

14

UNDERSTANDING DAILY LIFE WITH ECOLOGICAL MOMENTARY ASSESSMENT¹

David B. Newman and Arthur A. Stone

Introduction

Consumer behavior researchers are primarily interested in what happens in the real world. When are people likely to consume a product? How do decisions influence our emotions over time? Many questions such as these can only be answered by studying these topics in everyday life. That is, particular phenomena may not be able to be measured and studied in typical laboratory paradigms. Experiments are useful tools that can provide insights into why people behave in certain ways under various circumstances, but we are fundamentally interested in consumers' behaviors, choices, emotions, and preferences as they occur in real life.

Researchers often ask participants to reflect and introspect on their lives; they report how often they consume certain beverages, how satisfied they are with particular products, and how they view or feel about various brands (for the methodological issues involved, see Schwarz, this volume). Although such retrospective techniques can be informative, they unfortunately do not provide a means to understand dynamic processes as they unfold over time. In this chapter, we describe a technique known as ecological momentary assessment (EMA) that can capture life as it is lived. EMA has been particularly popular in the fields of health, clinical, social, and personality psychology, but has not been used as extensively in consumer behavior research. We describe what EMA is, the strengths and weaknesses of this technique, and how it can be applied to consumer behavior topics and settings.

What Is Ecological Momentary Assessment?

EMA is a broad umbrella term that can refer to a range of different real-world data collection methods, usually based on within-day sampling (including the experience sampling method; Csikszentmihalyi & Larson, 1987; Hektner, Schmidt, & Csikszentmihalyi, 2007; Larson & Csikszentmihalyi, 1983), but sometimes also includes end-of-day daily diaries. These methods share features that define the core attributes of EMA (Stone & Shiffman, 1994).

- The data are collected as participants go about their daily lives in real-world contexts. This allows for broad generalizations that extend to people's everyday lives, hence the term "ecological," which alludes to Brunswik's term, ecological validity (Brunswik, 1956).
- The questions or measures typically refer to the participants' current state, feeling, or behavior. In contrast to retrospective reports over long periods of time, EMA questions assess current or

recent states to avoid biases and errors associated with retrospection, hence the term “momentary.” However, some EMA studies employ relatively short-term recall periods, usually minutes or a few hours, in order to capture less frequent events. EMA techniques can also be applied to passive monitoring techniques such as ambulatory blood pressure, but the focus in this chapter is on self-reported experience.

- The selection of the assessed moments follows a strategic plan. Participants may be assessed each time a particular behavior occurs (e.g., purchasing a product), at randomly selected times throughout the day (to provide a representative sample of time points of their day), or some combination or variation of these two.
- To provide a more complete picture of how behaviors, thoughts, and feelings vary over time across situations, participants complete repeated assessments (e.g., five times per day for one week).

To provide an example of each of these four characteristics of an EMA study and the types of questions EMA studies can address, consider a study about Facebook use and well-being (Kross et al., 2013). Participants received text messages on their phone at five randomly selected times every day for two weeks. At each prompt, they were asked questions about their Facebook use (e.g., “How much have you used Facebook since the last time we asked?”) and their momentary feelings (e.g., “How do you feel right now?”). Note that these questions either referred to current states or reflections of behavior in the past few hours since the last notification in order to minimize recall biases. These questions did not require participants to rely on long-term retrospective recall. Moreover, to understand how Facebook use and emotions were related, the participants did not need to report to the laboratory (although they nevertheless came to campus to collect their payment and to complete exploratory trait reports). Rather, they were assessed as they lived their daily lives across a variety of situations and contexts. Presumably, these assessment points covered times when participants were working, enjoying leisure time, with friends, with strangers, by themselves, and so on, which captures a variety of contexts that would be difficult to create in a controlled laboratory setting.

The Kross et al. (2013) study design allowed the researchers to evaluate several interesting hypotheses. For example, does Facebook use predict affective well-being at a subsequent time? (Yes, they felt worse following Facebook use). Do momentary states of worry and loneliness predict increased Facebook use? (Worry did not, but loneliness did). Does gender moderate these effects? (No, and neither did global levels of self-esteem or motivations for using Facebook.)

This study is merely illustrative of the kinds of question that could be asked with an EMA design. Previous EMA studies have asked participants questions tapping a wide range of content using various assessment schedules, depending on the substantive and theoretical purposes of the study. But they all have in common the characteristics that we outlined above. Here we provide a fuller description of these attributes.

Real-World Data

One of the primary goals of EMA is to generalize findings to the real world in daily life situations. Therefore, the assessments need to occur in people’s everyday lives and should capture naturalistic contexts. Thoughts, feelings, and behaviors are highly context-sensitive, and EMA techniques allow researchers to understand these variations in ecologically valid contexts. In other words, “EMA emphasizes ecologically valid observations” (Shiffman, Stone, & Hufford, 2008).

EMA techniques can often be used to complement laboratory or experimental findings by attempting to generalize the findings to real-world situations. Here is one example: Kumar, Killingsworth, and Gilovich (2014) randomly assigned participants in an experimental setting

to think about an upcoming material (money spent on having) or experiential (money spent on doing) purchase. Those purchasers who were anticipating an experience reported more pleasant and excited feelings about their upcoming purchase than those anticipating a material purchase. In an attempt to generalize their findings to daily life, Kumar and colleagues sent text message notifications to participants at randomly selected times during the day and asked them how they were feeling, whether they were anticipating purchasing something in the near future (and if so, what kind of purchase they were considering), and how they felt about this upcoming purchase. Their findings in daily life were consistent with their findings from the experiment and thus generalized the results from the laboratory to real-world situations, which illustrates how EMA could be used to strengthen the conclusions from experimental research.

Momentary, Real-Time Assessment

Another goal of EMA is to eliminate or reduce any biases inherent in retrospective recall. A large body of research on autobiographical memory has shown that recalls are often inaccurate and often in systematically biased ways (Bradburn, Rips, & Shevell, 1987; Schwarz, this volume). For example, when people are in a negative mood, they are more likely to recall negatively valenced information (Clark & Teasdale, 1982). EMA aims to minimize these problems by removing retrospection from the reporting task – that is, participants report about their immediate states. Some EMA designs, however, allow participants to reflect on the past few hours or the current day, which introduces some retrospective bias, although this bias should be less problematic than traditional retrospective judgments over longer periods of time. In a later section, we present a more thorough discussion of biases induced by retrospection.

