Toward an ecology of emotion in everyday life

Ratna Kandala[1], Niels Vanhasbroeck[2], Bastiaan Tamm[3,4], Hugo Van hamme[3], Peter Kuppens[4], Batja Mesquita[4], & Katie Hoemann[1]

1. Department of Psychology, University of Kansas, Lawrence, KS, USA
2. Faculty of Social and Behavioral Sciences, University of Amsterdam, Amsterdam, NL
3. Department of Electrical Engineering, KU Leuven, Leuven, BE
4. Laboratory for Cognitive Neurology, KU Leuven, Leuven, BE
5. Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, BE

Corresponding author:

Katie Hoemann
Department of Psychology
1415 Jayhawk Blvd, Lawrence, KS, USA 66045
+1 785 864 2717
hoemann@ku.edu

Abstract

Human emotional life is rich and varied, yet psychological science often studies it in decontextualized ways, asking participants to rate feelings without capturing the situations that evoke them. Here, we map the ecology of everyday emotion by analyzing the link between affective experience and daily events. We collected thousands of real-time verbal descriptions of momentary affect from participants over 70 days. Using transformer-based language models, we identified granular topics corresponding to daily situations, from 'morning routines' to 'socializing with friends.' Our analysis revealed that the mapping between situations and emotions is not one-to-one, but many-to-many. For example, a single situation, such as 'reading books,' was associated with a wide spectrum of emotions, while a single emotion word, such as “angry,” appeared across a diverse range of situations. This data-driven approach moves beyond simple emotion labels to build a foundational ecology of human affective experience.

Keywords: experience sampling, ambulatory assessment, natural language processing, topic modeling, emotion categories

Our everyday lives are populated with a beautiful and dizzying array of emotional experiences. These experiences vary in the categories of emotion that they represent (e.g., pride, anger, sadness, amusement) and range along key dimensions, such as their degree of pleasantness (valence) and energy (arousal). To capture these moments, ambulatory assessment methods such as experience sampling have been developed and are increasingly implemented (Conner & Mehl, 2015; Fritz et al., 2024). Via smartphone apps, diaries, and more, researchers can get at the diverse, personally meaningful events at the core of emotion (Kuppens et al., 2022). At least, that is the idea. In practice, however, participants are often asked to label their feelings or to rate their levels of valence and arousal without providing any additional information about surrounding situations and activities. Potentially rich data about emotion in everyday life are reduced to a series of individual words, or even just points in coordinate space (Klipstein et al., 2023; Myin-Germeys et al., 2018). In this paper, we report on a project that seeks to catalogue the emotion-related events that constitute daily life, joining a growing line of work (Hoemann, Lee, et al., 2023; Sun et al., 2019). This ecological approach, we argue, is the first step toward linking everyday events to the mechanisms and outcomes that matter most.

Ambulatory assessment provides a powerful set of tools for studying psychological phenomena outside of the lab. In studies using the experience sampling method (ESM; Csikszentmihalyi & Larson, 1987) or ecological momentary assessment (EMA; Stone & Shiffman, 1994), participants report on their current or recent experiences multiple times per day. In studies using daily diary methods (e.g., the Day Reconstruction Method; Kahneman et al., 2004; Stone et al., 2006), participants reflect on their experiences over the course of the past day. Either approach can be implemented for a few days, a few weeks, or even a few months (Wrzus & Neubauer, 2022). The goal is to collect (near) real-time data that better reflects how feelings, thoughts, and behaviors unfold in real-world contexts in ways that are less influenced by beliefs or heuristics (Robinson & Clore, 2002; Schneider et al., 2020; Wilhelm & Grossman, 2010). Due to their high ecological validity and ability to recover within-person fluctuations over time, ambulatory assessment methods have seen remarkable uptake in the science of emotion (for a recent review, see Kuppens et al., 2022). These methods have been used to demonstrate relationships between emotion dynamics and key moment-to-moment and person-to-person differences in coping, decision-making, mental health, and more (e.g., Pond et al., 2012; Snippe et al., 2023; Tomko et al., 2015; for reviews and discussion, see Dejonckheere et al., 2019; O'Toole et al., 2020; Seah & Coifman, 2021).

At the same time, ambulatory assessment studies of emotion often abandon or constrain the collection of information about events in which feelings, thoughts, and behaviors are embedded. This is due in part to a shift away from the open-ended responses that ambulatory assessment methods were initially intended to capture (e.g., Csikszentmihalyi & Larson, 1987) and toward closed-ended items such as scales and lists that facilitate quantification (Mehl, 2017; Myin-Germeys & Kuppens, 2022; for discussion, see Bringmann et al., revise & resubmit). Aspects of context – location, activity, social company, etc. – are frequently ignored, under the assumption that the emotion is paramount: where, when, with whom, and other details are theoretically incidental (e.g., Roseman, 2001; Tracy & Randles, 2011) or methodologically noisy. When information about context is collected, this is typically accomplished using lists of broad categories. For instance, categories for activities may include "working/studying," "exercising," and "eating" (for a review, see Stadel et al., 2025). Depending on study aims and design, participants may be forced to select (only) one response option and may not have a means of further specification (e.g., an "other" category to indicate a different activity or a separate item to indicate that they are with someone else). This approach necessarily limits the situation-, person-, and sample-specific diversity and nuance in the events that can be observed. While this limitation is widely acknowledged and a source of frustration among scientists (e.g., Mehl, 2017; Myin-Germeys & Kuppens, 2022), a solution has not yet been brought forward.

Identifying a solution is essential to advancing emotion science theoretically as well as empirically (Aldao, 2013; Greenaway et al., 2018). According to constructionist accounts of emotion, for example, instances of emotion are created by comparing prior experiences and accrued knowledge (i.e., concepts) to the current array of sensory and mental features from its physical and social environment (Barrett, 2017, 2022; Hoemann et al., 2017; Hoemann, Lee, et al., 2023; Mesquita & Boiger, 2014). Context, from this perspective, is inherent to emotion (as to all psychological phenomena). A full understanding of the process and result of emotion construction requires attention to the situations and relationships that people are constantly navigating. These building blocks of emotional life vary naturally within and across individuals (e.g., Lydon-Staley et al., 2020; Oishi et al., 2004; Vogel et al., 2017), creating the dynamic, person-specific distributions of experiences and knowledge that form concepts and give rise to sensation and behavior. Surveying the terrain in which emotion is embedded can shed light on how people come to have the patterns of experience that they do. It also provides an opportunity to evaluate theoretical claims about the nature of emotion categories like anger and amusement. If experiences of emotion are assembled online as idiographic concepts are applied to momentary circumstances, the result should be categories that are populated by highly heterogeneous instances (e.g., Hoemann et al., 2020) rather than reliably tied to specific contexts (e.g., Cowen et al., 2021).

One way of capturing the events that constitute emotion, while maximizing naturalism and variation, is through a return to open-ended reports of experience. Conceptually, these reports are better aligned with the intentions of ambulatory assessment (e.g., Csikszentmihalyi & Larson, 1987). Practically, technological advances mean that the processing and analysis of natural language data (e.g., text fields, diary entries, voice recordings) are more accessible and scalable than ever before (for reviews, see Berger & Packard, 2021; Feuerriegel et al., 2025; Hoemann et al., 2025; Jackson et al., 2021). Recent ambulatory assessment studies of emotion that have collected natural language data have found that emergent themes from everyday speech and diary entries provide insight, respectively, into experienced valence (Sun et al., 2019) and the degree of differentiation among negative emotions (Hoemann, Lee, et al., 2023). These themes were discovered using topic modeling – an unsupervised machine learning technique that automatically identifies and organizes key themes within large collections of texts by grouping words and texts based on their similarities (Blei et al., 2003; Günther et al., 2019; Wilson et al., 2016). In these studies, emergent themes – including 'movie night,' 'dinner with friends,' 'studying for exams,' and 'work meetings' – can be understood as the situations and activities encountered by participants over the course of the study, lending support to the idea that knowing more about the context of emotion will shed light on momentary experience and individual differences therein.

In this article, we build on this recent work with the aim of answering three descriptive questions. First: (Q1) What are the types of emotion-related events that populate daily life for participants in our sample? Second: (Q2a) How are these events distributed across persons in our sample – are some mentioned more often or by more people? (Q2b) Do people vary in how many different types of emotion-related events they mention, and how regularly do they mention them? And third: (Q3) How do these events map onto common categories for emotion, like pride, anger, sadness, and amusement?

