To cite this article:

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

Some recommendations on the use of daily life methods in affective science

Peter Kuppens[1], Egon Dejonckheere[1], Elise K. Kalokerinos[2], Peter Koval[1,2]

[1] KU Leuven, Belgium

[2] University of Melbourne, Australia

Author Note

The writing of this article was supported by a KU Leuven Research Council grant (C14/19/054) and IBOF grant (IBOF/21/090) awarded to P. Kuppens and Discovery Early Career Researcher Awards (DE180100352; DE190100203) awarded to E. Kalokerinos and P. Koval. Egon Dejonckheere is an FWO postdoctoral fellow (1210621N). Correspondence concerning this article should be addressed to Peter Kuppens, email: peter.kuppens@kuleuven.be

Abstract

Real-world emotions are often more vivid, personally meaningful, and consequential than those evoked in the lab. Therefore, studying emotions in daily life is essential to test theories, discover new phenomena, and understand healthy emotional functioning; in short, to move affective science forward. The past decades have seen a surge of research using daily diary, experience sampling, or ecological momentary assessment methods to study emotional phenomena in daily life. In this paper, we will share some of the insights we have gained from our collective experience applying such daily life methods to study everyday affective processes. We highlight what we see as important considerations and caveats involved in using these methods and formulate recommendations to improve their use in future research. These insights focus on the importance of (i) theory and hypothesis-testing; (ii) measurement; (iii) timescale; and (iv) context, when studying emotions in their natural habitat.

Keywords: Affect, Emotions, Daily life methods, Experience Sampling, Ecological Momentary Assessment

Some recommendations on the use of daily life methods in affective science

In many areas of science, phenomena of interest are studied in artificial environments ranging from minuscule petri dishes to large hadron colliders. Although invaluable, this work must be complemented and validated against observations of how the phenomena under study behave outside the laboratory in their natural settings. Likewise, in affective science, lab studies are indispensable to isolate and manipulate emotional stimuli, test causal theoretical predictions, and observe overt and covert responses. Yet to fully capture the richness and complexity of affective phenomena, we must also study emotions in the muddiness of daily life. Daily life research is vital to discover new phenomena that may be difficult to observe under lab conditions, test theoretical predictions, and externally validate findings from experimental work (for an excellent discussion, see Wilhelm & Grossman, 2010) .

There are, of course, a number of essential differences between lab and daily life approaches. Lab methods allow us to systematically manipulate the environment, measure multiple response systems at fine-grained levels, and to draw causal conclusions. In contrast, daily life methods deal with emotions that are arguably more vivid, personally meaningful, and consequential than those evoked in the lab, and allow us to chart contingencies and variation in the complexity of uncontrolled real life. Notwithstanding these differences, few would disagree that lab and daily life are complementary approaches to move affective science forward (and if you're one of these few, please give us a call for an adversarial collaboration).

The past decades have seen an enthusiastic rise in the study of everyday affective processes using daily diary (e.g., completing an end-of-day report about events and experienced during that day), and experience sampling or ecological momentary assessment methods (typically involving multiple assessments per day, in which people report

momentary or recent feelings, behavior, thoughts, and/or context in daily life). This rise is likely driven both by technological advances (e.g., smartphones) that have made this type of research more feasible, and by the realization that relying exclusively on lab methods undermines the ecological validity of our findings. These methods have driven discoveries ranging from computational models of real-life affective changes (e.g., Loossens, et al., 2020) right through to the emotions experienced during a game of golf (e.g., Dickens, et al., 2018). Daily life studies have also made plain that when we measure people outside of experimental control, there is large variation in how emotions behave both within (e.g., Zelenski & Larsen, 2000) and between persons (e.g., Kuppens, et al., 2013).

In spite (or perhaps because) of this enthusiasm for daily life methods, we should not turn a blind eye towards the difficulties and potential drawbacks of using these methods. With this in mind, in this paper we share some of our insights gained from years of collective experience with daily life methods, highlight what we see as important considerations and caveats involved in these methods, and formulate recommendations to improve their future use. These insights focus on the importance of (i) theory and hypothesis-testing; (ii) measurement; (iii) timescale; and (iv) context, when studying emotions in their natural habitat (for an overview, see Table 1).

Theory and Hypothesis Testing

There is no shortage of theories in affective science. Some of the most prominent include theories that explain what causes emotions (for a review, see Moors, 2009) and moods (e.g., Bennett et al., 2020) and theories describing the processes governing emotion and mood regulation (e.g., Gross, 2015; Larsen, 2000). As in any discipline, the value of theories in affective science lies in their ability to explain the underlying causal mechanisms, and/or make accurate predictions about, naturally occurring affective processes. Thus, one might assume that the recent surge of studies applying daily life methods in affective science

would have allowed researchers to test the (ecological) validity of prominent theories, and perhaps even settle long-held debates. However, we believe this potential has not yet been realized. With some exceptions (e.g., Kalokerinos, Tamir, et al., 2017; Kuppens et al., 2010; Wirth et al., 2021), most research on everyday emotions can be classified as "descriptive" (Mottus et al., 2020) or "phenomenon-driven" (Fiedler, 2017). Such research is primarily concerned with describing effects or associations among variables.

Emotion research is not unique in its over-reliance on exploratory, descriptive research. For instance, several writers have expressed concern over a lack of theory-driven research in psychology more broadly (e.g., Fried, 2020; Muthukrishna & Henrich, 2019; Oberauer & Lewandowsky, 2019; Robinaugh et al., 2020), and such criticisms are not new (Meehl, 1978, 1990). As Cummins (2010) put it, the emphasis in psychology on discovering and replicating effects has left us "overwhelmed with things to explain, and somewhat underwhelmed by things to explain them with" (p.286). Although we recognize the importance of descriptive research for advancing (affective) science (Fiedler, 2017; Tukey, 1980), we argue that research on daily emotional processes would benefit from an increase in theory-driven hypothesis testing. Below, we use the example of research on affect dynamics to illustrate the limitations of relying too much on an exploratory approach. We then suggest some steps to counter these problems in future research, while also acknowledging the challenges inherent in theory-driven cumulative science (Fiedler, 2017).