Repeated Assessment

In EMA studies, participants are typically assessed repeatedly over time to cover a variety of contexts and situations. The exact nature of these repeated measures will vary from study to study. Some studies will prompt participants every 30 minutes over the course of a few days (e.g., Shapiro, Jamner, Davydov, & James, 2002), whereas others might prompt participants once a day, at the end of their day, over the course of one year (e.g., Jamison et al., 2001). The repeated measures feature of EMA designs allows researchers to examine between- and within-person variations. For example, researchers may aggregate momentary reports and find that certain people report higher levels of positive affect on average compared with others. More commonly though, researchers may examine within-person relationships and examine how positive affect might vary depending on the particular situation.

History of Ecological Momentary Assessment

Before we outline the historical roots of EMA, it is important to reiterate the wide scope of methods that fall under the term EMA. These techniques include experience sampling (Larson & Csikszentmihalyi, 1983), ambulatory physiological monitoring (Kop et al., 2001), and other variants typically used by health and clinical psychologists. They all share the particular characteristics outlined above and, hence, can be grouped into the EMA category.²

The term “ecological momentary assessment” was first coined in 1994 (Stone & Shiffman, 1994), but some of the methods existed prior to 1994. Dating as far back as the 1970s, the technique of experience sampling was created to study experiences of flow in everyday life (Larson & Csikszentmihalyi, 1983). These studies were truly innovative because they were the first to utilize

a random sampling of time points. Experience sampling studies were initially conducted with “beepers” that alerted participants to complete a questionnaire with paper and pencil. Technology has improved considerably since that time, which allows for more complicated and intricate designs. Importantly, the advent of electronic diaries and smartphones allowed for precise time/date stamping, an essential component for obtaining accurate sampling completion rates (Stone, Shiffman, Schwartz, Broderick, & Hufford, 2002, 2003).

EMA and related techniques were initially inspired by the work of Egon Brunswik (1956), who stressed the importance of the environment on social and cognitive functioning and not merely the organism itself. His term ecological validity referred to the notion that a representative design should include stimuli that are representative of the environment of the person. Representative design is achieved by administering questions to participants over time across a variety of situations. This theoretical framework has guided the various different types of EMA study designs.

Autobiographical Memory and Limitations of Recall

To gain a better appreciation for EMA techniques, it is important to outline the weaknesses of autobiographical memory. In this section, we expand on the problems and pitfalls of retrospective recalls. Retrospective recall is systematically biased. People are not able to perfectly and directly retrieve information from their past when asked. Rather, much of memory is a reconstruction based on fragmentary inputs and heuristics (Belli, 1998; Bradburn et al., 1987; Schwarz & Oyserman, 2001). The memories and pieces of information that are more likely to be recalled are the ones that are emotionally salient and unique. The current emotional state can also influence which memories are retrieved.

Retrospective questions not only require people to retrieve specific memories and feelings, but they also often ask people to aggregate or provide some average or typical estimate. For example, people might be asked how satisfied they have been this past month with their recent purchase of an iPhone. People are not able to recall every experience they have had with their phone this past month and then aggregate their average level of satisfaction with the product. Rather, people rely on various heuristics or mental shortcuts to answer such questions.

One particularly noteworthy heuristic is the availability heuristic (Tversky & Kahneman, 1973). If a particular instance or memory comes to mind and is available, people believe this memory is more typical and frequent than it actually is. In many cases, this is a fine strategy because it is not possible to replay every event from your life before answering a question. If an example comes to mind, it probably occurs frequently because rare events would be harder to find. This strategy could backfire, though, if the recalled memory is quite memorable but atypical. For example, if someone were to recall a particularly frustrating episode with their phone, they might report being less satisfied with their phone even though they were quite satisfied with their phone much of the time during the past month.

Related to this bias, people are also influenced by their momentary state at the time of retrieval. For example, when people are in a negative mood when answering a recall report, they are more likely to bring to mind negative experiences or states (e.g., Kihlstrom, Eich, Sandbrand, & Tobias, 2000). Another example of a cognitive trick the mind plays on us is exemplified in study of medically induced pain. In this study, colonoscopy patients who were awake during the procedure recorded their level of pain on a minute-to-minute basis (Redelmeier, Katz, & Kahneman, 2003). After the procedure, they were asked to summarize their level of pain during the colonoscopy. It was found that summary reports were influenced by the peak or worst pain during the procedure and by pain levels at the very end of the procedure (known as the “peak-end heuristic”). This illustrates two ways that retrospective reports can be biased in ways that would lead to incorrect conclusions about people’s experiences.

Strengths and Weaknesses of Retrospective and Real-Time Reports

Because retrospective reports may be biased in various ways, it is instructive to compare a retrospective report with aggregated EMA data. The surprising conclusion from these studies is that, despite the biases in retrospection, there are strengths and weaknesses with both approaches. Aggregated EMA data provide more accurate estimates of the rate of occurrence of the phenomena of interest (events, behaviors) and/or average levels of continuous phenomena (affect, symptom levels), whereas retrospective reports may provide more accurate estimates of future intentions of the behavior. In this section, we review studies that have compared these two approaches.

To illustrate the strength of aggregated EMA data, suppose a researcher is interested in knowing the average amount of time people spend watching television per week in various regions of the country. Participants in each region could complete multiple reports throughout the week by recording how much television they have watched since the prior assessment. At the end of the week, they additionally recall and estimate how much time they spent watching television. The aggregated reports throughout the week should provide a more accurate estimate of their total number of hours spent watching television, whereas the retrospective report should be biased for the various reasons described above. In this example, aggregated EMA data provide researchers with more accurate information about time spent watching television that could provide valuable information to marketing researchers who might want to know how to advertise certain television products.