To answer these questions, we used data from an innovative ESM study, in which a community sample of 102 Dutch-speaking Belgian adults completed a 10-week protocol. They received four prompts per day, at each one recording or typing a brief description of their current experience and how they felt about it. Participants provided up to 280 descriptions each, for a final total of nearly 25,000 texts for the sample. These data provide us with extensive group- and person-level corpora to examine the kinds of experiences organically reported and the emotion words spontaneously used to label these experiences. Using these data, we discovered the situations, activities, and events that were typical for our sample using a topic modeling approach that leverages state-of-the-art, transformer-based embedding models (Grootendorst, 2022). These models place texts closer in high-dimensional semantic space based on how their constituent words are used in context, enabling us to account for sentence structure and thematic content as well as word choice when discovering our topics.

We answered our first question (Q1) by inspecting the number and type of topics discovered and illustrating pertinent examples via word clouds. We answered our second research question by counting how often each topic was used per person and overall (Q2a), and by examining the range and evenness with which each person sampled from the events common to the group (Q2b). We answered our third question (Q3) by identifying high-frequency emotion words in our data and examining the number and type of topics each word was used in conjunction with.

Results

Daily Life Events

A topic modeling analysis of all texts revealed 139 highly granular themes. Figure 1 presents a subset of word clouds translated to English for visualization; full results are available via our OSF repository. Participants talked about everyday moments (column A), like ‘morning routines,’ ‘train commutes,’ and ‘exercising.’ They described work and school (column B): ‘studying,’ ‘sitting in meetings,’ and ‘applying for jobs.’ They discussed leisure activities (column C): ‘reading books,’ ‘watching TV,’ and ‘enjoying music.’ Participants also spoke about their relationships (column D): ‘playing games with friends,’ ‘having lunch with colleagues,’ and ‘parties with family.’ They mentioned mental and physical health concerns, as well (column E): ‘going to therapy,’ ‘headaches and colds,’ and ‘dental issues.’ References to salient events, like ‘Christmas,’ ‘vacations,’ and ‘medical procedures,’ were also among the topics (column F). These word clouds showcase the diverse range of meaningful topics that context-free approaches to sampling emotion in everyday life miss out on.

Figure 1. Example themes from a topic modeling analysis of all texts. Each column presents a set of themes related to: A) everyday moments; B) work and school; C) leisure activities; D) relationships; E) health concerns; F) salient events. Topic words have been translated from Dutch to English for presentation only.

Distribution Across and Within Participants

Examining topic use over the sample, we found that individual themes differed in how broadly and evenly they were mentioned. As can be seen in Figure 2, themes such as 'watching TV' were mentioned by all participants and heavily used by some (Panel A), whereas themes such as 'meetings' were brought up a few times by most participants (Panel B), and themes such as 'dental issues' were used regularly by fewer participants (Panel C). Topics were used on average 1,289 times (*SD* = 48; range: 6-12,816) by on average 65 (*SD* = 7; range: 1-102) of the 102 participants.

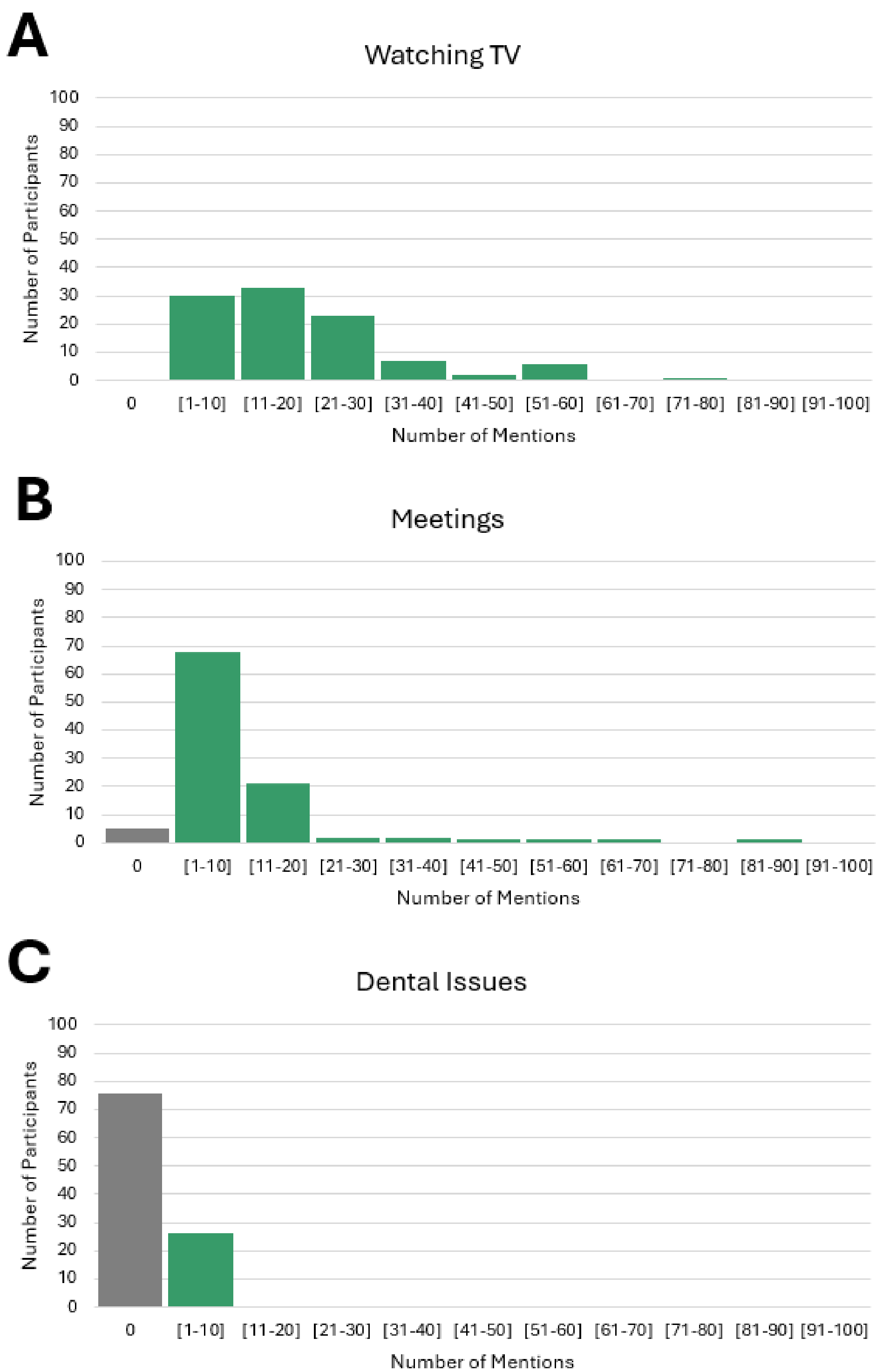

Figure 2. Distributions of use of example themes: A) 'watching TV'; B) 'meetings'; C) 'dental issues'.

We also found that participants varied in how many themes they used throughout the study and how often they did so. To quantify this variation, we calculated a Gini coefficient for each participant (following Benson et al., 2018), capturing the relative spread of topics across their descriptions of momentary experience (Hoemann, Lee, et al., 2023). As can be seen in Figure 3, participants varied in their topic diversity, with some regularly using many of the themes common to the sample (example participant in Panel A: 77 topics across 266 event descriptions; Gini coefficient = 0.21) and others using fewer themes more of the time (example participant in Panel B: 48 topics across 251 event descriptions; Gini coefficient = 0.09). On average, participants used 65 (*SD* = 7; range: 47-79) of the 139 themes (Gini coefficient *M* = 0.18, *SD* = 0.02).