Research on affect dynamics seeks to describe how affective states (i.e., moods, emotions) fluctuate over time, to explain the mechanisms underlying affective fluctuations, and to understand the consequences of different patterns of affect dynamics for psychological functioning (Butler, 2015; Coifman, et al., 2007; Ebner-Priemer & Sawitzki, 2007; Heller et al., 2015; Kuppens, 2015; Kuppens & Verduyn, 2015; Larsen, 1987; Trull, et al., 2015; Waugh & Kuppens, 2021). Over the past 20-30 years, a fairly large body of research on

affect dynamics has accumulated—much of it using experience sampling and other daily life methods. Yet, despite hundreds of studies, synthesized in narrative reviews (e.g., Trull et al., 2015) and a meta-analysis (Houben et al., 2015), we seem to be no closer to a consensus on what constitutes healthy versus unhealthy patterns of affect dynamics. In our view, this is at least partly because research on affect dynamics has not been designed to test theory-derived predictions. Instead, researchers have adopted a largely exploratory and pluralistic approach to studying affect dynamics. For instance, Dejonckheere et al. (2019) identified 16 statistical indicators of affect dynamics and Houben et al. (2015) reviewed studies adopting several others. This diversity in how affect dynamics are quantified not only makes it difficult to navigate the literature, but also raises questions about (i) the uniqueness and (ii) the theoretical or substantive meaning of different indices of affect dynamics.

For example, some researchers use the within-person standard deviation (*SD*) to measure so-called *variability* in affect, while others use the mean squared successive difference (*MSSD*) to operationalize affective *instability*, and still others estimate the autoregressive (*AR*) slope as a measure of affective *inertia*. Each of these indices correlates reliably (and in the same direction) with depressive symptoms (Houben et al., 2015), leading to the paradoxical conclusion that depression involves more variable, more unstable, but also more inert (i.e., self-predictable) affect. Such findings are difficult to interpret without considering the mathematical overlap among these indices (see Bos et al., 2018; Koval et al., 2013). More importantly, how these indices of affect dynamics map onto specific psychological mechanisms or processes remains unclear. Without a strong theory of affect dynamics from which to derive hypotheses to guide study design and statistical analysis, we risk being left with a veritable cornucopia of affect dynamic indicators without a way to determine which measures are theoretically meaningful and which may be redundant (Dejonckheere et al., 2019).

To overcome this, we need to strive for more theory-driven cumulative science (Fiedler, 2017). This requires affective scientists to become more familiar with theory construction, involving "systematic derivation of diagnostic hypotheses from incontestable laws and logical constraints" (Fiedler, 2017, p.53), translation of verbal theories into formal or computational models (Grahek, Schaller, & Tackett, 2021), and mapping of theoretical to statistical models (Robinaugh et al., 2020). A number of useful resources have recently become available to help in this regard, including reviews (e.g., Muthukrishna & Henrich, 2019; Oberauer & Lewandowsky, 2019; Robinaugh et al., 2020; P. Smaldino, 2019) and tutorials (Borsboom et al., 2020; Smaldino, 2020; van Rooij & Blokpoel, 2020; see also https://osf.io/uscfx/ for a reading list compiled by Fried, 2020). As is so often the case, this is easier said than done. Developing explanatory theories may be especially challenging when trying to explain the functioning of a complex (possibly chaotic) system, such as that which may characterize the dynamics of affect in everyday life.

Given this, it may also be fruitful to incorporate scientific approaches that emphasize prediction rather than (or in addition to) explanation into research on everyday affective processes (see e.g., Carlier et al., in press; Dejonckheere et al., 2019). To clarify: whereas explanatory research focuses on precise estimation of causal effects with the goal of understanding the mechanisms underlying observable phenomena, predictive research is primarily concerned with accurately forecasting as-yet-unobserved outcomes regardless of their causes (Shmueli, 2010; Yarkoni & Westfall, 2017). Although explanation and prediction have often been cast as fundamentally opposing scientific goals, Hofman et al. (2021) propose that they can and should be integrated.

Based on the above considerations, we offer some practical recommendations for affective scientists to increase the theoretical rigor of our research: First and foremost, we should strive to clearly identify the scientific goals of our research, be they

descriptive/exploratory, explanatory, predictive, or some combination (see Hofman et al., 2021 for a useful labeling framework). Second, if relevant, we should specify within which theoretical framework a given study is operating and articulate the theory's core assumptions. Third, in the case of explanatory research, we should derive testable hypotheses from our chosen theory and translate these into precise mathematically quantifiable terms. Fourth, in the case of predictive research, we should aim to specify the practical (and/or theoretical) utility of accurately predicting our outcome, and justify how we evaluate predictive accuracy (for more on this, see Yarkoni & Westfall, 2017).

Measurement

To empirically test theoretical predictions, we must be able to adequately measure the variables involved. As most of daily life emotion research has focused on the experience of emotion, we will focus here on the measurement of the experiential component of emotion (and, to a lesser extent, related self-reported information regarding context and behavior or physiology). While self-report can be readily criticized on a number of fronts, it is indispensable in affective science for two reasons.

First, self-report is - as far as our science stands - the only way to get direct insight into people's subjective experience of emotion. As definitively stated by Barrett, Mesquita, Ochsner, and Gross (2007): "The most direct way to measure the contents of a mental representation of emotion is to examine people's verbal behaviors regarding their own mental state, in the form of self-reports" (p. 377). We certainly have at our disposal other, more objective, means to measure emotional behavior and physiology, and technical advances are increasingly allowing their measurement in the context of daily life (e.g., Hoemann et al., 2020). Yet, based on current knowledge, these cannot be considered direct proxies of emotional experience. Rather, they reflect other contexts, components or corollaries of emotions that only loosely covary with experience and one another during emotional

episodes and across contexts, individuals, and culture (Barrett, et al., 2007; Mauss and Robinson, 2009; Van Halem, et al., 2020). Also when it comes to the newest technology, there is so far no clear indication that for instance passive smartphone sensing or wearable data will be able to predict emotional experience with great precision (e.g., Niemeijer, et al., 2021; Tag, et al., 2021).