Understanding whether retrospective reports lead to overestimates or underestimates of the phenomenon of interest is clearly important if one would like to use them in place of more granular data such as EMA, as well as to understand how memory works. In some studies that have compared aggregated EMA scores and recall reports, researchers have found these two scores to be quite similar (Shrier, Shih, & Beardslee, 2005). In other studies, recalls have been lower than the aggregated EMA scores (Carney, Tennen, Affleck, Del Boca, & Kranzler, 1998; Litt, Cooney, & Morse, 2000). In most cases, however, the recall report is higher than the daily or momentary average reports (e.g., van den Brink, Bandell-Hoekstra, & Abu-Saad, 2001), especially in the case of negatively valenced phenomena. People often remember the target experience as having been more frequently occurring, more intense, and longer lasting.

The correlations between the aggregated momentary reports and the recall have ranged from .20 to .70 (Shiffman et al., 2008). There are several possible reasons for the wide range of associations. One could be due to individual differences. For example, neurotic people overestimate how much stress they experience (Feldman-Barrett, 1997), whereas extraverted people overestimate how much positive affect they experience (van den Brink et al., 2001). Another reason for the variation in correlations between aggregated EMA reports and recalls is because the target experience might vary. People can more accurately recall a target experience if that target was relatively stable over time (Stone, Schwartz, Broderick, & Shiffman, 2005). There are other potential reasons for this variation, and this is an ongoing area of research.

To illustrate the strength of retrospective reports over aggregated EMA data, some studies have found that retrospective reports can be better predictors of future behavior or intention. In the colonoscopy study described above, some participants were randomly assigned to a condition to receive a slightly longer colonoscopy but with less intense pain at the very end (Redelmeier et al., 2003). Compared with the participants in the other condition, these participants experienced more pain in absolute terms because their colonoscopy lasted longer. But these participants recalled their colonoscopy as less painful because the ending was less painful. Importantly, they were more likely to return for a follow-up colonoscopy. In other words, recalled colonoscopy pain was a better predictor of future behavior than the more “accurate” aggregated average level of concurrently assessed pain from the EMA reports.

This pattern of findings has been replicated in some other settings. In a study involving romantic couples, participants were asked to keep daily records of the satisfaction with their relationship and they were also asked to make a retrospective summary of their relationship satisfaction (Oishi & Sullivan, 2006). The retrospective summary was a better predictor of a future breakup than the aggregated dissatisfaction from the daily reports. Similarly, participants who were on a spring-break trip recorded their satisfaction and affect during their trip at randomly selected time points (Wirtz, Kruger, Scollon, & Diener, 2003). They also evaluated how happy they were with their vacation after it ended and rated their intention of going on a similar trip in the future. Consistent with the previous examples, the recall report was a better predictor of future intention than the average momentary reports during the vacation.

These studies highlight a key point about these different reports. EMA methods perform better at describing experience *at the time it occurred*, whereas recall reports perform better at predicting future behavior. If the goal is to understand life as it is lived in real time and to summarize experiences in ways that map closely onto the actual experience, EMA methods should be used. On the other hand, if the goal is to understand how people might behave in the future, recall reports might be a better option, because an experience that is not remembered will not influence the decision. Researchers must carefully think about their question of interest and their theoretical framework when deciding which method to use.

The Uses of Ecological Momentary Assessment

EMA techniques can serve multiple purposes depending on the goal of the study. In this section, we describe four of the key uses of EMA: Characterizing individual differences, describing natural history, assessing contextual associations, and documenting temporal sequences.

Individual Differences

Historically, one common way of assessing individual differences has been through trait reports, particularly in personality psychology (Funder, 1995). Trait reports typically ask participants to evaluate their lives in some manner, but there are systematic biases inherent in such recall judgments (see Schwarz, 2012 for a review). One remedy to this issue was suggested by Fleeson (2001), who argued that personality could be measured by aggregating daily states. Daily or momentary states are contextualized reports that are not biased (or are less biased) by recall errors. When aggregated, certain people behave more often in extraverted or introverted ways, for example. This can be a more reliable and more valid way to measure individual differences. This technique is not restricted to personality topics, of course, and topics pertinent to marketing, for instance, would appear to be relevant here, such as the characterization of individual purchasing preferences.

In one study, Newman, Schwarz, Graham, and Stone (in press) aggregated daily and momentary reports of meaning and purpose in life. They found that conservatives, in comparison with liberals, reported greater daily and momentary states of meaning and purpose in life. A similar aggregation technique was used to categorize students based on daily diary reports of how they spent their time over the course of one week (Nonis, Philhours, & Hudson, 2006). They found that campus-centered students spent more time studying, watching TV, and participating in various other leisure activities and less time working at a paid job than life-centered students. Not surprisingly, campus-centered students earned higher GPAs than life-centered students. These studies illustrate some of the ways EMA techniques could be used to characterize and study individual differences in natural contexts.

Natural History

Natural history refers to the trend of variables over time. Some researchers are interested in the within-person variation, where time is the predictor variable. For example, participants in one study were asked to recall a recent experiential or material gift that they received for Christmas (Weidman & Dunn, 2015, Study 2). Next, they received three to five text messages each day for two weeks, beginning on Christmas Day. They were asked how much their gift was contributing to their happiness at that moment. The researchers were able to examine how momentary happiness changed over time, and they found a slight negative decrease over time. This decrease was not moderated by the type of experience they were asked to recall at the outset of the study, suggesting that the happiness people derive from gifts decreases as time moves on, regardless of the type of gift received. This illustrates one way in which natural trends over time can inform researchers about consumer behavior over time.

Contextual Associations

A strength of EMA techniques is their ability to examine contextual associations within a given individual, which allows researchers to determine, *for an individual*, how context impacted them. In these analyses, researchers examine the co-occurrence of two states, experiences, or events. In contrast to the natural history purpose, time is not of primary interest in contextual associations, but serves as a forum where states and experiences interact. In one study of college undergraduates, for example, participants were asked each day over the course of two weeks to report the extent to which they maximized while selecting who they spent their time with (Newman, Schug, Yuki, Yamada, & Nezlek, 2018). Maximizing typically refers to an individual difference characteristic describing a tendency to find the very best option in a given set of choices (Schwartz et al., 2002). The concept of maximizing has been applied to job selection situations (Iyengar, Wells, & Schwartz, 2006) as well as various consumer decision situations (e.g., Dar-Nimrod, Rawn, Lehman, & Schwartz, 2009; Lai, 2011). Newman and colleagues (2018) referred specifically to maximizing in the domain of friendship selection, and they examined within-person relationships between daily states of maximizing (e.g., “No matter how satisfied I was with my social group today, I was always on the lookout for better opportunities to spend time with other friends”) and daily states of well-being. They found that people were less satisfied with their lives and experienced greater negative affect on days when they maximized in selecting friends.