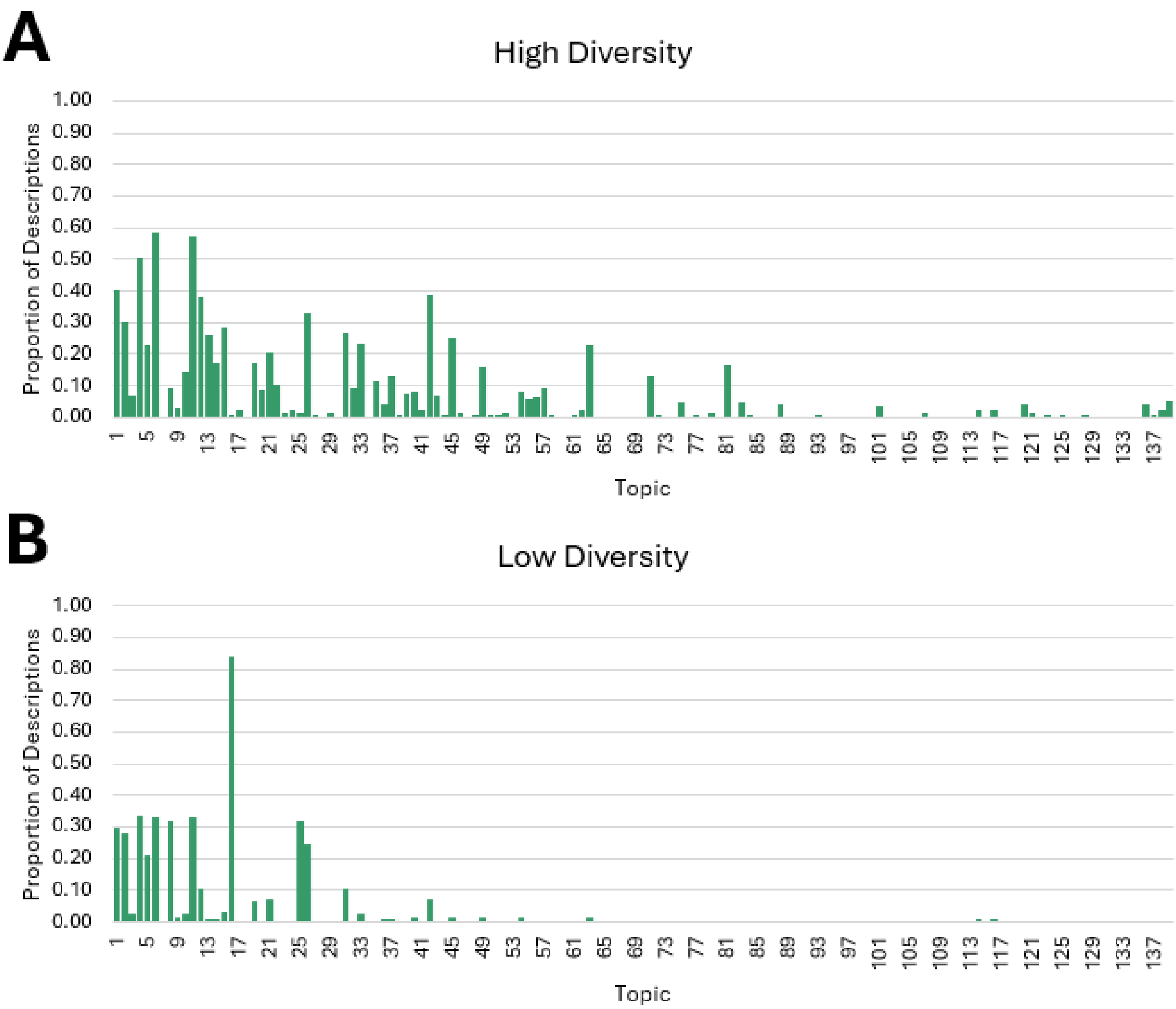

Figure 3. Example participants with high (A) and low (B) diversity in topic use, here depicted as the proportion of descriptions of momentary experience that mentioned a given topic.

Relationship with Categories for Emotion

A topic modeling analysis of texts containing 27 commonly-used emotion words yielded a set of 102 themes (see Methods for details of word selection). Words and themes exhibited a many-to-many relationship, with each word used to label events corresponding with multiple themes (*M* = 61, *SD* = 18; range: 5-98). Figure 4 illustrates these relationships for a subset of words and themes selected to show that this finding was not specific to pleasant (positive) or unpleasant (negative) emotions or to a particular domain of themes (e.g., everyday moments, relationships); for the full results, please see our OSF repository (https://osf.io/8mxs7/overview?view_only=c4abcc6156e5451fa560f76f7d276aea). For example, “proud” (“trots”) was associated with 67 topics, including ‘exercise’ and ‘exams.’ “Angry” (“boos”) was associated with 68 topics, including ‘playing games’ and ‘train commutes.’ “Sad” (“verdrietig”) was associated with 61 topics, including ‘reading books’ and ‘relationships.’ “Amused” (“geamuseerd”) was associated with 60 topics, including ‘Christmas’ and ‘movies.’ Conversely, individual themes were generally associated with labels for multiple emotion categories (*M* = 16, *SD* = 6; range: 1-26). For example, ‘reading books’ occurred in descriptions that mentioned sadness as well as those that mentioned amusement. ‘Exercise’ occurred alongside references to pride as well as to anger. These many-to-many relationships between words and themes also emerged in a topic modeling analysis of texts containing an expanded list of emotion words, suggesting that they were not due to the specific emotion categories selected. See Supplemental Materials for details.

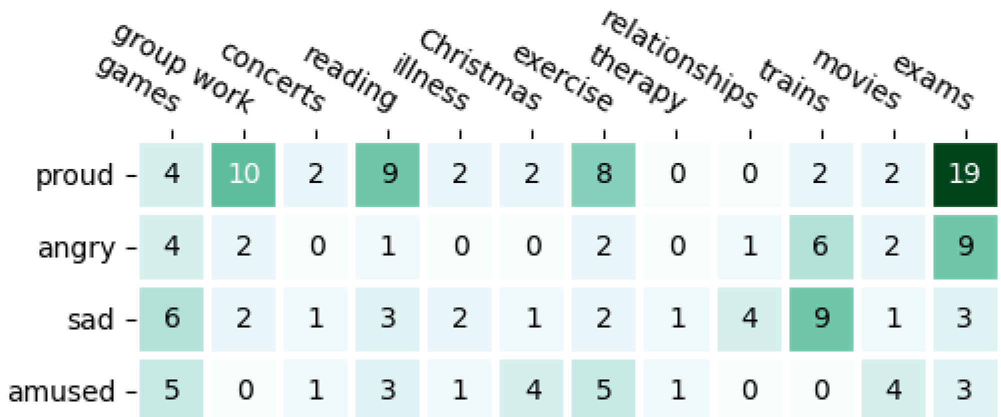

Figure 4. Example emotions and examples of their associated themes. Emotion words have been translated from Dutch to English for presentation only.

Discussion

Leveraging a unique experience sampling data set, in which participants verbally described their current experience and how they felt about it several times a day and for almost three months, we used topic modeling to investigate the diverse, personally-meaningful events at the heart of emotion. We first described the types of themes that populated daily life for our sample: ‘train commutes,’ ‘job applications,’ ‘reading books,’ ‘family parties,’ ‘headaches and colds,’ ‘vacation,’ and more. We then examined how these themes were distributed, finding differences in how broadly and evenly individual themes were mentioned as well as in how broadly and evenly participants sampled among the set of themes. Lastly, we evaluated the mapping between themes and categories for emotion that were frequently referenced. We found a many-to-many relationship between emotion words and themes, with each word used to label events corresponding with multiple themes and each theme associated with labels for multiple emotion categories.

The main takeaway from this work is perhaps intuitive: People navigate an extensive variety of situations, activities, and events in their everyday lives. Topic modeling applied to ambulatory assessment data appears to capture this intuition (e.g., Hoemann, Lee, et al., 2023; Sun et al., 2019). Variation in themes related to emotion has also been substantiated by topic modeling of social media posts (e.g., De Choudhury et al., 2017; Eichstaedt et al., 2018; Ramírez-Esparza et al.,

2008) and by analyses of event descriptions elicited through questionnaires (e.g., Davitz, 1969; Scherer et al., 1983; Wallbott & Scherer, 1988) and interviews (Hoemann, Gendron, et al., 2023; e.g., Hoemann, Şencan, et al., 2024; Mesquita, 2001). Nevertheless, studies that explicitly aim to document the specific contexts of emotion are rare (as discussed by e.g., Mehl, 2017; Myin-Germeys & Kuppens, 2022). This may be due to assumptions that emotions are inherently yoked to certain types of situations (e.g., Tracy & Randles, 2011) or that they are generated solely by situational evaluations (i.e., appraisals; e.g., Roseman, 2001; see also Tekoppele et al., 2023). In either case, details about the circumstances themselves are rendered less theoretically relevant. Rich contextual description in emotion science has also been hindered by methodological considerations, such as the need for large, open-response data sets and replicable ways of automating analysis (for discussion, see Bringmann et al., revise & resubmit; Hoemann et al., 2025; Stadel et al., 2025).