Second, among these different components of emotion, subjective experience is arguably the most defining of people's emotions (see Lieberman, 2019). It may not be a sufficient condition to speak of an emotion occurring, but it certainly seems a necessary condition. One can perhaps imagine a person saying to be angry without displaying any behavioral or physiological changes to be challenged on the truthfulness of his or her words. But we doubt that one would agree to speak of an emotion occurring in the absence of subjective experience (Lieberman, 2019, even argues experience is both necessary and sufficient for emotion). These reasons imply that if we want to study how people feel in daily life, we can't do without self-reports of emotional experience. Add to these reasons the fact that self-report is relatively easy to collect (but see below), it comes as no surprise that it is the most used method to study emotion in daily life.

So, what is the current state of the art in emotion measurement in daily life? Here, we stumble on a remarkable situation: We have an enormous psychometric tradition in psychological science devoted to measuring stable properties of the mind. Over the last century, think of all the attention that has been devoted to measuring individual differences in intelligence, personality, values, and related constructs. Influential theories of measurement (e.g., classic test theory, item response theory), widely adopted statistical procedures (e.g., factor analysis, SEM), and eager business models (e.g., testing and statistical services) have been developed to evaluate and guarantee our measurement of these properties. Likewise, neither cost nor effort have been spared to develop objective measurement instruments and

machinery to get insight into the workings of the mind. Yet, this all stands in stark contrast to the little attention devoted to the repeated measurement of more transient states of the mind (e.g., Horstmann & Ziegler, 2020), of which emotional experience is a prominent example.

Indeed, the history, tenacity, and pride that accompanies our field's insistence on the importance of good measurement seems almost absent in daily life research. In a convincing but sobering review of the literature, Brose et al. (2020) recently highlighted the lack of consensus on self-report instruments to measure daily life emotions: a majority of scales lack theoretical justification (or do not apply their theory when scoring scales), resulting in a large variation in how to operationalize, assess, score, and interpret measures of momentary affect (see also Weidman, Steckler, & Tracy, 2017). Most certainly, there are many examples where researchers have undertaken efforts to systematically improve the measurement of psychological constructs in the context of daily life, and this should not be minimized (e.g., Schimmack, 2003; Shiffman, Stone, & Hufford, 2008; Schuurman & Hamaker, 2019; Stone & Shiffman, 1994). Yet, when we focus on the measurement of emotional experience, there is simply not a lot of systematicity; and the same holds a fortiori when it comes to measuring emotion regulation (but see Medland, et al., 2020), quality of emotional events or contexts (see below), and so on.

Of course, this is not coincidental. First, we lack a consensual definition of emotion (or mood, affect, etc.), and this impedes consensus on measurement. Second, how do you assess the reliability and validity for measures of transient states? Unlike classic psychometrics, where one assumes the temporal stability of its object (Lord & Novick, 1968), how do you capture and validate a measurement of something that is temporally dynamic by its very nature? We return to this issue below.

Yet, the fact that is difficult does not mean it is not important, nor that we should not pursue it. We agree with Brose et al. (2020) that affective science would hugely benefit from

a reliable, validated, consensually adopted instrument to measure emotional experience in daily life. After all, along with theory, good measurement is the foundation on which empirical findings that truly advance our science are built. Bad measurement, in contrast, does the opposite. In the next section, we share two examples from our own research where sloppy measurement has been or is impeding progress.

The first example, relating to reliability, comes from research using the affect grid to measure people's momentary feelings as coordinates along the valence and arousal dimensions of core affective space (Russell, Weiss, & Mendelsohn, 1989). In our research (e.g., Kuppens, et al., 2007; Kuppens, Oravecz, & Tuerlinkcx, 2010), we have used different versions of this grid to repeatedly measure valence and arousal in daily life. One version asked participants to indicate their current affective state on a fully visible 9x9 grid. A second version consisted of a continuous 99x99 grid that showed valence and arousal axes on a blank background (i.e., no visible grid). Participants' responses were later recoded to values ranging from 0 to 99 for each dimension. With the latter version, we noticed in pilot studies that heat maps of collected data showed an X-shaped form, indicating that participants tended to use the diagonals of the space as implicit anchor points for reporting their feelings. We did not observe this pattern of responding when using the original fully visible 9x9 affect grid. This measurement issue was problematic, and needs for instance to be carefully considered when examining the relation between valence and arousal to avoid a measurement bias lying at the basis of substantive conclusions (e.g., Kuppens, Tuerlinckx, Russell, & Barrett, 2013 used data from different sources to rule out this possibility).

The second example, relating to validity, comes from a recent study in which we measured students' emotional responses to receiving exam grades as an ecologically valid, high-stakes emotional event in Belgian students' daily lives (see Kalokerinos et al., 2019). Here, we noticed that when asking participants to report their feelings in the period

surrounding such an event, it mattered for the findings and conclusions whether participants reported how they felt “in general” or in relation to the specific event (e.g., see also Dejonckheere, et al., 2019). This means that specifying the object of people’s feelings (or not) when measuring emotions in daily life is consequential and should be carefully considered.

These two examples make clear that even small measurement changes can change our conclusions and suggest the need for a consensus in emotion measurement. If we were to go for such a consensual instrument, what should this instrument look like? We see three key generic and specific criteria. First, this instrument should be able to assess a number of differentiated emotional states, as well as capturing some of the more abstract dimensions capturing emotional experience. While there may be no consensual definition of emotion (Coan, 2010), there seems little disagreement that with (1) a limited number of typical categories we call anger, sadness, anxiety, happiness, etc. and (2) dimensions of valence and arousal and/or positive and negative affectivity, we should be able to capture a large portion of the qualitative variation in emotional experience (see also, Weidman, et al., 2017).