Highlighting the importance of the EMA technique for gathering information, the results showed that the within-person relationships between daily maximizing and daily well-being were orthogonal to the between-person relationships documented in previous research (e.g., Iyengar et al., 2006; Schwartz et al., 2002). In general, within-person relationships are mathematically distinct from between-person relationships (Nezlek, 2001), and they often represent distinct psychological processes (Affleck, Zautra, Tennen, & Armeli, 1999). Consider the association between exercise and heart attack risk. At a between-person or individual difference level of analysis, you would find that people who frequently exercise are less likely to experience a heart attack (a negative relationship). In contrast, at a within-person level of analysis, you would find that, on days or at times when people exercise, they are more likely to experience a heart attack (a positive relationship; Curran & Bauer, 2011). The within-person level of analysis can therefore considerably extend many areas of research and provide unique insights. The ability to understand within-person contextual associations is one of the main uses and a key strength of EMA techniques.

Temporal Sequences

Finally, EMA data allow researchers to examine putative causes and effects of certain events and experiences that play out over time. Whereas historical trends are concerned with the trends over

time, temporal sequences refer specifically to lagged associations between two or more events and outcomes. In a study about quitting smoking, participants were tracked over time and they recorded any relapses that occurred, along with their momentary affect and self-efficacy (Shiffman et al., 1997). Relapses occurred at different times, but the researchers were able to study how affect and self-efficacy changed over time after relapses occurred. In a study on food consumption, participants completed daily diary reports by recording what they ate at each meal and snack time during the day (Khare & Inman, 2006). This allowed the researchers to examine how food intake at one time influenced what people ate at a later time. Similar techniques could be used in other consumer behavior domains to study how regret about a purchase may change over time, for instance.

Summary

EMA studies address one or more of these four purposes. As a hypothetical example, a within-person contextual association between eating candy and regret may be moderated by an individual difference measured as the aggregate of daily or momentary reports of eating candy. That is, regarding the average within-person association across all participants, people might feel regretful after eating candy. This relationship might be moderated by how much candy people eat on average, such that the positive within-person relationship between eating candy and regret might be stronger among people who do not eat candy very often. This would be combining purposes of individual differences and contextual associations.

This section highlighted some potential uses of EMA and illustrated them with select examples. Clearly, many hypotheses concerning the temporal dynamics and interplay of variables over time can be tested with EMA. The ability to ask questions about dynamic processes is akin to watching a movie that plays out over time. In contrast, recall reports provide a snapshot or photograph, which may nevertheless be very useful. Yet we argue that addressing the temporal sequence of events and the covariations among variables measured over time can extend and enhance the research agenda in many areas and provide nuanced or novel understandings that were not possible with other techniques.

Ecological Momentary Assessment Designs and Approaches

A key feature of EMA is that EMA designs typically sample multiple time points, whereas other more traditional designs typically sample a single time point. The assumption is that a single response in recall assessments covers the entire experience of interest encompassed by the specified recall period (e.g., one's entire life, the past month, during a particular experience). The timing of that single assessment is typically not a primary concern. In EMA, in contrast, selecting the timing of questions is critical because the repeated measures should constitute either a representative sample of the experience of interest or be specifically selected for testing hypotheses about particular types of events or experiences.

There are two primary types of assessment schemes in EMA designs: event-based sampling and time-based sampling (Shiffman, 2007; Wheeler & Reis, 1991). Event-based schemes focus on a particular event that may occur repeatedly, such as eating or drinking (Todd et al., 2005). Time-based schemes typically do not focus on specific, discrete events, but sampling is based on investigator-specified temporal rules (e.g., random sampling or at particular hours of the day). There are strengths and weakness to each type of sampling scheme, and researchers need to carefully determine which schemes meet their study's needs.

Event-Based Monitoring

Consumer behavior researchers are often interested in specific behaviors, such as eating, watching television, purchasing a product, and so on. When one wishes to ensure that such events are captured

by the study design, then event-based schemes are well-suited. Researchers could ask participants to complete an assessment each time they eat food at a restaurant, use a particular app on their phone, or purchase something online, for instance. One of the earliest versions of this type of study was the Rochester Interaction Record (Reis & Wheeler, 1991), in which participants completed an assessment each time they engaged in a social interaction lasting more than 10 minutes. Similar techniques could be applied to study professional networking and emotional responses following these interactions. In these studies, participants must decide when the defined event has occurred and they must remember to complete an assessment. This means that the researchers must clearly define what constitutes an event and determine if participants are capable of self-identifying them (as reporting will be initiated by participants). Although this may seem rather straightforward, many ill-defined situations are likely to arise in daily life. If a study's goal was to examine emotions during food ordering at a restaurant, then does it count if one person orders, but another person pays? These types of ambiguous situations must be carefully considered by the researcher before the study begins.

One limitation of event-based schemes is that compliance is often very difficult or even impossible to gauge. If participants are asked to record an assessment each time an event occurs, there is no way of knowing how many events the participants missed or failed to record. In some cases, personal technological devices, such as smartphones, may be able to record how often a person is in a particular location (e.g., grocery store or a restaurant), which could provide an estimate for how many assessments the participant should have recorded. Another limitation of event-based schemes is that the assessments will rarely provide a representative sample of time points during the duration of the study.

Time-Based Designs

In contrast to event-based monitoring, time-based designs focus on sampling time points rather than events. Participants are not asked to record an entry every time an event occurs; rather, participants receive notifications and complete reports at certain times selected by the researcher. Various different sampling schemes may be used, depending on the frequency of the variable of interest.

In a fixed-interval schema, participants receive notifications at regularly spaced intervals – for example, every 90 minutes. A unique version of this type of design is a daily diary study (discussed in detail below) in which the assessments are distributed every 24 hours, typically at the end of the day. One advantage of fixed-interval designs is that researchers can examine lagged effects from one time to the next, as such analyses often assume and require the intervals between data points to be equal.

