by the Institutional Review Board at Northeastern University (IRB#s 13-03-16, 16-11-32). Approval for Study 3 was granted by the KU Leuven Social and Societal Ethics Committee (protocol G-2018-01-1095).

Consent to Participate Informed consent was obtained from all individual participants included in the studies.

Consent for Publication Not applicable.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42761-023-00185-2.

References

Alves, H., Koch, A., & Unkelbach, C. (2017). Why good is more alike than bad: Processing implications. *Trends in Cognitive Sciences, 21*(2), 69–79. https://doi.org/10.1016/j.tics.2016.12.006

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health, 22*, 153–160. https://doi.org/10.1136/ebmental-2019-300117

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*(1), 20–46. https://doi.org/10.1207/s15327957pspr1001_2

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion, 15*(6), 713–724. https://doi.org/10.1080/02699930143000239

Bates, D., Mächler, 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

Benson, L., Ram, N., Almeida, D. M., Zautra, A. J., & Ong, A. D. (2018). Fusing biodiversity metrics into investigations of daily life: Illustrations and recommendations with emodiversity. *The Journals of Gerontology: Series B, 73*(1), 75–86. https://doi.org/10.1093/geronb/gbx025

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement, 65*, 676–696.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons.

Boyd, R. L. (2017). Psychological text analysis in the digital humanities. In S. Hai-Jew (Ed.), *Data analytics in digital humanities* (pp. 161–189). Springer International Publishing. https://doi.org/10.1007/978-3-319-54499-1_7

Boyd, R. L. (2018). *MEH: Meaning Extraction Helper* (2.2.03). https://www.ryanboyd.io/software/meh

Carlier, C., Niemeijer, K., Mestdagh, M., Bauwens, M., Vanbrabant, P., Geurts, L., van Waterschoot, T., & Kuppens, P. (2021). In search of state and trait emotion markers in mobile-sensed language: A field study. *JMIR Mental Health*. https://doi.org/10.2196/31724

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276. https://doi.org/10.1207/s15327906mbr0102_10

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality, 42*(1), 96–132. https://doi.org/10.1016/j.jrp.2007.04.006

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods, 14*(2), 81–100. https://doi.org/10.1037/a0015914

Cutts, M. (2009). *Oxford guide to plain English (3rd ed.)*. Oxford University Press.

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature Human Behaviour, 3*(5), 478–491. https://doi.org/10.1038/s41562-019-0555-0

Eisele, G., Vachon, H., Lafit, G., Tuyaerts, D., Houben, M., Kuppens, P., Myin-Germeys, I., & Viechtbauer, W. (2022). A mixed-method investigation into measurement reactivity to the experience sampling method: The role of sampling protocol and individual characteristics. *Psychological Assessment,* https://doi.org/10.1037/pas0001177

Entwistle, C., Horn, A. B., Meier, T., & Boyd, R. L. (2021). Dirty laundry: The nature and substance of seeking relationship help from strangers online. *Journal of Social and Personal Relationships.* https://doi.org/10.1177/02654075211046635

Erbas, Y., Ceulemans, E., Lee Pe, M., Koval, P., & Kuppens, P. (2014). Negative emotion differentiation: Its personality and well-being correlates and a comparison of different assessment methods. *Cognition and Emotion, 28*(7), 1196–1213. https://doi.org/10.1080/02699931.2013.875890

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass, 10*(10), 535–549. https://doi.org/10.1111/spc3.12267

Hedges, L. V., & Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6*(3), 203–217. https://doi.org/10.1037/1082-989X.6.3.203

Heller, A. S., Shi, T. C., Ezie, C. E. C., Reneau, T. R., Baez, L. M., Gibbons, C. J., & Hartley, C. A. (2020). Association between real-world experiential diversity and positive affect relates to hippocampal–striatal functional connectivity. *Nature Neuroscience, 23*(7), 800–804. https://doi.org/10/ghhvms.

Hoemann, K., Fan, M., Engen, H., Chou, C.-A., Quigley, K. S., Gendron, M., & Barrett, L. F. (2020). A network analytic approach to measuring emotional granularity. *PsyArXiv.* https://doi.org/10.31234/osf.io/yt9cv

Hoemann, K., Khan, Z., Feldman, M. J., Nielson, C., Devlin, M., Dy, J., Barrett, L. F., Wormwood, J. B., & Quigley, K. S. (2020). Context-aware experience sampling reveals the scale of variation in affective experience. *Scientific Reports, 10*, 12459. https://doi.org/10.1038/s41598-020-69180-y

Hoemann, K., Barrett, L. F., & Quigley, K. S. (2021). Emotional granularity increases over the course of experience sampling: Methodological and individual factors influence how much. *Frontiers in Psychology, 12*, 704125. https://doi.org/10.3389/fpsyg.2021.704125