Second, assessing emotion in daily life requires attention to low intensity emotional states. Sampling people’s emotions regularly in everyday life may not give rise to an abundance of strong emotional responses to momentous events, but will rather capture less intense and more mood-like states, like feeling *tense*, *upbeat*, or *content*.

Third, a good instrument for assessing momentary affect must conform to the classic criteria of reliability and validity. A reliable instrument produces the same values whenever it measures the same emotional state, and this can be empirically verified (it should be said that in the context of single item measures, which are often used in daily life research to limit participant burden, this is more difficult to assess, but not impossible). A valid instrument measures what it is supposed to assess, and this can be evaluated by looking at distributional

characteristics of emotional responses, the systematic covariation of emotional responses (in specific contexts), and predicted relations with third variables. We feel that the development of standard, consensual, reliable and valid instruments to measure people's emotional experiences should be high on the agenda of affective science, if daily life research is to provide cumulative contributions.

Until such an instrument has been developed, the concrete recommendations we can offer are to (1) combine specific emotion labels with more general dimensions of emotional experience, (2) take into account that using intense emotion labels may result in little meaningful variance in the data, (3) carefully consider whether you will assess emotions tied to a certain context or not, and (4) design your study such that you are able to verify elements of reliability (for instance, in terms of split half consistency or scale reliability) and validity (by making sure to include variables to evaluate predictive, convergent and/or discriminant validity).

Timescale

Emotions change across seconds, minutes, hours, and days. Indeed, the dynamic nature of emotions is central to most emotion theories (Kuppens, 2015; Kuppens & Waugh, 2021). A benefit of daily life methods is that they allow us to capture these emotions as they unfold across time. However, achieving this goal requires grappling with several factors including the interval between assessments, temporal dependencies between variables, and incorporating time in measurement.

A first issue is how we decide to space the interval between momentary assessments. Emotion processes continuously unfold across time, but the active self-report methods most common in the field entail discrete measurement (Hamaker & Wichers, 2017). This means we are effectively capturing stills from a film, and by necessity, missing much of the storyline. Of course, such criticism does not apply to passive sensing methods (e.g., GPS,

heart-rate), but as mentioned, many emotion processes are difficult to map directly onto such methods. This leaves researchers to wrestle with timing their measures to capture as much of the story as possible.

The most common design in emotion research is an experimenter-initiated scheme, meaning that the researcher (and/or a computer algorithm) sets the interval between surveys. The intervals used in existing studies vary widely, from very dense designs with 15 minutes between measurement occasions (e.g., Ebner-Priemer & Sawitzi, 2007; Koval & Kuppens, 2012) right through to diary designs with 24 hours between occasions (e.g., Kalokerinos, Tamir, & Kuppens, 2017). This choice of interval is often not justified in papers (Haslbeck & Ryan, 2020; Trull & Ebner-Priemer, 2020), which is surprising, given that it is a non-trivial decision. If we wish to capture change, measurements must be spaced at the timescale at which changes occur: if they are too sparse, we will be unable to capture the dynamics in our data (Hamaker & Wichers, 2017). The spacing decision should be based on a thoughtful consideration of theory and evidence surrounding the timescale of the target processes, but this evidence is often lacking, making such decisions difficult.

One set of evidence that affective scientists can turn to for inspiration investigates emotion duration. Diary studies of everyday emotional episodes suggest that most end quickly: for example, in one study, 80% of emotions had dropped back to baseline within the first hour (Verduyn et al., 2009). In a study focusing on the most intense negative event of the day, and thus likely capturing stronger emotions, episodes ranged in duration from 10 seconds through to 24 hours, with a median duration of 1 hour 25 minutes (Kalokerinos, Rèsibois, Verduyn, & Kuppens, 2017). This research suggests that a typical design (e.g., with between 5 or 10 measurement occasions per day) may capture only the unfolding of the most intense emotional episodes, and such episodes are uncommon in daily life (Kalokerinos et al., 2020), meaning a shorter sampling schedule might improve our understanding. In addition,

emotion duration varied with the specific emotion, situational features, and individual differences (Verduyn et al., 2015), suggesting that thought must be given to the specific context in which the study is conducted.

There are also practicalities to be considered, and a trade-off needs to be made between fidelity and participant burden in completing the study. If we begin to shorten the intervals between surveys, we may have to compensate by sampling for fewer days or including fewer items – a decision that may come with other costs. Alternate options could be to have short bursts of more frequent measurements, perhaps by having a trigger of interest (e.g., a stressor) prompt more densely spaced assessments (Bolger & Laurenceau, 2013; Dejonckheere & Mestdagh, 2021).

Another consideration is how to effectively capture change in time-series data. The most prevalent methods include incorporating values from previous measurement occasions - often referred to as "lagged variables" – into statistical models predicting values at the current occasion (Ariens, Ceulemans, & Adolf, 2020). Lagged associations capitalize on the ability of ESM designs to investigate change across time and to determine directionality, and so are often preferred by researchers over contemporaneous associations, in which variables are associated within one measurement occasion.

However, the strength of lagged relationships is known to depend on the spacing between time-points (Gollob & Reichardt, 1987), which can lead to researchers making conclusions primarily driven by the length of the intervals in their data. Thus, if the timescale of a process is not known, researchers should be careful using lagged associations – these associations will capture any process that is *longer* than the interval between surveys, but anything *shorter* than that interval will be captured only in contemporaneous effects (Granger, 1969). For example, a typical experience-sampling study with a 90-minute interval between surveys may find no association between lagged itchiness and current scratching

behavior: from that, we should not conclude that being itchy does not lead to scratching. The timescale of change between these two variables is such that their association would only be reflected in the contemporaneous association. In sum, lags should be considered carefully and with reference to theory, and one should be very careful about using lagged associations to drive causal inference (Eichler, 2012).