Another time-based option is a variable schedule, which refers to a random selection of time points throughout the day or period of study. This schedule ensures that the selection of time points constitutes a random sample of participants' states. One commonly used variation of this technique is called stratified random sampling. In this design, researchers create time windows (e.g., 9:00 a.m. to 10:30 a.m., 10:30 a.m. to 12:00 p.m., etc.) and randomly select times from these windows. This ensures that assessments will cover the entire day while still providing random sampling. For example, Verduyn et al. (2015) sent participants text messages five times per day between 10:00 a.m. and midnight. Time windows of 168 minutes were used so that only one text message was sent within each window, thereby ensuring text messages were received throughout the whole day. Random sampling (with high participant compliance) has the unique characteristic that it can produce unbiased estimates of the phenomenon studied (just as randomly sampling individuals creates an unbiased view of the population from which they were drawn). Of course, low rates of compliance can create biased estimates of the phenomenon, and so researchers should carefully consider the trade-off between participant compliance and desired density of assessments (see the section on compliance below).

In each of these types of designs, it is important to consider the frequency of occurrence of the construct of interest. If variable X only changes every four or five hours, it would not make sense to assess participants every 45 minutes. Of course, how frequently a variable varies and when it changes might not be known until the study is conducted, hence the importance of pilot testing. A common assessment strategy is to prompt participants three to five times per day, although some studies have successfully sent participants notifications 20 times per day (Goldstein, Jamner, & Shapiro, 1992; Kamarck et al., 1998). In other cases, participants could be notified twice a day, once in the morning and once in the evening (e.g., Thrash, Elliot, Maruskin, & Cassidy, 2010).

Combination Designs

A combination of event-based and time-based designs can be fruitfully employed to study various topics. For example, if researchers wish to study the types of emotions that covary with difficult daily decisions, they might choose to use a combination of these designs. In this example, participants would complete an assessment each time they faced a daily decision and they would record their current emotions. This information alone might not be terribly meaningful though, if there are no control states with which to compare emotional experiences. Therefore, participants might also receive notifications at randomly selected times during the day when they are not faced with a decision, and these times would serve as a baseline condition to compare emotional states. These types of designs have sometimes been called case-control designs (Maclure & Mittleman, 2000).

Use of Recall in EMA

Even though the goal of EMA studies is to diminish the biases of recall judgments, some EMA designs involve a level of retrospection. For example, Hofmann, Wisneski, Brandt, and Skitka (2014) administered text message notifications five times per day at random times over the course of three days and asked participants if they committed, were the target of, witnessed, or learned about a moral or immoral act within the past hour. In social interaction event-based designs discussed above (Reis & Wheeler, 1991), participants were asked to recall the social interaction that had just occurred, which also relies on recall. In many cases, EMA techniques simply try to minimize recall biases by asking participants to recall very recent experiences. However, we also believe that more research about this technique, and specifically whether or not it introduces significant bias, is in order.

Daily Diaries: A Special Case of EMA

As mentioned in the introduction, daily diary studies (Bolger, Davis, Rafaeli, 2003; Nezlek, 2012) require short-term retrospection and are considered by some scientists to be a specialized form of EMA. These studies have been popular techniques among health and clinical researchers and have been gaining in popularity among social and personality psychologists, and they could be readily applied to many consumer behavior domains as well. Part of the reason these techniques have been popular is that administration is relatively easy and participant burden is fairly low (compared with many other sampling schedules). Because daily diaries are distributed once per day at the end of the day, they would be considered a fixed-interval design with a retrospective coverage strategy. Even though they are not momentary reports, they are administered repeatedly and provide a means to examine dynamic processes.

Despite their advantage of allowing dynamic processes to be studied over time, diary studies carry some of the biases inherent in retrospective reports. Moreover, any type of within-day variations cannot be captured with diary studies. Mood, stress, pain, and fatigue may vary throughout the day and could even influence the daily reports. In fact, peak and end reports of pain uniquely predicted daily recalls of pain (Schneider, Stone, Schwartz, & Broderick, 2011).

One related technique worth mentioning in this section is the day reconstruction method (DRM), because it attempts to obtain EMA-like data using retrospective methods (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004). As the name suggests, participants are asked to reconstruct their day by listing and describing each episode that occurred during each hour of the day. They are often asked various questions about their thoughts and feelings during each episode. This tool allows researchers to examine how long people were engaged in various activities, as well as how they were feeling during the activities. Some recent studies have shown that the end-of-day reconstruction can reasonably approximate what is observed in EMA data (Dockray et al., 2010; Stone et al., 2006).

The DRM could be fruitfully used to examine how much time people spend shopping, watching TV, working, commuting to work, and so on. In fact, the Princeton Affect and Time Use Survey and the American Time Use Survey have employed a version of the DRM by asking a representative sample of Americans to report how they felt during randomly selected specific activities they had engaged in during the day. For example, Stone and Schneider (2016) found that people felt stressed and tired when commuting to and from work, and they rated commuting as less meaningful than other activities during the day.

One disadvantage of DRM, though, is that reconstruction can take up to 45 minutes, which means DRM is not an ideal tool for repeated measurement as it becomes too burdensome on participants. This rules out the possibility of examining within-person relationships in daily life. Nevertheless, the DRM could provide novel insights into consumer habits and activities and how they compare with other types of daily experiences.

Measurement Considerations for EMA Self-Reports

EMA differs from traditional self-report techniques, which means there are several measurement issues worth considering. For instance, many scales and questions were initially designed to refer to one's life as a whole. At the very least, these items would need to be rephrased to assess a momentary state. In some cases, the variable may not make sense when converted to a momentary state. Pilot-testing and cognitive interviews (Willis, 2005) are often required before questions are used in EMA that have typically been used in global self-report techniques. Relatedly, the amount of variation of the construct needs to be considered in the wording of the items. If a particular type of decision is only made once or twice a day, the wording of the questions should reflect this.

Because participants answer the same questions repeatedly over time, it is important to make sure that the administration of the questions goes smoothly so that the participants are not annoyed. If the questions or the application device create frustration, the resulting negative mood could influence the responses. It is also important to limit the number of questions and the number of notifications received as much as possible so as not to unduly burden the participants.