A second takeaway follows from the observation that themes were distributed differently within and across participants. This could make them a useful means of identifying atypical moments in a person or a group, or of identifying changes or deviations in patterns of experience. For example, the tendency to use particular themes (e.g., 'headaches and colds,' 'work stress') has been shown to relate to mental health (e.g., depression; Eichstaedt et al., 2018). Individual differences in the diversity of theme use have been shown to relate to emotional functioning, with people who experience or report on a restricted set of events in daily life also reporting less differentiated negative emotions (Hoemann, Lee, et al., 2023). As such, the present approach and findings can be used to support translational aims of emotion science. Interventions and clinical applications of emotion science cannot hope to provide tailored care to patients without idiographic approaches to tracking the contexts associated with symptoms and behaviors (Aldao et al., 2015; Bringmann, 2024; Klipstein et al., 2023; Piot et al., 2022). Knowing what someone is feeling, and the dynamics therein, is insufficient if the ultimate objective is to ascertain why they would feel that way and how this can be avoided or facilitated (as the case may be; e.g., Myin-Germeys et al., 2016; Snippe et al., 2016).

This is not to say, of course, that knowing the context of emotional experience is tantamount to knowing how someone is feeling. Indeed, a final takeaway from the present analyses is that themes did not have fixed relationships with emotion categories. This observation is consistent with constructionist accounts of emotion, in which categories are collections of instances that can vary widely because each is constructed in a situation-specific manner (Barrett, 2022; Hoemann et al., 2020; Mesquita & Boiger, 2014; Wilson-Mendenhall et al., 2015). This observation is also partially consistent with a recent findings from Kamiloğlu and colleagues (2025), in which participants were provided with definitions and examples of 22 positive emotions and asked to describe a personal experience for each. Topic modeling revealed 21 clusters of subjective experience that mapped onto the queried emotions with variable consistency. For example, amusement was strongly associated with the theme of 'humor,' interest with 'curiosity,' and tenderness with 'nurturing'; whereas elation, inspiration, and triumph (among others) were weakly associated with multiple themes. The difference between these findings and the present may be attributed to how descriptions were elicited (e.g., a single prompted exemplar vs multiple unprompted per person). Ultimately, however, both works illustrate flexibility in the context-to-emotion mapping.

In interpreting the present findings, there are several notable considerations to bear in mind. Chief among them is that the results are not generalizable beyond the data set analyzed. This is by design, as it were: Topic modeling is a data-driven approach to recovering themes from natural language, and so the topics that emerge are not prespecified (for discussion, see Schwartz et al., 2013; Sun et al., 2019). Likewise, the situations, activities, and events captured by ambulatory assessment will differ by person, sample, culture, and even time of year (Kuppens et al., 2022). While this makes the current approach useful for investigating person- and sample-specific diversity in emotion-related events, it challenges direct comparison of thematic content across samples. Another consideration, relevant to our third set of findings, is that emotion words do not necessarily index a person's own experience of emotion. Although we selected the emotion words conservatively (see Methods for details), it is still possible that participants were talking about other people ("he was happy") or in general ("to be happy"), that they negated the word ("not happy") or used it with a non-emotional meaning ("a happy coincidence"; for discussion, see Boyd & Schwartz, 2021). More contextualized tagging of emotion words, or another method of accessing subjective experience, would be necessary to distinguish between these various use cases. Equally, emotions can be indexed in other ways, beyond simple labels. A fuller scope of linguistic, subjective, and contextual features (e.g., vocal prosody, situational evaluations, peripheral physiology) can be used in future research to discover the context-specific patterns of experience that correspond with daily life events (Hoemann et al., 2025; Hoemann, Warfel, et al., 2024).

Despite these considerations, the present study makes important contributions to the science of emotion and of human behavior more broadly. Empirically, our findings give voice to the fact that emotion – as any psychological phenomenon – is fundamentally embedded in context. Methodologically, pairing open-ended, verbal descriptions of momentary experience with data-driven methods of discovering patterns in these data can answer core theoretical questions about the construction of experience, while minimizing the constraints imposed by researcher expectations. This general approach can be extended to other cultural and functional contexts to gain insight into the situational shaping of feelings, thoughts, behaviors, and interactions, and how these give rise over time to relevant (inter)personal outcomes such as mental, physical, and relational health (e.g., Hurlburt et al., 2002, 2021; Kaplan et al., 2022, 2025). The physical, social, and psychological environment in which experiences of emotion unfold is a complex and dynamic ecosystem. Ultimately, we see this approach as a pathway toward studying this ecosystem in everyday life.

Methods

Participants

Participants were a community sample of Dutch-speaking Belgian adults recruited through flyers, online posts, and word-of-mouth. To be eligible, participants needed to be native Dutch speakers, at least 18 years old, living in Belgium, and in possession of a smartphone in good condition. Interested individuals were directed to an eligibility survey, after which these criteria were verified during an online introduction session.

Participants were compensated up to €250 for completing a 70-day (10-week) experience sampling protocol including biweekly online surveys. They earned €0.50 for each completed experience sampling prompt (4 prompts x 70 days = 280 prompts or €140 max), €10 for each of 4 short online surveys (2, 4, 6, and 8 weeks; €40 max), and €15 for each of 2 long online surveys (0 and 10 weeks, €30 max). Participants who remained in the study for at least 60 days received a bonus of €40. Participants were remunerated once, in full, at the end of the study. To remain in the study, participants must have completed 75% of prompts on average, and the verbal descriptions of experience gathered via these prompts must have been at least 25 words long. Periodic compliance checks and summary reports were used to inform participants of their progress, alert those who needed to increase their engagement, and determine if participants needed to be dismissed.

A total of 115 participants enrolled in the study (age: 18-65, M = 27.26, SD = 9.86; gender: 58 women, 56 men, 1 other). Of these, 10 chose to leave the study after partial completion and 3 were dismissed for poor compliance (i.e., response rate under 50%). The remaining 102 participants ranged in age from 18 to 65 (*M* = 26.47, *SD* = 8.87) and comprised 52 women, 49 men, and 1 other.

Study procedure and materials were reviewed and approved by the KU Leuven Social and Societal Ethics Committee (SMEC), protocol G-2023-6379-R3(AMD). Data collection occurred from August 2023 through July 2024. All study instructions and materials were presented in Dutch.

Procedure and Materials

Participants completed 70 days of experience sampling, during which they received 4 prompts per day via a dedicated smartphone app (m-Path; Mestdagh et al., 2023). Prompts were sent at pseudorandom times between 9 am and 9 pm with at least one hour between each prompt. At each prompt, participants responded to the following question (in Dutch): “What is going on now or since the last prompt, and how do you feel about it?” (“Wat speelt er nu of sinds de vorige beep, en hoe voel je je daarover?”). They were asked to record a voice message of about 1 minute into the app but could choose to type a 3-4-sentence response instead. Participants also rated their current valence and arousal, while phone sensors were used to register passive sensing data (e.g., GPS coordinates). These latter data are not analyzed in the present report.

The day prior to experience sampling, participants attended a brief online introduction session during which they provided informed consent and basic demographic information, installed m-Path, and received study instructions. At the end of the session, they received an email with the link to an online survey including two measures of well-being and three measures of emotional functioning that are not analyzed in the present report. Participants repeated a subset of these measures after 2, 4, 6, and 8 weeks of experience sampling. After 10 weeks (i.e., after their last day), participants again received an email with the link to an online survey including all measures, as well as additional demographic information.

During experience sampling, participants received periodic reports summarizing their participation to date. These reports were generated every other Monday (i.e., every 14 days) for all currently enrolled participants and sent as PDFs attached to an email from the study account. The summary report included: the cumulative percentage of prompts completed (along with a reminder about compliance) and the amount of money earned for these prompts; the cumulative percentage of descriptions that were sufficiently long; the cumulative percentage of descriptions recorded versus typed (not assessed for compliance); a graph of valence and arousal ratings over time; a graph of high-level description contents over time (generated via Linguistic Inquiry and Word Count [LIWC] using the 2015 Dutch translation; Van Wissen & Boot, 2017). Participants who had recently completed the study protocol additionally received information about the number of online surveys completed and the amount of money earned for these, whether they had earned the bonus for completing more than 60 study days, and their total amount earned.

Data Preparation

Recorded descriptions were automatically transcribed using a Dutch automatic speech recognition (ASR) system developed in-house (Tamm et al., 2024). This model was trained on approximately 270 hours of Flemish Dutch speech from the Spoken Dutch Corpus (Oostdijk, 2000), following protocols similar to the baseline described by Poncelet and Van hamme (2023). The resulting transcripts contain system-generated punctuation and adhere to the orthographic transcription conventions of the Spoken Dutch Corpus, which include detailed annotation tags for foreign words, dialect, disfluencies, and other phenomena. Subsequent preprocessing removed these annotation tags and non-content markers, retaining only the core textual content for further analysis (see below for details). These transcriptions were integrated into a master experience sampling data file including the typed descriptions. Participants provided a total of 25,151 descriptions, 10,362 (41%) of which were typed. Texts shorter than 15 words long were dropped prior to analysis, for a final total of 24,852.