Daily life data can also be understood at higher-level timescales. Experience sampling data reflect cyclical patterns, and emotion data is no exception – for example, positive affect follows a circadian rhythm (Clark, Watson, & Leeka, 1989). Ignoring such cycles can lead to misleading conclusions (for an excellent discussion of this, we refer to Liu & West, 2015). There are cyclical analytic methods to detect this kind of change for within-day emotion processes (van de Maat, Lataster, & Verboon, 2020) and for weekly cycles (Liu & West, 2015). However, detecting cycles will depend on the length of the data collection period, and so must also be considered in study design.

Finally, timescale also intersects with measurement. Many items can be assessed with a momentary frame (e.g., "*right now, how angry do you feel?*"), but other items are best assessed retrospectively, often because they are not common enough to be captured momentarily. For example, in our work, we have often asked participants to report emotion regulation strategy use "*since the last survey*" (e.g., Kalokerinos et al., 2019). This means we must consider the temporal ordering created by our item wording: with this example, emotion regulation strategies precede emotion, even when they are assessed simultaneously. It should be noted that using "*since the last survey*" may be problematic because the intervals between surveys are often different (because of missed surveys, or random-interval designs). It may sometimes be more effective to think about the timescale of processes and anchor the items using real time, for example, "*in the last hour*" (see e.g., Medland et al., 2020). When using these kinds of items, it is important to note that people draw on qualitatively different types

of information when judging their experience in the moment versus over a longer time period (Robinson & Clore, 2002), echoing the distinction between the experiencing self and the remembering self. Especially in the case of daily life research that often uses both momentary items and those assessing a longer time period, this distinction is important consider when designing and interpreting studies (see Conner & Barrett, 2012).

Measurement may also intersect with timescale in yet another manner. As recently shown by Shrout and colleagues (2018), repeated self-report measures of a same variable may show an initial elevation bias, in which initial measurements show higher levels compared to subsequent measurements. When handling repeated measures data, it is therefore important to check the presence of such initial elevation bias, and correct for it if needed, in order to avoid spurious effects or correlations.

In sum, we need to carefully theorize about timescale in conducting and analyzing our daily life studies. In terms of concrete recommendations, first, it will be important to begin by spacing measurements appropriately to capture change in the phenomena being tested, if capturing change is central to the research. However, given the lack of strong descriptive data or theory surrounding timescale in many areas, this may not always be possible. Here, it might be useful to consider an exploratory analytic method to investigate the scale of change (e.g., Jacobson et al., 2019), keeping in mind that analytic methods cannot compensate for overly sparse sampling (Haslbeck & Ryan, 2020), and using this as a basis to design studies and test theories. Second, researchers should consider using contemporaneous associations as well as lagged associations when the timescale of a process is not known, since lagged associations can mask effects that are more fast-moving than the spacing between measurements. Third, researchers should give thought to cycles across days and weeks, and model them where warranted. Fourth, researchers should consider item wording

carefully, using semi-retrospective wording only where it maps both analyses and the goals of the data collection.

Context

Finally, emotions do not occur in a vacuum, but are strongly contextually driven (Greenaway et al., 2018). While the emotional responses captured with daily life methods are often praised for their personal relevance and ecological quality, we know surprisingly little about these personal ecologies themselves (Mestdagh & Dejonckheere, 2021). This is problematic, not in the least because affective scientists agree that the adaptiveness or functionality of an emotional reaction can only be evaluated considering the context in which it takes place (e.g., Barker, 1968; Frijda, 1988). Nevertheless, it is common for daily life emotion research to study emotional responses without proper recognition of the specific stimuli or events that underlie those responses (Lapate & Heller, 2020). This disregard for context may explain, for instance, why linking distinctive emotional patterns to indicators of psychological (mal)adjustment has not always been fruitful (Dejonckheere et al., 2019). But why do daily life methods struggle with the incorporation of context information, and how can we resolve this issue?

First, traditional self-report prohibits a fine-grained assessment of the relevant situational parameters in everyday life. This is because the need for standardized items and/or response formats requires researchers to construct abstract context categories that generally fail to capture the subtle nuances that define the emotional relevance of an event (e.g., *What are you doing at the moment?* [*Work – Hobby – Sports – Therapy – Rest*]; Heininga et al., 2019). Equally questionable is the fact that many self-reported context items exhibit an emotional undertone themselves (e.g., *Did you have a* [*negative – neutral – positive*] *social interaction since the last prompt?*; Rivera et al., 2020), mixing a person's affective response with features of the stimulus. Passive sensing via wearables or smartphone sensors may

provide both the objective window and the level of detail that is required to accurately assess contextual information (Jacobson & Chung, 2020; Laport-López, et al., 2020; Niemeijer & Kuppens, 2021), yet combining these rich but noisy data streams into a coherent and meaningful contextual signal poses a major challenge (Ram et al., 2017).

But even when passive sensors provide precise and objective insights into participants' natural habitat, a second concern is that, for the vast majority of people, their everyday lives mostly comprise a chain of mundane events that hardly trigger a sizable, lasting emotional reaction (see also above). Only sporadically are people's daily routines interrupted by unusual and impactful events that elicit strong felt emotions (e.g., a positive encounter, success or failure at work, conflict with partner, etc.). The intensity and duration of these emotional reactions reflect a thrust out of emotional equilibrium (Chow et al., 2005; Loossens, et al., 2020; Kuppens et al., 2010), and trigger an emotional trajectory that is sufficiently pronounced to emerge against the background of inevitable measurement noise (Dejonckheere et al., 2020). In combination with their personal relevance and real-life status, these affective responses are extremely valuable for emotion researchers, but not always easy to capture in daily life.