The specific device that participants use to answer questions throughout the day should be carefully considered. In the earlier days of EMA, researchers would frequently provide participants with a personal digital assistant (PDA; e.g., Palm Pilot) with a preprogrammed schedule of notifications. Participants would often come to the lab to collect the device and would learn how to respond to questions. Although this would often be expensive for researchers, the advantage was that participants would respond to answers using the exact same type of device. Now that smartphones are widely used, researchers have begun to use applications that participants can download on their phones (so-called bring-your-own-device). Although this is more convenient, it is important to pretest the questions on different types of phones to make sure the questions are displayed in similar manners across the different devices. There are also a host of issues concerning the performance of smartphone apps when operating systems are modified by the telecom provider or even when a phone is rebooted, but these considerations go beyond the scope of this paper. Two websites allow users to design their own EMA apps using prepared toolboxes of EMA components (i.e., different

types of response options and branching routine): www.experiencesampler.com and www.pacoapp.com. Commercial providers of EMA services and data collection are also available.

Most EMA studies ask participants questions that can be quantified, which allows researchers to easily analyze the data. Qualitative data could also be collected though by asking participants open-ended questions which could be analyzed with various text analysis techniques, such as LIWC (Pennebaker, Francis, & Booth, 2001) or semantic analyses or latent Dirichlet allocation (Iliev, Dehghani, & Sagi, 2014). Of course, if the data set is manageable, research assistants can code responses for various features or topics. EMA techniques could also be used to record auditory sound bites or snippets of daily conversations (Mehl, Pennebaker, Crow, Dabbs, & Price, 2001). Given the widespread use of smartphones, participants could be asked to take pictures during their day when prompted and these could be analyzed in various ways.

Methodological Considerations in EMA Studies

Reactivity

One methodological concern of EMA studies is reactivity, defined as the influence of the instrument on the actual behavior or experience. Of course, any self-report question or instrument will influence the participant in some way (for a review, see Schwarz, 1999, and this volume). In EMA, the issue of reactivity refers more specifically to the effect of self-monitoring. Some studies have shown that self-monitoring can influence the enjoyment of an experience (e.g., Schooler, Ariely, & Loewenstein, 2003), although this method of repeated assessments differed from typical EMA procedures. In some EMA studies, reactivity effects were found when the participants were specifically trying to change the target behavior (Korotitsch & Nelson-Gray, 1999).

In most examinations of reactivity in EMA studies, however, very few or no reactivity effects have been found (e.g., Hufford & Shiffman, 2002). In one careful examination of reactivity effects, Stone et al. (2003) randomly assigned participants experiencing high levels of pain to complete either 0, 3, 6, or 12 assessments per day. They found no systematic evidence of reactivity on pain reports.

A related concern is that repeated assessments might create a burden on the participants which could influence their reports gradually over time. For example, if the assessments contain an exorbitant number of questions, participants might get tired of answering them and begin to carelessly respond later in the study. In a recent examination of this possibility, Shrout and colleagues (2018) found a very weak negative temporal trend of numerous variables over time in four daily diary studies. More fine-grained analyses showed that there was an initial spike in reported measures at the very beginning of the study, but everything leveled out after this initial spike. It is unclear from their analyses, however, whether the initial elevation bias had any influence on any substantive within-person relationships, which tend to be the primary interest in most EMA and daily diary studies. In fact, in a review of dozens of daily diary studies, Nezlek (2012) reported no effect of time on any theoretically meaningful within-person relationships. That is, within-person relationships did not differ from one week to the next or from the beginning of the study to the end. Although reactivity appears to be of only modest concern for most EMA designs, it is certainly possible that reactivity effects may be observed. Researchers should carefully consider these possibilities when designing and analyzing studies.

Compliance

Another methodological consideration in EMA studies is compliance, which typically refers to the rate of completed assessments. When participants fail to complete several or many of the assessments, it could bias the sample. The completed assessments may not be truly representative of the time

points in the study. For example, if participants fail to complete assessments when they are feeling particularly sad or depressed, the sample of time points would only be representative of time points when the participants are happy, biasing the assessment of mood.

One compliance issue that used to be a primary concern was known as hoarding. Before the widespread use of smartphones and laptops, participants were often asked to complete assessments on paper. If participants forgot to complete reports at the appropriate time, they would sometimes complete multiple reports at the same time, presumably trying to recall their thoughts, feelings, and behaviors from previous days. Stone et al. (2002) found that 90 percent of completed entries were not completed at the time they should have been completed! With the use of electronic diaries and notifications, this is less problematic because everything is time-stamped automatically, so the researcher is at least aware of completion rates. Nevertheless, participants may still forget to complete entries, and so compliance is an issue that requires serious attention.

Analysis of EMA Data

The statistics used to analyze EMA data are typically more complicated than the analyses used in most experimental or cross-sectional designs. In typical EMA data sets, participants have provided many observations. This means that individual observations are not independent, a requirement needed for ordinary least squares regression. Repeated measures ANOVA designs are also not recommended for a few reasons, most notably because the number of completed observations varies among participants.

The most common way to analyze EMA data is to use multilevel modeling, also known as random effects models, hierarchical linear modeling, or multilevel random coefficients modeling. Although the models can be a bit more complicated than the general linear model (e.g., regression, ANOVA), the general framework and logic of multilevel modeling are fairly straightforward. EMA data are typically nested, such that observations are clustered or grouped within individuals. In a daily diary study, for example, participants complete multiple daily reports, which means that daily reports are clustered within individuals. The days are presumably a representative random sample from the larger population of days in the year, and the participants are presumably a random sample from the larger population of people. Multilevel modeling takes the nested data structure into account and allows for these generalizations at multiple levels of analysis. A thorough overview of multilevel modeling is beyond the scope of this review. Detailed descriptions, practical guidelines, and applications of multilevel modeling for EMA designs can be found in several articles (Nezlek, 2001; Schwartz & Stone, 2007) and books (Bolger & Laurenceau, 2013; Nezlek, 2012).

In addition to multilevel modeling, there are several less cumbersome ways to analyze nested data sets, such as generalized estimating equations and cluster-robust standard errors. These techniques correct for the biased standard errors, but the types of questions they can address tend to be limited (for a review, see McNeish, Stapleton, & Silverman, 2017). As an analogy, multilevel modeling is similar to watching a movie in color, whereas other techniques are similar to watching a movie in black and white. Different analytic techniques can be selected depending on the goals of the research.