Hoemann, K., Khan, Z., Kamona, N., Dy, J., Barrett, L. F., & Quigley, K. S. (2021). Investigating the relationship between emotional granularity and cardiorespiratory physiological activity in daily life. *Psychophysiology, 58*(6), e13818. https://doi.org/10.1111/psyp.13818

Hoemann, K., Nielson, C., Yuen, A., Gurera, J. W., Quigley, K. S., & Barrett, L. F. (2021). Expertise in emotion: A scoping review and unifying framework for individual differences in the mental representation of emotional experience. *Psychological Bulletin, 147*(11), 1159–1183. https://doi.org/10.1037/bul0000327

Israelashvili, J., Oosterwijk, S., Sauter, D., & Fischer, A. (2019). Knowing me, knowing you: Emotion differentiation in oneself is associated with recognition of others' emotions. *Cognition and Emotion, 33*(7), 1461–1471. https://doi.org/10.1080/02699931.2019.1577221

Kalokerinos, E. K., Erbas, Y., Ceulemans, E., & Kuppens, P. (2019). Differentiate to regulate: Low negative emotion differentiation is associated with ineffective use but not selection of emotion-regulation strategies. *Psychological Science, 30*(6), 863–879. https://doi.org/10.1177/0956797619838763

Kashdan, T. B., Ferssizidis, P., Collins, R. L., & Muraven, M. (2010). Emotion differentiation as resilience against excessive alcohol use: An ecological momentary assessment in underage social drinkers. *Psychological Science, 21*(9), 1341–1347. https://doi.org/10.1177/0956797610379863

Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and visualize the results of multivariate data analyses* (R package version 1.0.7). https://CRAN.R-project.org/package=factoextra

Killick, R., & Eckley, I. (2014). changepoint: An R package for changepoint analysis. *Journal of Statistical Software, 58*(3), 1–19. https://doi.org/10/gddvvn.

Koo, T. K., & Li, M. Y. (2017). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 16*(4), 346–346. https://doi.org/10.1016/j.jcm.2017.10.001

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12*(2), 72–79. https://doi.org/10/bk23jd.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, S., Urban-Wojcik, E. J., Charles, S. T., & Almeida, D. M. (2021). Rich and balanced experiences of daily emotions are associated with activity diversity across adulthood. *The Journals of Gerontology: Series B*, gbab144.

MATLAB. (2018). *9.5.0.1033004 (R2018b Update 2)*. The Mathworks, Inc.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex, 48*(7), 788–804. https://doi.org/10.1016/j.cortex.2010.11.002

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry, 17*(2), 123–132. https://doi.org/10/gdnp29.

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour, 1*(12), 881–889. https://doi.org/10.1038/s41562-017-0238-7

Nook, E. C., Satpute, A. B., & Ochsner, K. N. (2021). Emotion naming impedes both cognitive reappraisal and mindful acceptance strategies of emotion regulation. *Affective Science, 2*(2), 187–198. https://doi.org/10/gk85jv.

Oishi, S., & Westgate, E. C. (2021). A psychologically rich life: Beyond happiness and meaning. *Psychological Review.*

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour, 3*(2), 173–182. https://doi.org/10.1038/s41562-018-0506-1

O'Toole, M. S., Renna, M. E., Elkjær, E., Mikkelsen, M. B., & Mennin, D. S. (2020). A systematic review and meta-analysis of the association between complexity of emotion experience and behavioral adaptation. *Emotion Review, 12*(1), 23–38. https://doi.org/10/ggcb24.

Ottenstein, C., & Lischetzke, T. (2019). Development of a novel method of emotion differentiation that uses open-ended descriptions of momentary affective states. *Assessment*. https://doi.org/10.1177/1073191119839138. 107319111983913.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*(3), 12.

Pulverman, C. S., Boyd, R. L., Stanton, A. M., & Meston, C. M. (2017). Changes in the sexual self-schema of women with a history of childhood sexual abuse following expressive writing treatment. *Psychological Trauma: Theory, Research, Practice, and Policy, 9*(2), 181. https://doi.org/10/f9zm2q.

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

Ramirez-Esparza, N., Chung, C. K., Kacewicz, E., & Pennebaker, J. W. (2008). *The psychology of word use in depression forums in English and in Spanish: Testing two text analytic approaches*. 7.

Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research* (2.0.12). https://CRAN.R-project.org/package=psych

Richmond, L. L., & Zacks, J. M. (2017). Constructing experience: Event models from perception to action. *Trends in Cognitive Sciences, 21*(12), 962–980. https://doi.org/10.1016/j.tics.2017.08.005

Rodríguez-Arauz, G., Ramírez-Esparza, N., Pérez-Brena, N., & Boyd, R. L. (2017). Hablo inglés y español: Cultural self-schemas as a function of language. *Frontiers in Psychology, 8*, 885. https://doi.org/10.3389/fpsyg.2017.00885

Saito, Y., Sozu, T., Hamada, C., & Yoshimura, I. (2006). Effective number of subjects and number of raters for inter-rater reliability studies. *Statistics in Medicine, 25*(9), 1547–1560. https://doi.org/10.1002/sim.2294

Salarian, A. (2016). *Intraclass Correlation Coefficient (ICC)* (1.3.0.0) [MATLAB]. https://www.mathworks.com/matlabcentral/fileexchange/22099-intraclass-correlation-coefficient-icc

Schwartz, H. A., Sap, M., Kern, M. L., Eichstaedt, J. C., Kapelner, A., Agrawal, M., Blanco, E., Dziurzynski, L., Park, G., Stillwell, D., Kosinski, M., Seligman, M. E. P., & Ungar, L. H. (2016). Predicting individual well-being through the language of social media. *Biocomputing, 2016*, 516–527. https://doi.org/10.1142/9789814749411_0047

Schwarz, N., & Bless, H. (1991). Happy and mindless, but sad and smart? The impact of affective states on analytic reasoning. In J. P. Forgas (Ed.), *Emotion and social judgments* (pp. 55–71). Taylor & Francis.

Seah, T. H. S., Aurora, P., & Coifman, K. G. (2020). Emotion differentiation as a protective factor against the behavioral consequences of rumination: A conceptual replication and extension in the context of social anxiety. *Behavior Therapy, 51*(1), 135–148. https://doi.org/10.1016/j.beth.2019.05.011

Seah, T. H. S., & Coifman, K. (2021). Emotion differentiation and behavioral dysregulation in clinical and non-clinical samples: A meta-analysis. *Emotion*. https://doi.org/10.1037/emo0000968

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 1–7.

Statistics and Machine Learning Toolbox (11.4). (2019). The Mathworks, Inc.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702–712. https://doi.org/10/f88j4q.

Sun, J., Kern, M. L., Schwartz, H. A., Son, Y., & Vazire, S. (2019). The language of well-being: Tracking fluctuations in emotion experience through everyday speech. *Journal of Personality and Social Psychology, 118*(2), 364–387.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization-minimization hypothesis. *Psychological Bulletin, 110*(1), 67–85. https://doi.org/10.1037/0033-2909.110.1.67

Thompson, R. J., Springstein, T., & Boden, M. (2021). Gaining clarity about emotion differentiation. *Social and Personality Psychology Compass, 15*(3), e12584. https://doi.org/10/gjjq58.

Tong, E. M. W., & Keng, S.-L. (2017). The relationship between mindfulness and negative emotion differentiation: A test of multiple mediation pathways. *Mindfulness, 8*(4), 933–942. https://doi.org/10/gbpk2b.

Tugade, M. M., Fredrickson, B. L., & Barrett, L. F. (2004). Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of Personality, 72*(6), 1161–1190. https://doi.org/10.1111/j.1467-6494.2004.00294.x

Van der Gucht, K., Dejonckheere, E., Erbas, Y., Takano, K., Vandemoortele, M., Maex, E., Raes, F., & Kuppens, P. (2019). An experience sampling study examining the potential impact of a mindfulness-based intervention on emotion differentiation. *Emotion, 19*(1), 123–131. https://doi.org/10.1037/emo0000406

Vedernikova, E., Kuppens, P., & Erbas, Y. (2021). From knowledge to differentiation: Increasing emotion knowledge through an intervention increases negative emotion differentiation. *Frontiers in Psychology, 12*.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). Springer. http://www.stats.ox.ac.uk/pub/MASS4/

Vine, V., Bernstein, E. E., & Nolen-Hoeksema, S. (2019). Less is more? Effects of exhaustive vs minimal emotion labelling on emotion regulation strategy planning. *Cognition and Emotion, 33*(4), 855–862. https://doi.org/10.1080/02699931.2018.1486286

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Widdershoven, R. L., Wichers, M., Kuppens, P., Hartmann, J. A., Menne-Lothmann, C., Simons, C. J., & Bastiaansen, J. A. (2019). Effect of self-monitoring through experience sampling on emotion differentiation in depression. *Journal of Affective Disorders, 244*, 71–77. https://doi.org/10/gfj6cz.

Williams, G. E., & Uliaszek, A. A. (2021). Measuring negative emotion differentiation via coded descriptions of emotional experience. *Assessment*, 10731911211003948.

Willroth, E. C., Flett, J. A. M., & Mauss, I. B. (2019). Depressive symptoms and deficits in stress-reactive negative, positive, and within-emotion-category differentiation: A daily diary study. *Journal of Personality*. https://doi.org/10.1111/jopy.12475. jopy.12475.