To overcome this difficulty, we see two primary options that we can recommend. First, daily life researchers could give up on the premise of complete non-interference and introduce an impactful stressor into the daily lives of participants themselves (e.g., a Trier Social Stress Task; Koval & Kuppens, 2012). Second, researchers could design a study around the occurrence of an upcoming event that is relatively easy to predict (e.g., the release of students' exam results; Dejonckheere et al., 2019; Kalokerinos et al., 2019; Metalsky et al., 1993; the birth of a child; Belisario, et al., 2017). An additional benefit of these quasi-experiments in the wild is that the contextual input that participants receive is, at least to some extent, comparable across individuals. This makes between-person comparisons in

emotional reactivity or recovery more feasible (e.g., the hospital setting could partly explain variation in emotional responding between in- and out-patients with BPD features; Houben et al., 2017). Furthermore, in both cases, if the research question allows, the practice of anchoring all emotion items to the event, stimulus or situation of interest (e.g., How stressed are you right now with respect to your exams?) may further fine-tune the contextual signal underlying people's affective response (Dejonckheere & Mestdagh, 2021).

In sum, these recommendations, summarized in Table 1, are an invitation for daily life researchers to further open the contextual black box of everyday life. We advise daily life researchers to take emotional event information seriously in all of its aspects (i.e., accurate assessment, strength of stimuli, differential input, etc.). Only then will we able to truly quantify and judge the appropriateness of emotional reactions in real-life contexts.

Conclusion

Daily life emotion research is burgeoning. This can be a blessing for the field of affective science, as daily life research can complement lab research the way clinical research complements petri dishes and telescope observations complement particle accelerators. Yet, it is in our common interest that this daily life research is theoretically driven, relies on proper measurement, and takes seriously the temporal and contextual nature of emotions. We hope that this article may help to motivate daily life affective scientists to take on this task, and provide some helpful guidance should they choose to do so.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Funding

The writing of this article was supported by a KU Leuven Research Council grant (C14/19/054) and IBOF grant (IBOF/21/090) awarded to P. Kuppens and Discovery Early Career Researcher Awards (DE180100352; DE190100203) awarded to E. Kalokerinos and P. Koval. Egon Dejonckheere is an FWO postdoctoral fellow (1210621N).

Ethics approval

As this paper does not report data, matters related to ethics approval, informed consent, or data/code availability are not applicable.

Author statement

PK, ED, EK, and PK drafted and finalised the text for this manuscript.

Acknowledgments

We acknowledge the help of Leonie Cloos for helpful comments on a draft of this manuscript.

References

Ariens, S., Ceulemans, E., & Adolf, J. K. (2020). Time series analysis of intensive longitudinal data in psychosomatic research: A methodological overview. *Journal of Psychosomatic Research, 137*, 110191.

Barker, R. G. (1968). *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford: Stanford University Press.

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, 58, 373-403.

Belisario, J. S. M., Doherty, K., O'Donoghue, J., Ramchandani, P., Majeed, A., Doherty, G., ... & Car, J. (2017). A bespoke mobile application for the longitudinal assessment of depression and mood during pregnancy: protocol of a feasibility study. *BMJ open, 7*, e014469.

Bennett, D., Davidson, G., & Niv, Y. (2021). A model of mood as integrated advantage. Manuscript submitted for publication. https://doi.org/10.31234/osf.io/dzsme

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York: Guilford Press.

Borsboom, D., van der Maas, H., Dalege, J., Kievit, R., & Haig, B. (2021). Theory Construction Methodology: A practical framework for theory formation in psychology. https://doi.org/10.31234/osf.io/w5tp8

Bos, E. H., de Jonge, P., & Cox, R. F. A. (2019). Affective variability in depression: Revisiting the inertia-instability paradox. *British Journal of Psychology, 110*, 814-827. https://doi.org/10.1111/bjop.12372

Brose, A., Schmiedek, F., Gerstorf, D., & Voelkle, M. C. (2020). The measurement of within-person affect variation. *Emotion, 20*, 677.

Butler, E. A. (2015). Interpersonal affect dynamics: It takes two (and time) to tango. *Emotion Review, 7*, 336-341.

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

Chow, S.-M., Ram, N., Boker, S. M., Fujita, F., & Clore, G. (2005). Emotion as a thermostat: Representing emotion regulation using a damped oscillator model. *Emotion*, *5*, 208–225.

Clark, L. A., Watson, D., & Leeka, J. (1989). Diurnal variation in the positive affects. *Motivation and Emotion, 13*(3), 205–234.

Coan, J. A. (2010). Emergent ghosts of the emotion machine. *Emotion Review, 2*, 274–285.

Coifman, K. G., Bonanno, G. A., & Rafaeli, E. (2007). Affect dynamics, bereavement and resilience to loss. *Journal of Happiness Studies, 8*, 371-392.

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic medicine, 74*, 327-337.

Cummins, R. (2013). *The world in the head*. Oxford: Oxford University Press.

Dejonckheere, E., & Mestdagh, M. (2021). On the signal-to-noise ratio in emotional time series. In C. Waugh & P. Kuppens (Eds.), *Affect dynamics*. Springer Nature.

Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P., & Tuerlinckx, F. (2019). Complex affect dynamic measures add limited information to the prediction of psychological well-being. *Nature Human Behaviour, 3*, 478–491.

Dejonckheere, E., Mestdagh, M., Kuppens, P., & Tuerlinckx, F. (2020). Reply to: Context matters for affective chronometry. *Nature Human Behavior, 4*, 690–693.

Dejonckheere, E., Mestdagh, M., Verdonck, S., Lafit, G., Ceulemans, E., Bastian, B., & Kalokerinos, E. K. (2019). The relation between positive and negative affect becomes more negative in response to personally relevant events. Emotion. Advanced online publication.

Dickens, Y. L., Van Raalte, J., & Hurlburt, R. T. (2018). On investigating self-talk: A descriptive experience sampling study of inner experience during golf performance. *The Sport Psychologist, 32*, 66-73.

Ebner-Priemer, U. W., & Sawitzki, G. (2007). Ambulatory assessment of affective instability in borderline personality disorder: The effect of the sampling frequency. European *Journal of Psychological Assessment, 23*, 238–247. doi: 10.1027/1015-5759.23.4.238

Eichler, M. (2012). Causal inference in time series analysis. In Berzuini, C., Dawid, P., Bernardinelli, L. (Eds.), *Causality: Statistical perspectives and applications* (pp. 327–354). Chichester, England: John Wiley & Sons. doi:10.1002/9781119945710.ch22

Fiedler, K. (2017). What constitutes strong psychological science? The (neglected) role of diagnosticity and a priori theorizing. *Perspectives on Psychological Science, 12*, 46–61.