Conclusion

EMA techniques can be usefully applied to help researchers understand consumers' decisions, emotions, preferences, and attitudes in daily life. Contextual influences unfold in dynamic ways over time and can be captured with EMA. Processes and effects discovered in laboratory settings can be tested in real-world settings to examine the generalizability of the findings. As Joe McGrath (1982) noted, "all research strategies and methods are seriously flawed." Therefore, it is often prudent to combine

techniques (that each have different flaws) to improve our understanding of consumer-related phenomena. EMA techniques are one method that can be used in a researcher's toolkit. We hope this review can serve as a guide to help researchers in consumer behavior in familiarizing themselves with EMA techniques, which are currently underutilized in consumer research.

Notes

- 1 This work was supported, in part, by a grant from the National Institute on Aging (grant number R01 AG042407; AAS, PI).
- 2 Other techniques, such as daily diaries (Bolger, Davis, & Rafaeli, 2003; Nezlek, 2012) and interpersonal interaction diaries (Reis & Wheeler, 1991), were not initially included in the definition of EMA, but some researchers have morphed the definition of EMA to include them as a type of EMA. We provide examples of daily diaries throughout this chapter, given their previous use and potential applications in consumer behavior research, and we note that diary studies have sometimes been considered a special case of EMA in a later section.

References

- Affleck, G., Zautra, A., Tennen, H., & Armeli, S. (1999). Multilevel daily process designs for consulting and clinical psychology: A preface for the perplexed. *Journal of Consulting & Clinical Psychology, 67*(5), 746–754.
- Belli, R. (1998). The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory, 6*, 383–406.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616.
- Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford.
- Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236*(4798), 157–161.
- Brunswick, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.
- Carney, M. A., Tennen, H., Affleck, G., Del Boca, F. K., & Kranzler, H. R. (1998). Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time “electronic interviews.” *Journal of Studies on Alcohol, 59*(4), 447–454.
- Clark, D. M., & Teasdale, J. D. (1982). Diurnal variation in clinical depression and accessibility of memories of positive and negative experiences. *Journal of Abnormal Psychology, 91*(2), 87–95.
- Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583–619.
- Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience sampling method. *Journal of Nervous & Mental Disease, 775*, 526–536.
- Dar-Nimrod, I., Rawn, C. D., Lehman, D. R., & Schwartz, B. (2009). The maximization paradox: The costs of seeking alternatives. *Personality & Individual Differences, 46*(5–6), 631–635.
- Dockray, S., Grant, N., Stone, A. A., Kahneman, D., Wardle, J., & Steptoe, A. (2010). A comparison of affect ratings obtained with ecological momentary assessment and the day reconstruction method. *Social Indicators Research, 99*(2), 269–283.
- Feldman-Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality & Social Psychology Bulletin, 23*(10), 1100–1110.
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality & Social Psychology, 80*(6), 1011–1027.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*(4), 652–670.
- Goldstein, I. B., Jamner, L. D., & Shapiro, D. (1992). Ambulatory blood pressure and heart rate in healthy male paramedics during a workday and a nonworkday. *Health Psychology, 11*(1), 48–54.
- Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage.
- Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science, 345*(6202), 1340–1343.

- Hufford, M. R., & Shiffman, S. (2002). Methodological issues affecting the value of patient-reported outcomes data. *Expert Review of Pharmacoeconomics & Outcomes Research*, 2(2), 119–128.
- Iliev, R., Dehghani, M., & Sagi, E. (2014). Automated text analysis in psychology: Methods, applications, and future developments. *Language & Cognition*, 7(02), 265–290.
- Iyengar, S. S., Wells, R. E., & Schwartz, B. (2006). Doing better but feeling worse. Looking for the “best” job undermines satisfaction. *Psychological Science*, 17(2), 143–150.
- Jamison, R. N., Raymond, S. A., Levine, J. G., Slawsby, E. A., Nedeljkovic, S. S., & Katz, N. P. (2001). Electronic diaries for monitoring chronic pain: 1-year validation study. *Pain*, 91(3), 277–285.
- 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(5702), 1776–1780.
- Kamarck, T. W., Shiffman, S. M., Smithline, L., Goodie, J. L., Paty, J. A., Gnys, M., & Jong, J. Y.-K. (1998). Effects of task strain, social conflict, and emotional activation on ambulatory cardiovascular activity: Daily life consequences of recurring stress in a multiethnic adult sample. *Health Psychology*, 17(1), 17–29.
- Khare, A., & Inman, J. J. (2006). Habitual behavior in American eating patterns: The role of meal occasions. *The Journal of Consumer Research*, 32(4), 567–575.
- Kihlstrom, J. F., Eich, E., Sandbrand, D., & Tobias, B. A. (2000). Emotion and memory: Implications for self-report. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 81–100). Mahwah, NJ: Erlbaum.
- Kop, W. J., Verdino, R. J., Gottdiener, J. S., O’Leary, S. T., Bairey Merz, C. N., & Krantz, D. S. (2001). Changes in heart rate and heart rate variability before ambulatory ischemic events. *Journal of the American College of Cardiology*, 38(3), 742–749.
- Korotitsch, W. J., & Nelson-Gray, R. O. (1999). An overview of self-monitoring research in assessment and treatment. *Psychological Assessment*, 11(4), 415–425.
- Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., et al. (2013). Facebook use predicts declines in subjective well-being in young adults. *PLoS One*, 8(8), e69841.
- Kumar, A., Killingsworth, M. A., & Gilovich, T. (2014). Waiting for Merlot: Anticipatory consumption of experiential and material purchases. *Psychological Science*, 25(10), 1924–1931.
- Lai, L. (2011). Maximizing and customer loyalty: Are maximizers less loyal? *Judgment & Decision Making*, 6, 307–313.
- Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. In H. T. Reis (Ed.), *Naturalistic approaches to studying social interaction. New directions for methodology of social and behavioral sciences* (pp. 41–56). San Francisco, CA: Jossey-Bass.
- Litt, M. D., Cooney, N. L., & Morse, P. (2000). Reactivity to alcohol-related stimuli in the laboratory and in the field: Predictors of craving in treated alcoholics. *Addiction*, 95(6), 889–900.
- Maclure, M., & Mittleman, A. M. A. (2000). Should we use a case-crossover design? *Annual Review of Public Health*, 21(1), 193–221.
- McGrath, J. E. (1982). Dilemmas: The study of research choices and dilemmas. In J. E. McGrath, J. Martin, & R. A. Kulka (Eds.), *Judgment calls in research* (pp. 69–102). Beverly Hills, CA: Sage.
- McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, 22(1), 114–140.
- Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, 33(4), 517–523.
- Newman, D. B., Schug, J., Yuki, M., Yamada, J., & Nezlek, J. B. (2018). The negative consequences of maximizing in friendship selection. *Journal of Personality & Social Psychology*, 114(5), 804–824.
- Newman, D. B., Schwarz, N., Graham, J., & Stone, A. A. (in press). Conservatives report greater meaning in life than liberals. *Social Psychological & Personality Science*.
- Nezlek, J. B. (2001). Multilevel random coefficient analyses of event- and interval-contingent data in social and personality psychology research. *Personality & Social Psychology Bulletin*, 27(7), 771–785.
- Nezlek, J. B. (2012). Diary methods for social and personality psychology. In J. B. Nezlek (Ed.), *The SAGE library in social and personality psychology methods*. London: Sage.
- Nonis, S. A., Philhours, M. J., & Hudson, G. I. (2006). Where does the time go? A diary approach to business and marketing students’ time use. *Journal of Marketing Education*, 28(2), 121–134.
- Oishi, S., & Sullivan, H. W. (2006). The predictive value of daily vs. retrospective well-being judgments in relationship stability. *Journal of Experimental Social Psychology*, 42(4), 460–470.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count (LIWC)*. Mahwah, NJ: Erlbaum.
- Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: A randomized trial. *Pain*, 104(1–2), 187–194.