Data Cleaning and Preprocessing

Descriptions of momentary experience were preprocessed by a custom pipeline in Python (version 3.9.13) using the natural language processing packages NLTK (Loper & Bird, 2002) and Stanza (Qi et al., 2020). We first removed the annotation tags generated by the automatic transcription, including markers for foreign words (*v), dialect or accented words (*d and *z), slips of the tongue (*u), and other placeholders. For words cut short (*a) and those with uncertain transcription (*x), we additionally removed the preceding word or string. Some participants chose to begin their recordings by greeting the researcher (author K.H.; e.g., "Good morning, Katie."). These strings were also removed.

Each cleaned description was then divided into sentences and stripped of punctuation, with each sentence further divided into words and lemmatized. Lemmatization is the process of reducing a word to its root form (i.e., a lemma) based on its intended meaning in context, such as when "meeting" becomes "meet" when used as a verb (e.g., "we are meeting tomorrow") but remains "meeting" when used as a noun (e.g., "our meeting is tomorrow"). Here, lemmatization was used to remove inflected word forms from the descriptions, resulting in a base set of 'core', unique meanings for analysis (Qi et al., 2020).

Evident typos and lemmatization errors were manually identified by reviewing a list of all lemmatized word forms occurring at least twice across texts. These typos and errors were entered into a list of conversions applied to the data. During this review process, common closed-class or 'stop' words (e.g., articles, auxiliary verbs, prepositions, pronouns) were also identified. Using stop lists to exclude generic or non-content words helps reduce noise, enhance topic specificity, and improve interpretability, as these common words occur frequently across many contexts and do not contribute meaningful information about the unique content of the topics. We validated and augmented our Dutch stop word list by merging it with several publicly available lists (e.g., https://github.com/stopwords-iso/stopwords-nl). Final conversion and stop word lists are available via our OSF repository (https://osf.io/8mxs7/overview?view_only=c4abcc6156e5451fa560f76f7d276aea) .

Emotion Words

The emotion words used by participants were also identified through manual review. First, all raw, inflected word forms were extracted along with their frequency of use. Then, words used at least 10 times in the data were coded by author K.H. and trained Dutch-speaking job students for whether they represented an emotion (e.g., "blij" ["happy"], "boos" ["angry"], "fier" ["proud"], "zenuwachtig" ["nervous"]). 'Emotion' was defined narrowly using only specific (e.g., "blij" ["happy"]) and not generic states (e.g., "goed" ["good"]). Bodily states (e.g., "moe" ["tired"]), cognitive states (e.g., "nieuwsgierig" ["curious"]), demeanors (e.g., "verlegen" ["shy"]), and circumstances (e.g., "achtergelaten" ["abandoned"]) were not considered. Emotion words also had to be adjectives (e.g., "geïrriteerd" ["irritated"; adjective]), with other parts of speech ignored (e.g., "irritatie" ["irritation"; noun], "irriteren" ["to irritate"; verb]). Equally, emotion words had to refer in principle to the state of the self (e.g., "geïrriteerd" ["irritated"]) rather than to the experience of a situation (e.g., "irriterend" ["irritating"]). Words in other languages (e.g., "happy") and contemporary vernacular (e.g., "chill") were retained. This data-driven process yielded 80 emotion words used at least 10 times in the data, and 27 used at least 100 times. Lastly, words were lemmatized to reduce inflected forms (e.g., "blij" ["happy"] vs. "blijer" ["happier"]). These word lists were used as input into a custom Python script that was used to compute the length of each description of momentary experience and to count the number of times each emotion lemma was used.

Topic Modeling

We employed the transformer-based embedding model BERTopic to extract and analyze themes from our Dutch textual data (Figure 5). The topic modeling pipeline consisted of three main steps: 1) vectorization, 2) dimensionality reduction, and 3) clustering (see Figure 5). First, we evaluated three multilingual sentence-level embedding models to identify the optimal model for processing Flemish (see Supplemental Material for more details). We ultimately selected the “jinaai/jina-embeddings-v3” model (Sturua et al., 2024) via the “SentenceTransformers” library (Reimers & Gurevych, 2019) to generate dense vector representations (embeddings) of the text. This model, available publicly on Hugging Face, is optimized for multilingual and domain-specific embedding tasks, making it suitable for Dutch text analysis. Second, given the high-dimensional nature of text embeddings (i.e. the numerical/vector representations of the text), we applied the neighbor graph based Uniform Manifold Approximation and Projection (UMAP; McInnes et al., 2018) to reduce the embeddings to a lower-dimensional space while preserving the semantic relationships between texts. UMAP constructs a representation of the data that reflects the closeness of similar texts, allowing semantically related responses to group together in a lower-dimensional feature space. We selected UMAP due to its superior performance in preserving both local and global structure in the data, which has been shown to produce higher-quality clustering results (Kamiloğlu et al., 2025). We used cosine distance (i.e., one minus the cosine similarity) as the distance metric between embeddings. We systematically evaluated multiple parameter configurations to optimize the balance between preserving local neighborhood structure and maintaining global data relationships. Specifically, we tested different numbers of neighbors (10, 15, 20, 25, and 30) and minimum distance values (0.0 and 0.1). This dimensionality reduction enabled us to cluster texts with similar semantic content, thereby identifying coherent themes within the corpus.

Third, following dimensionality reduction, we clustered the embeddings to group texts with similar semantic features into coherent topics. We employed the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm (Campello et al., 2013), which identifies clusters based on regions of high density in the data space. HDBSCAN offers several advantages for text analysis: it automatically determines the optimal number and size of clusters based on the data structure, adapts to clusters of varying densities, and identifies outliers (texts that do not fit well into any cluster, labeled as "-1") rather than forcing all data points into a cluster. We systematically evaluated different minimum cluster size parameters (5, 10, and 15), which determine the smallest number of texts required to form a distinct topic. Additionally, we tested minimum topic size parameters (5, 10, and 15). For the number of topics (option “nr_topics”), we evaluated both automatic determination (“auto”: allowing the algorithm to identify the optimal number based on data structure) and a predetermined value of 150 topics. Based on this evaluation, we set the minimum cluster size to 5, which balanced the need for meaningful topic interpretation with sufficient granularity to capture nuanced thematic distinctions. This parameter choice also enabled soft clustering, retaining probabilistic assignment information that allowed for subsequent refinement of topic boundaries.

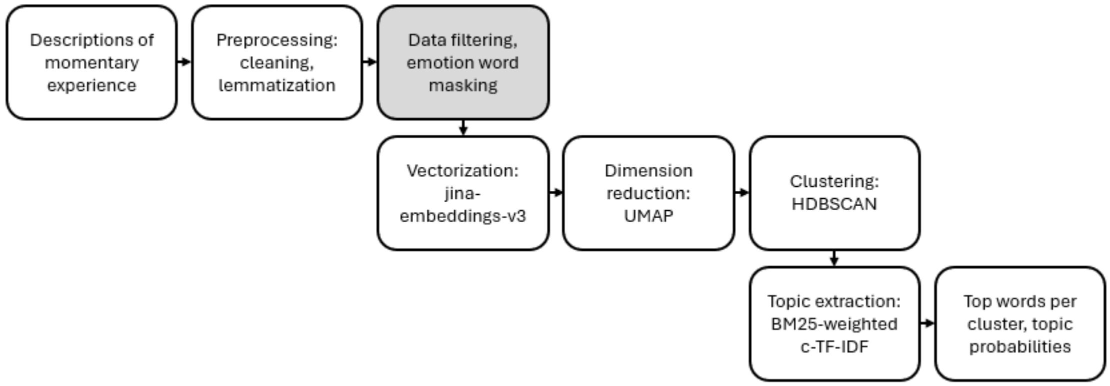

Figure 5. Schematic of the BERTopic pipeline. Data filtering and emotion word masking step (gray box) only applied to assess the relationship between events and categories of emotion (third research question).