Fried, E. I. (2021). Lack of theory building and testing impedes progress in the factor and network literature. https://doi.org/10.31234/osf.io/zg84s

Frijda, N. H. (1988). The laws of emotion. *American Psychologist, 43*, 349–358.

Greenaway, K. H., Kalokerinos, E. K., & Williams, L. A. (2018). Context is everything (in emotion research). *Social and Personality Psychology Compass, 12*, e12393.

Gollob, H., & Reichardt, C. (1987). Taking account of time lags in causal models. *Child Development, 58*, 80-92. doi:10.2307/1130293

Grahek, I., Schaller, M., & Tackett, J. L. (2021). Anatomy of a psychological theory: Integrating construct-validation and computational-modeling methods to advance theorizing. *Perspectives on Psychological Science*, 1745691620966794.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424-438.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, 26, 1–26.

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science, 26*, 10–15. 10.1177/0963721416666518

Haslbeck, J. M. B., & Ryan, O. (2020, August 4). Recovering within-person dynamics from psychological time series. https://doi.org/10.31234/osf.io/dymhw

Heininga, V. E., Dejonckheere, E., Houben, M., Obbels, J., Sienaert, P., Leroy, B., van Roy, J., & Kuppens, P. (2019). The dynamical signature of anhedonia in major depressive disorder: Positive emotion dynamics, reactivity, and recovery. *BMC Psychiatry*, *19*, 59.

Heller, A. S., Fox, A. S., Wing, E. K., McQuisition, K. M., Vack, N. J., & Davidson, R. J. (2015). The neurodynamics of affect in the laboratory predicts persistence of real-world emotional responses. *Journal of Neuroscience*, *35*(29), 10503-10509.

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*(1), 12459.

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., ... & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, *595*(7866), 181-188.

Horstmann, K. T., & Ziegler, M. (2021). Assessing personality states: What to consider when constructing personality state measures. *European Journal of Personality, per.2266*. https://doi.org/10.1002/per.2266

Houben, M., Claes, L., Vansteelandt, K., Berens, A., Sleuwaegen, E., & Kuppens, P. (2017). The emotion regulation function of nonsuicidal self-injury: A momentary assessment study in inpatients with borderline personality disorder features. *Journal of Abnormal Psychology*, *126*, 89–95.

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, 141, 901–930.

Jacobson, N. C., Chow, S. M., & Newman, M. G. (2019). The Differential Time-Varying Effect Model (DTVEM): A tool for diagnosing and modeling time lags in intensive longitudinal data. *Behavior Research Methods, 51*(1), 295-315.

Jacobson, N. C., & Chung, Y. J. (2020). Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors, 20*(12), 3572.

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*, 863–879.

Kalokerinos, E. K., Murphy, S. C., Koval, P., Bailen, N. H., Crombez, G., Hollenstein, T., ... & Bastian, B. (2020). Neuroticism may not reflect emotional variability. *Proceedings of the National Academy of Sciences*, *117*, 9270-9276.

Kalokerinos, E. K., Résibois, M., Verduyn, P., & Kuppens, P. (2017). The temporal deployment of emotion regulation strategies during negative emotional episodes. *Emotion, 17*, 450–458. doi: 10.1037/emo0000248

Kalokerinos, E. K., Tamir, M., & Kuppens, P. (2017). Instrumental motives in negative emotion regulation in daily life: Frequency, consistency, and predictors. *Emotion, 17,* 648-657. doi: 10.1037/emo0000269

Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion, 12*, 256–267.

Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: variable, unstable or inert? *Emotion , 13*, 1132–1141.

Kuppens, P. (2015). It's about time: A special section on affect dynamics. *Emotion Review, 7*, 297–300.

Kuppens, P., & Verduyn, P. (2015). Looking at emotion regulation through the window of emotion dynamics. *Psychological Inquiry, 26*, 72–79.

Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. *Journal of Personality and Social Psychology, 99*, 1042-1060.

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin, 139*, 917-940.

Kuppens, P., Van Mechelen, I., Nezlek, J. B., Dossche, D., & Timmermans, T. (2007). Individual differences in core affect variability and their relationship to personality and psychological adjustment. *Emotion, 7*, 262-274.

Lapate, R. C., & Heller, A. S. (2020). Context matters for affective chronometry. *Nature Human Behaviour, 4,* 688–689.

Laport-López, F., Serrano, E., Bajo, J., & Campbell, A. T. (2020). A review of mobile sensing systems, applications, and opportunities. *Knowledge and Information Systems*, *62*(1), 145-174.

Larsen, R. J. (1987). The stability of mood variability: A spectral analytic approach to daily mood assessments. *Journal of Personality and Social Psychology, 52*, 1195.

Larsen, R. J. (2000). Toward a science of mood regulation. *Psychological Inquiry, 11*, 129–141.

Lieberman, M. D. (2019). Boo! The consciousness problem in emotion. Cognition and *Emotion, 33*, 24-30.

Liu, Y., & West, S. G. (2016). Weekly cycles in daily report data: An overlooked issue. *Journal of Personality*, *84*, 560-579.

Lord, F. M., & Novick, R. (1968). *Statistical theories of mental test scores*. Reading MA: AddisonWesley

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*, 209-237.

Medland, H., De France, K., Hollenstein, T., Mussoff, D., & Koval, P. (2020). Regulating emotion systems in everyday life. *European Journal of Psychological Assessment, 36*, 437–446.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806–834.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66*, 195–244.

Mestdagh, M., & Dejonckheere, E. (2021). Ambulatory assessment in psychopathology research: Current achievements and future ambitions. *Current Opinion in Psychology*.

Metalsky, G. I., Joiner, T. E., Hardin, T. S., & Abramson, L. Y. (1993). Depressive reactions to failure in a naturalistic setting: A test of the hopelessness and self-esteem theories of depression. *Journal of Abnormal Psychology*, *102*, 101–109.