- Reis, H. T., & Wheeler, L. (1991). Studying social interaction with the Rochester Interaction Record. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 269–318). San Diego, CA: Academic Press.
- Schneider, S., Stone, A. A., Schwartz, J. E., & Broderick, J. E. (2011). Peak and end effects in patients' daily recall of pain and fatigue: A within-subjects analysis. *Journal of Pain, 12*, 228–235.
- Schooler, J. W., Ariely, D., & Loewenstein, G. (2003). The pursuit and assessment of happiness can be self-defeating. *The psychology of economic decisions, 1*, 41–70.
- Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality & Social Psychology, 83*(5), 1178–1197.
- Schwartz, J. E., & Stone, A. A. (2007). The analysis of real-time momentary data: A practical guide. In A. Stone, S. Shiffman, A. Arienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 76–113). New York: Oxford University Press.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *The American Psychologist, 54*(2), 93–105.
- Schwarz, N. (2012). Why researchers should think “real-time”: A cognitive rationale. In M. R. Mehl & T. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 22–42). New York: Guilford.
- Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication and questionnaire construction. *American Journal of Evaluation, 22*, 127–160.
- Shapiro, D., Jamner, L. D., Davydov, D. M., & James, P. (2002). Situations and moods associated with smoking in everyday life. *Psychology of Addictive Behaviors, 16*(4), 342–345.
- Shiffman, S. (2007). Designing protocols for ecological momentary assessment. In A. S. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 27–53). New York: Oxford University Press.
- Shiffman, S., Hickcox, M., Paty, J. A., Gnys, M., Kassel, J. D., & Richards, T. J. (1997). The abstinence violation effect following smoking lapses and temptations. *Cognitive Therapy & Research, 21*(5), 497–523.
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32.
- Shrier, L. A., Shih, M.-C., & Beardslee, W. R. (2005). Affect and sexual behavior in adolescents: A review of the literature and comparison of momentary sampling with diary and retrospective self-report methods of measurement. *Pediatrics, 115*(5), e573–e581.
- Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavé, F. D., et al. (2018). Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences of the United States of America, 115*(1), E15–E23.
- Stone, A. A., Broderick, J. E., Schwartz, J. E., Shiffman, S., Litcher-Kelly, L., & Calvanese, P. (2003). Intensive momentary reporting of pain with an electronic diary: Reactivity, compliance, and patient satisfaction. *Pain, 104*(1–2), 343–351.
- Stone, A. A., & Schneider, S. (2016). Commuting episodes in the United States: Their correlates with experiential wellbeing from the American Time Use Survey. *Transportation Research. Part F, Traffic Psychology & Behaviour, 42*, 117–124.
- Stone, A. A., Schwartz, J. E., Broderick, J. E., & Shiffman, S. S. (2005). Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. *Personality & Social Psychology Bulletin, 31*(10), 1340–1346.
- 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.
- Stone, A. A., & Shiffman, S. (1994). Ecological Momentary Assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine, 16*(3), 199–202.
- Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2002). Patient non-compliance with paper diaries. *British Medical Journal, 324*, 1193–1194.
- Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials, 24*(2), 182–199.
- Thrash, T. M., Elliot, A. J., Maruskin, L. A., & Cassidy, S. E. (2010). Inspiration and the promotion of wellbeing: Tests of causality and mediation. *Journal of Personality & Social Psychology, 98*(3), 488–506.
- Todd, M., Armeli, S., Tennen, H., Carney, M. A., Ball, S. A., Kranzler, H. R., & Affleck, G. (2005). Drinking to cope: A comparison of questionnaire and electronic diary reports. *Journal of Studies on Alcohol, 66*(1), 121–129.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232.

- van den Brink, M., Bandell-Hoekstra, E. N., & Abu-Saad, H. H. (2001). The occurrence of recall bias in pediatric headache: a comparison of questionnaire and diary data. *Headache, 41*(1), 11–20.
- Verduyn, P., Lee, D. S., Park, J., Shablack, H., Orvell, A., Bayer, J., et al. (2015). Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology. General, 144*(2), 480–488.
- Weidman, A. C., & Dunn, E. W. (2015). The unsung benefits of material things. *Social Psychological & Personality Science, 7*(4), 390–399.
- Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality, 59*(3), 339–354.
- Willis, G. (2005). *Cognitive interviewing*. Thousand Oaks, CA: Sage.
- Wirtz, D., Kruger, J., Napa Scollon, C., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science, 14*(5), 520–524.