Topic Extraction and Evaluation

We extracted the topics from the clustering solution using BM25-weighted c-TF-IDF (Grootendorst, 2022). This strategy produces interpretable and semantically coherent topics by dynamically weighting terms for their importance by considering their frequency within a topic relative to their corpus-wide distribution. For example, if the term “rain” appeared 120 times within a topic but only in 30 out of 24,852 documents, it would receive a high weight, indicating that it is particularly significant for that topic. That is, the term’s local frequency within the cluster across the corpus (here, an incidence of just 0.12%) increases its weight as a distinctive topic identifier.

We further generated word clouds for each topic using the term importance scores provided by BERTopic’s get_topic() method, which returns the weighted representation of each term within the topic. We extracted the top 10 words and their associated importance scores (based on BM25-weighted c-TF-IDF) and visualized them using the WordCloud library, where word size reflects terms’ relative importance within the topic. These word clouds provide an interpretable visual summary of the most distinctive terms in each cluster.

We evaluated model performance by assessing topic coherence and diversity. Coherence (C_v; Röder et al., 2015) is a metric that represents how well the keywords in a topic stick together by checking: how often words appear together (by using normalized pointwise mutual information), how similar they are (using cosine similarity), and how these patterns hold up in small chunks of text (using a sliding window). Additionally, we assessed topic diversity by comparing the number of unique words in each topic to the total number of words, ensuring that each topic is distinct and not overly redundant.

We also evaluated model performance qualitatively, by holistically examining the interpretability, uniqueness, and representativeness of the extracted topics. Based on our knowledge of the data and the contexts of data collection (e.g., sample characteristics, locations), we assessed whether probable activities and situations were captured by the topics and at a useful level of granularity. For example, a model solution with only 3 topics would not provide enough of a range of topics to give insight into the events that populate our sample’s daily life. Equally, a solution with many overlapping or participant-specific topics would not illustrate the distinct themes common to the group.

Data Filtering and Emotion Word Masking

To answer our first and second questions – what kinds of events populated the daily lives of our sample, and how were they distributed across and within participants – we ran an initial topic model on all 24,852 texts. For our third research question – how do events correspond with emotions – we filtered the data to only include those texts that contained at least one of the emotion words identified (see above for details). This produced two subsets: one for the emotion words occurring at least 10 times in the data (11,705 texts) and one for the emotion words occurring at least 100 times in the data (8,302 texts). Emotion words were masked out of the texts prior to topic modeling, allowing us to examine the intersection between the emotion categories being referenced (i.e., emotion words used) and the types of events being described (i.e., topics), without the former directly influencing the latter (following Kamiloğlu et al., 2025).

Extraction of Variables

For all topic models, we extracted the highest-loading topic for each text (i.e., the topic with the highest probability score in the document-topic probability matrix) as well as the continuous probability [0,1] that each text loaded onto each topic. To account for the fact that texts often included content corresponding to multiple themes or events, we considered a topic to be mentioned if its loading probability surpassed a threshold of 0.01 (i.e., binarizing the data), with this threshold established based on an examination of the distribution of loading probabilities (see Supplemental Materials for details). A custom Python script was then used to count the number of times each topic was mentioned by each participant, and to calculate a Gini coefficient for each participant based on these counts (following Benson et al., 2018). We visualized distributions of topics across and within participants using histograms.

References

Aldao, A. (2013). The future of emotion regulation research: Capturing context. *Perspectives on Psychological Science*, *8*(2), 155–172. https://doi.org/10.1177/1745691612459518

Aldao, A., Sheppes, G., & Gross, J. J. (2015). Emotion regulation flexibility. *Cognitive Therapy and Research*, *39*(3), 263–278. https://doi.org/10.1007/s10608-014-9662-4

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

Barrett, L. F. (2022). Context reconsidered: Complex signal ensembles, relational meaning, and population thinking in psychological science. *American Psychologist*, *77*(8), 894–920. https://doi.org/10.1037/amp0001054

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

Berger, J., & Packard, G. (2021). Using natural language processing to understand people and culture. *American Psychologist*. https://doi.org/10.1037/amp0000882

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.

Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, *40*(1), 21–41. https://doi.org/10/ghhzjd

Bringmann, L. F. (2024). The future of dynamic networks in research and clinical practice. *World Psychiatry*, *23*(2), 288–289. https://doi.org/10.1002/wps.21209

Bringmann, L. F., Gudmundsdottir, G. R., Schorlepp, L., Vogelsmeier, L. V. D. E., De Smet, M. M., Vrijen, C., Van Roekel, E., Fritz, J., Chen, J. X., Kirtley, O., Eronen, M. I., Chung, J. M., Stadel, M., & Hoemann, K. (revise & resubmit). *Beyond numbers: Revisiting the role of open-response data in experience sampling methodology*.

Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, *7819*, 160–172. https://doi.org/10.1007/978-3-642-37456-2_14

Conner, T. S., & Mehl, M. R. (2015). Ambulatory assessment: Methods for studying everyday life. In R. Scott, S. Kosslyn, & N. Pinkerton (Eds.), *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource* (pp. 1–15). Wiley.

Cowen, A. S., Keltner, D., Schroff, F., Jou, B., Adam, H., & Prasad, G. (2021). Sixteen facial expressions occur in similar contexts worldwide. *Nature*, *589*(7841), 251–257. https://doi.org/10/ghp7fg

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, *175*, 526–536.

Davitz, J. R. (1969). *The language of emotion*. Academic Press.

De Choudhury, M., Sharma, S. S., Logar, T., Eekhout, W., & Nielsen, R. C. (2017). Gender and cross-cultural differences in social media disclosures of mental illness. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 353–369. https://doi.org/10/gmzfzm

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

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records.

Proceedings of the National Academy of Sciences, *115*(44), 11203–11208. https://doi.org/10.1073/pnas.1802331115

Feuerriegel, S., Maarouf, A., Bär, D., Geissler, D., Schweisthal, J., Pröllochs, N., Robertson, C. E., Rathje, S., Hartmann, J., Mohammad, S. M., Netzer, O., Siegel, A. A., Plank, B., & Van Bavel, J. J. (2025). Using natural language processing to analyse text data in behavioural science. *Nature Reviews Psychology*, *4*(2), 96–111. https://doi.org/10.1038/s44159-024-00392-z

Fritz, J., Piccirillo, M. L., Cohen, Z. D., Frumkin, M., Kirtley, O., Moeller, J., Neubauer, A. B., Norris, L. A., Schuurman, N. K., & Snippe, E. (2024). So you want to do ESM? 10 essential topics for implementing the experience-sampling method. *Advances in Methods and Practices in Psychological Science*, *7*(3), 25152459241267912. https://doi.org/10.1177/25152459241267912

Greenaway, K. H., Kalokerinos, E. K., & Williams, L. A. (2018). Context is everything (in emotion research). *Social and Personality Psychology Compass*, *12*(6), e12393. https://doi.org/10.1111/spc3.12393

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (No. abs/2203.05794). arXiv.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Hoemann, K., Gendron, M., & Barrett, L. F. (2017). Mixed emotions in the predictive brain. *Current Opinion in Behavioral Sciences*, *15*, 51–57. https://doi.org/10.1016/j.cobeha.2017.05.013

Hoemann, K., Gendron, M., Crittenden, A. N., Mangola, S. M., Endeko, E. S., Dussault, È., Barrett, L. F., & Mesquita, B. (2023). What we can learn about emotion by talking with the Hadza. *Perspectives on Psychological Science*. https://doi.org/10.1177/17456916231178555

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., Lee, Y., Dussault, È., Devylder, S., Ungar, L. H., Geeraerts, D., & Mesquita, B. G. de. (2025). The construction of emotional meaning in language. *Communications Psychology*, *3*, 99. https://doi.org/10.1038/s44271-025-00255-0

Hoemann, K., Lee, Y., Kuppens, P., Gendron, M., & Boyd, R. L. (2023). Emotional granularity is associated with daily experiential diversity. *Affective Science*. https://doi.org/10.1007/s42761-023-00185-2

Hoemann, K., Şencan, R. S., Cochez, A., Beckers, C., & Mesquita, B. G. de. (2024). *Cultural models of emotion manifest in descriptions of everyday experience: A case study of the US and Belgium*. OSF. https://osf.io/mtd4a

Hoemann, K., Warfel, E., Mills, C., Allen, L., Kuppens, P., & Wormwood, J. B. (2024). Using freely generated labels instead of rating scales to assess emotion in everyday life. *Assessment*, 10731911241283623. https://doi.org/10.1177/10731911241283623

Hurlburt, R. T., Heavey, C. L., Lapping-Carr, L., Krumm, A. E., Moynihan, S. A., Kaneshiro, C., Brouwers, V. P., Turner, D. K., & Kelsey, J. M. (2021). Measuring the frequency of inner-experience characteristics. *Perspectives on Psychological Science*, 1745691621990379. https://doi.org/10/gmg8ff

Hurlburt, R. T., Koch, M., & Heavey, C. L. (2002). Descriptive experience sampling demonstrates the connection of thinking to externally observable behavior. *Cognitive Therapy and Research*, *26*(1), 117–134.