Moors, A. (2009). Theories of emotion causation: A review. *Cognition and Emotion, 23*, 625–662.

Mõttus, R., Wood, D., Condon, D. M., Back, M., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., Lukaszewski, A., & al., E. (2020). Descriptive, predictive and explanatory personality research: Different goals, different approaches, but a shared need to move beyond the Big Few traits. https://doi.org/10.31234/osf.io/hvk5p

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour, 3*, 221–229.

Niemeijer, K., & Kuppens, P. (in revision). Emotion detection with mobile sensing. In Mehl, M. R., Eid, M., Wrzus, C., Harari, G., & Ebner-Priemer, U. (Eds.), *Mobile Sensing in Psychology: Methods and application*.

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review, 26*, 1596–1618.

Ram, N., Brinberg, M., Pincus, A. L., & Conroy, D. E. (2017). The questionable ecological validity of ecological momentary assessment: considerations for design and analysis. *Research in Human Development, 14*, 253–270.

Rivera, N. V., Parmelee, P. A., & Smith, D. M. (2020). The impact of social interactions and pain on daily positive and negative affect in adults with osteoarthritis of the knee. *Aging and Mental Health, 24*, 8-14.

Robinaugh, D., Haslbeck, J. M. B., Ryan, O., Fried, E. I., & Waldorp, L. (2020). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. https://doi.org/10.31234/osf.io/ugz7y

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

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*, 493-502.

Schimmack, U. (2003). Affect measurement in experience sampling research. *Journal of Happiness Studies*, 4, 79-106.

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological Methods, 24*, 70–91.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1-32.

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*, 289–310.

Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavél, F. D., Iida, M., Gleason, M. E. J., Xu, J. H., & Bolger, N. (2018). Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences, 115*(1), E15-E23.

Smaldino, P. (2019). Better methods can't make up for mediocre theory [Review of Better methods can't make up for mediocre theory]. Nature, 575(7781), 9.

Smaldino, P. E. (2020). How to translate a verbal theory into a formal model. *Social Psychology, 51*, 207–218.

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

Tag, B., Webber, S., Wadley, G., Bartlett, V., Goncalves, J., Koval, P., ..., & Kostakos, V. (2021). Making sense of emotion-sensing: Workshop on quantifying human emotions. UbiComp/ISWC 2021 - *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 226-229. doi:10.1145/3460418.3479272

Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology, 129*, 56–63. doi: 10.1037/abn0000473

Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective dynamics in psychopathology. *Emotion Review, 7*, 355–361.

Tukey, J. W. (1980). We need both exploratory and confirmatory. *The American Statistician, 34*, 23–25.

van Berkel, N., Dennis, S., Zyphur, M., Li, J., Heathcote, A., & Kostakos, V. (2021). Modeling interaction as a complex system. *Human–Computer Interaction*, 1–27.

van de Maat, R., Lataster, J., & Verboon, P. (2021). Why and how to deal with diurnal cyclic patterns in ambulatory assessment of emotions. *European Journal of Psychological Assessment*.

van Halem, S., van Roekel, E., Kroencke, L., Kuper, N., & Denissen, J. (2020). Moments that matter? On the complexity of using triggers based on skin conductance to sample arousing events within an experience sampling framework. *European Journal of Personality*, *34*, 794–807.

van Rooij, I., & Blokpoel, M. (2020). Formalizing verbal theories. *Social Psychology*, *51*, 285–298.

Verduyn, P., Delaveau, P., Rotgé, J.-Y., Fossati, P., & Van Mechelen, I. (2015). Determinants of emotion duration and underlying psychological and neural mechanisms. *Emotion Review, 7*, 330–335. doi: 10.1177/1754073915590618

Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Van Mechelen, I. (2009). Predicting the duration of emotional experience: Two experience sampling studies. *Emotion, 9*, 83–91. doi: 10.1037/a0014610

Waugh, C., & Kuppens, P. (Eds.). (2021). *Affect Dynamics*. Springer.

Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion, 17*, 267.

Wilhelm, F. H., & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology, 84*, 552-569.

Wirth, M., Voss, A., Wirth, S., & Rothermund, K. (in press). Affect dynamics and well-being: explanatory power of the model of intraindividual variability in affect. *Cognition & Emotion*.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*, 1100–1122.

Zelenski, J. M., & Larsen, R. J. (2000). The distribution of basic emotions in everyday life: A state and trait perspective from experience sampling data. *Journal of Research in Personality, 34*, 178-197.

Table 1

List of recommendations to improve the contribution of daily life methods for affective science per domain and associated concern

Domain	Concern	Recommendation
Theory and hypothesis testing	Scientific goal	*Explicitly state whether your study is descriptive/exploratory, explanatory, predictive, or some combination*
	Theoretical basis	*Specify theoretical framework and core assumptions behind the study*
	Hypothesis testing	*Derive testable hypotheses from your chosen theory and translate these into precise mathematically quantifiable terms*
	Prediction	*Specify the practical (and/or theoretical) utility of accurately predicting your outcome, and justify how you evaluate predictive accuracy*
Measurement	Type of items	*Consider and/or combine specific emotion labels and general affective dimension items*
	Frequency of occurrence	*Low-frequency items may result in very low variance*
	Context items	*Consider whether you want to measure context-specific or context-free emotion*
	Reliability and validity	*Design your study such that reliability and validity can be verified where possible*
Timescale	Frequency of measurement	*Adapt frequency of measurement to timescale of phenomenon under study, or build in ways to verify this*
	Measurement lags	*Examine both contemporaneous and lagged relations when timescale is unknown*
	Cycles and trends	*Consider, verify, and model the presence of time trends and cycles in your data*
	Timescale of item	*Match the time wording of your items to theory, study goals, and analyses*
Context	Capturing sufficiently informative context	*Design your study around introduced or anticipated emotionally impactful event(s)*
	Capturing context-specific emotion	*Anchor items to your context or event of interest*