Jackson, J. C., Watts, J., List, J.-M., Puryear, C., Drabble, R., & Lindquist, K. A. (2021). From text to thought: How analyzing language can advance psychological science. *Perspectives on Psychological Science*, 22.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The Day Reconstruction Method. *Science*, *306*, 1776–1780. https://doi.org/10.1126/science.1103572

Kamiloğlu, R. G., Türkmen, İ. U., Sarnıç, T. E., Landman, D., & Sauter, D. (2025). What makes us feel good? A data-driven investigation of positive emotion experience. *Emotion*, *25*(1), 271–276. https://doi.org/10.1037/emo0001417

Kaplan, D. M., Arconada Alvarez, S. J., Palitsky, R., Choi, H., Clifford, G. D., Crozier, M., Dunlop, B. W., Grant, G. H., Greenleaf, M. N., Johnson, L. M., Maples-Keller, J., Levin-Aspenson, H. F., Mascaro, J. S., McDowall, A., Pozzo, N. S., Raison, C. L., Rothbaum, B. O., Zarrabi, A. J., & Lam, W. A. (2025). *Fabla: A voice-based ecological assessment method for securely collecting spoken responses to researcher questions*. https://doi.org/10.31234/osf.io/qd3um_v1

Kaplan, D. M., Hughes, C. D., Schatten, H. T., Mehl, M. R., Armey, M. F., & Nugent, N. R. (2022). Emotional change in its "natural habitat": Measuring everyday emotion regulation with passive and active ambulatory assessment methods. *Journal of Psychotherapy Integration*.

Klipstein, L. von, Stadel, M., Bos, F., Bringmann, L., Riese, H., & Servaas, M. (2023). *Opening the contextual black box: A case for idiographic experience sampling of context for clinical applications*. PsyArXiv. https://doi.org/10.31234/osf.io/nv7c6

Kuppens, P., Dejonckheere, E., Kalokerinos, E. K., & Koval, P. (2022). Some recommendations on the use of daily life methods in affective science. *Affective Science*, *3*(2), 505–515. https://doi.org/10.1007/s42761-022-00101-0

Loper, E., & Bird, S. (2002). *NLTK: The natural language toolkit* (No. cs/0205028). arXiv. https://doi.org/10.3115/1118108.1118117

Lydon-Staley, D. M., Zurn, P., & Bassett, D. S. (2020). Within-person variability in curiosity during daily life and associations with well-being. *Journal of Personality*, *88*(4), 625–641. https://doi.org/10.1111/jopy.12515

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform manifold approximation and projection. *Journal of Open Source Software*, *3*(29), 861. https://doi.org/10.21105/joss.00861

Mehl, M. R. (2017). The Electronically Activated Recorder (EAR): A method for the naturalistic observation of daily social behavior. *Current Directions in Psychological Science*, *26*(2), 184–190. https://doi.org/10/f92zdf

Mesquita, B. (2001). Emotions in collectivist and individualist contexts. *Journal of Personality and Social Psychology*, *80*(1), 68–74. https://doi.org/10.1037/0022-3514.80.1.68

Mesquita, B., & Boiger, M. (2014). Emotions in context: A sociodynamic model of emotions. *Emotion Review*, *6*(4), 298–302. https://doi.org/10/cvjn

Mestdagh, M., Verdonck, S., Piot, M., Niemeijer, K., Kilani, G., Tuerlinckx, F., Kuppens, P., & Dejonckheere, E. (2023). m-Path: An easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. *Frontiers in Digital Health*, *5*. https://doi.org/10.3389/fdgth.2023.1182175

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

Myin-Germeys, I., Klippel, A., Steinhart, H., & Reininghaus, U. (2016). Ecological momentary interventions in psychiatry. *Current Opinion in Psychiatry*, *29*(4), 258–263. https://doi.org/10/gbpg28

Myin-Germeys, I., & Kuppens, P. (Eds.). (2022). *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (2nd ed.). Center for Research on Experience Sampling and Ambulatory Methods Leuven.

Oishi, S., Diener, C. N. S., & Biswas-Diener, R. (2004). Cross-situational consistency of affective experiences across cultures. *Journal of Personality and Social Psychology*, *86*(3), 460–472. https://doi.org/10.1037/0022-3514.86.3.460

Oostdijk, N. (2000). The Spoken Dutch Corpus. Overview and first evaluation. *Proceedings of LREC-2000, Athens*, *2*, 887.

O'Toole, M. S., Renna, Megan. 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

Piot, M., Mestdagh, M., Riese, H., Weermeijer, J., Brouwer, J. M. A., Kuppens, P., Dejonckheere, E., & Bos, F. M. (2022). Practitioner and researcher perspectives on the utility of ecological momentary assessment in mental health care: A survey study. *Internet Interventions*, *30*, 100575. https://doi.org/10.1016/j.invent.2022.100575

Poncelet, J., & Van Hamme, H. (2023). Learning to jointly transcribe and subtitle for end-to-end spontaneous speech recognition. *2022 IEEE Spoken Language Technology Workshop (SLT)*, 182–189.

Pond, R. S., Jr., Kashdan, T. B., DeWall, C. N., Savostyanova, A., Lambert, N. M., & Fincham, F. D. (2012). Emotion differentiation moderates aggressive tendencies in angry people: A daily diary analysis. *Emotion*, *12*(2), 326–337. https://doi.org/10.1037/a0025762

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). *Stanza: A Python natural language processing toolkit for many human languages* (No. arXiv:2003.07082). arXiv. http://arxiv.org/abs/2003.07082

Ramírez-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.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 3982–3992. https://doi.org/10.18653/v1/D19-1410

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934–960. https://doi.org/10.1037//0033-2909.128.6.934

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408. https://doi.org/10.1145/2684822.2685324

Roseman, I. J. (2001). A model of appraisal in the emotion system. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 68–91). Oxford University Press.

Scherer, K. R., Summerfield, A. B., & Wallbott, H. G. (1983). Cross-national research on antecedents and components of emotion: A progress report. *Social Science Information*, *22*(3), 355–385. https://doi.org/10/cvfbnw

Schneider, S., Junghaenel, D. U., Gutsche, T., Mak, H. W., & Stone, A. A. (2020). Comparability of emotion dynamics derived from ecological momentary assessments, daily diaries, and the day reconstruction method: Observational study. *Journal of Medical Internet Research*, *22*(9), e19201. https://doi.org/10/ghf465

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E. P., & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, *8*(9), e73791. https://doi.org/10.1371/journal.pone.0073791

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

Snippe, E., Simons, C., Hartmann, J., Menne-Lothmann, C., Kramer, I., Booij, S., Viechtbauer, W., Delespaul, P., Myin-Germeys, I., & Wichers, M. (2016). Change in daily life behaviors and depression: Within-person and between-person associations. *Health Psychology*, *35*, 433–441. https://doi.org/10.1037/hea0000312

Snippe, E., Smit, A. C., Kuppens, P., Burger, H., & Ceulemans, E. (2023). Recurrence of depression can be foreseen by monitoring mental states with statistical process control. *Journal of Psychopathology and Clinical Science*, *132*, 145–155. https://doi.org/10.1037/abn0000812

Stadel, M., Langener, A. M., Hoemann, K., & Bringmann, L. F. (2025). Assessing daily life activities with experience sampling methodology (ESM): Scoring predefined categories or qualitative analysis of open-ended responses? *Methods in Psychology*, *12*, 100177. https://doi.org/10.1016/j.metip.2025.100177

Stone, A. A., Schwartz, J. E., Schkade, D., Schwarz, N., Krueger, A., & Kahneman, D. (2006). A population approach to the study of emotion: Diurnal rhythms of a working day examined with the Day Reconstruction Method. *Emotion*, *6*(1), 139–149. https://doi.org/10.1037/1528-3542.6.1.139

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, *16*(3), 199–202.

Sturua, S., Mohr, I., Akram, M. K., Günther, M., Wang, B., Krimmel, M., Wang, F., Mastrapas, G., Koukounas, A., & Wang, N. (2024). *jina-embeddings-v3: Multilingual embeddings with task LoRA* (No. arXiv:2409.10173). arXiv.

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.

Tamm, B., Poncelet, J., Barberis, M., & Vandermosten, M. (2024). *Weakly supervised training improves Flemish ASR of non-standard speech*. Dutch Speech Tech Day 2024, Location: Netherlands Institute for Sound & Vision, Hilversum.

Tekoppele, J. L., De Hooge, I. E., & van Trijp, H. C. M. (2023). We've got a situation here! – How situation-perception dimensions and appraisal dimensions of emotion overlap. *Personality and Individual Differences*, *200*, 111878. https://doi.org/10.1016/j.paid.2022.111878

Tomko, R. L., Lane, S. P., Pronove, L. M., Treloar, H. R., Brown, W. C., Solhan, M. B., Wood, P. K., & Trull, T. J. (2015). Undifferentiated negative affect and impulsivity in borderline personality and depressive disorders: A momentary perspective. *Journal of Abnormal Psychology*, *124*(3), 740–753. https://doi.org/10.1037/abn0000064

Tracy, J. L., & Randles, D. (2011). Four models of basic emotions: A review of Ekman and Cordaro, Izard, Levenson, and Panksepp and Watt. *Emotion Review*, *3*(4), 397–405.

Van Wissen, L., & Boot, P. (2017). An electronic translation of the LIWC Dictionary into Dutch. *Electronic Lexicography in the 21st Century: Proceedings of eLex 2017 Conference*, 703–715.

Vogel, N., Ram, N., Conroy, D. E., Pincus, A. L., & Gerstorf, D. (2017). How the social ecology and social situation shape individuals' affect valence and arousal. *Emotion*, *17*(3), 509–527. https://doi.org/10.1037/emo0000244

Wallbott, H. G., & Scherer, K. R. (1988). How universal and specific is emotional experience? Evidence from 27 countries on five continents. In K. R. Scherer (Ed.), *Facets of emotion: Recent research* (pp. 31–56). Lawrence Erlbaum Associates.

Wilhelm, F. H., & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*, *84*(3), 552–569. https://doi.org/10.1016/j.biopsycho.2010.01.017

Wilson, S., Mihalcea, R., Boyd, R., & Pennebaker, J. W. (2016). Disentangling topic models: A cross-cultural analysis of personal values through words. *Proceedings of the First Workshop on NLP and Computational Social Science*, 143–152. https://doi.org/10.18653/v1/W16-5619

Wilson-Mendenhall, C. D., Barrett, L. F., & Barsalou, L. W. (2015). Variety in emotional life: Within-category typicality of emotional experiences is associated with neural activity in large-scale brain networks. *Social Cognitive and Affective Neuroscience*, *10*(1), 62–71. https://doi.org/10.1093/scan/nsu037

Wrzus, C., & Neubauer, A. B. (2022). Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment*, *30*(3), 825–846. https://doi.org/10.1177/10731911211067538

Author Notes

We would like to thank Lena Dognaux, Lies Raveel, and Ewald Verfaillie for their assistance with data processing.

K.H. and data collection were supported by the Research Foundation – Flanders (FWO 12A3923N). P. K. was supported by KU Leuven Research Council grant C14/23/062 and VLIR funding iBOF/21/090. B.M. was supported by the European Research Council (Advanced Grant 834587) under the European Union's Horizon 2020 research and innovation program. This article reflects only the authors' views; the European Research Council is not liable for any use that may be made of the contained information.

P.K. is shareholder and member of the board of directors of m-Path Software BV, the spin-off company commercializing m-Path, a platform for experience sampling research and clinical practice.

Supplemental Methods

Selection of the Embedding Model

We evaluated three multilingual sentence-level embedding models to identify the optimal approach for processing Flemish language data. The first model, robert-2022-dutch-sentence-transformers (Netherlands Forensic Institute, 2024), is a Dutch-specific model fine-tuned for semantic similarity tasks. The second model, gte-multilingual-base (Zhang et al., 2024), is a general-purpose multilingual encoder designed for long-context retrieval. The third model, jina-embeddings-v3 (Sturua et al., 2024), is a multilingual model that employs Task LoRA (Low-Rank Adaptation) to enable adaptable semantic representations.

Based on qualitative assessment of topic coherence within our corpus, jina-embeddings-v3 demonstrated superior performance in generating meaningful thematic structures. We attribute this advantage to its XLM-RoBERTa architecture, which is augmented with Rotary Position Embeddings and Task LoRA adapters. All model evaluations were performed using an NVIDIA RTX-5000 GPU and implemented in PyTorch.

Determination of the Binarization Threshold

For all topic models, we extracted the highest-loading topic for each text (i.e., the topic with the highest probability score in the document-topic probability matrix) as well as the continuous probability [0,1] that each text loaded onto each topic. Because texts often included content corresponding to multiple themes or events, we considered a topic to be mentioned if its loading probability surpassed a threshold of 0.01 (i.e., binarizing the data), with this threshold established based on an examination of the distribution of loading probabilities (Figure S1).

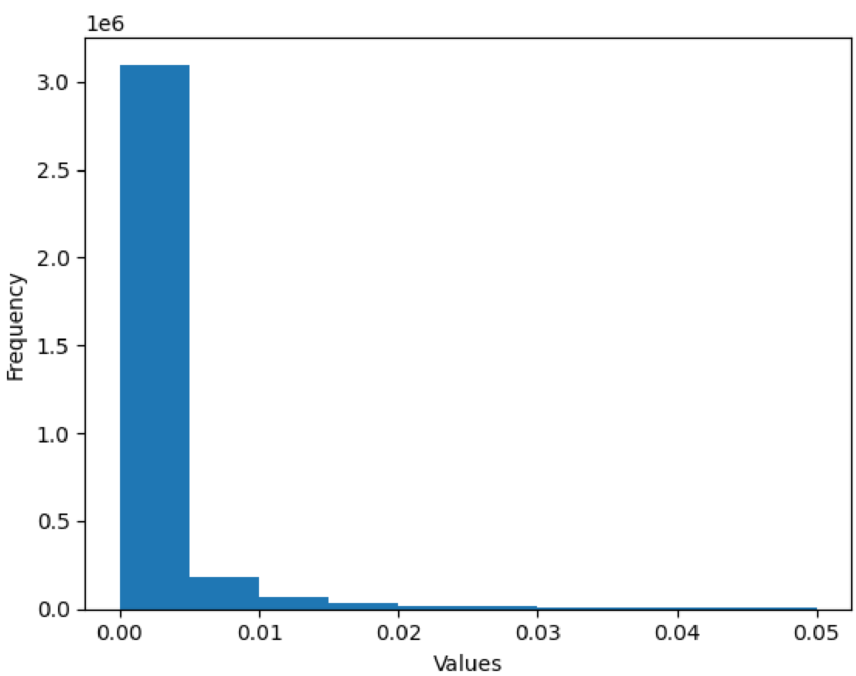

Figure S1. Histogram of text loading probabilities.

Supplemental Results

Relationship between Daily Life Events and Categories for Emotion

To examine whether the many-to-many relationship observed between daily life events and categories for emotion (e.g., pride, anger, sadness, amusement) in our data was due to the specific emotion categories selected, we conducted a robustness check using an expanded set of 69 emotion words that occurred at least 10 times in the data. A topic modeling analysis of the texts containing these words yielded 118 themes. As in the primary analyses, each of the included emotion words was used to label events corresponding with multiple themes (M = 42, SD = 29; range: 2-118). Conversely, individual themes were generally associated with labels for multiple emotion categories (M = 24, SD = 12; range: 1-67). Please see our OSF repository (https://osf.io/8mxs7/overview?view_only=c4abcc6156e5451fa560f76f7d276aea) for full correspondence tables.

References

Netherlands Forensic Institute. (2024). *Robbert-2022-dutch-sentence-transformers* (Version Revision cdf42f6) [Computer software]. HuggingFace. https://doi.org/10.57967/hf/3457

Sturua, S., Mohr, I., Akram, M. K., Günther, M., Wang, B., Krimmel, M., Wang, F., Mastrapas, G., Koukounas, A., & Wang, N. (2024). *jina-embeddings-v3: Multilingual embeddings with task LoRA* (No. arXiv:2409.10173). arXiv.

Zhang, X., Zhang, Y., Long, D., Xie, W., Dai, Z., Tang, J., Lin, H., Yang, B., Xie, P., & Huang, F. (2024). *mgte: Generalized long-context text representation and reranking models for multilingual text retrieval* (No. arXiv:2407.19669). arXiv.