The Design of Ecological Momentary Assessment Technologies

Doherty, Kevin Christopher; Balaskas, Andreas; Doherty, Gavin

Published in:
Interacting with Computers

Link to article, DOI:
10.1093/iwcomp/iwaa019

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Doherty, K. C., Balaskas, A., & Doherty, G. (2020). The Design of Ecological Momentary Assessment Technologies. Interacting with Computers, 32(3), 257-278. https://doi.org/10.1093/iwcomp/iwaa019
The Design of Ecological Momentary Assessment Technologies

KEVIN DOHERTY1,2, ANDREAS BALASKAS1 AND GAVIN DOHERTY*,1

1School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland
2Copenhagen Center for Health Technology, Technical University of Denmark, Lyngby, Copenhagen, Denmark
*Corresponding author: Gavin.Doherty@tcd.ie

Ecological Momentary Assessment (EMA) methods and technologies, designed to support the self-report of experience in the moment of daily life, have long been considered poised to revolutionize human-centred research, the practice of design and mental healthcare. The history of EMA is inextricably linked to technology, and mobile devices embody many of the characteristics required to support these methods. However, significant barriers to the design and adoption of these systems remain, including challenges of user engagement, reporting burden, data validity and honest disclosure. While prior research has examined the feasibility of a variety of EMA systems, few reviews have attended to their design. Through inter-disciplinary narrative literature review (n = 342), this paper presents a characterization of the EMA technology design space, drawing upon a diverse set of literatures, contexts, applications and demographic groups. This paper describes the options and strategies available to the EMA systems designer, with an eye towards supporting the design and deployment of EMA technologies for research and clinical practice.

RESEARCH HIGHLIGHTS
• Adoption of EMA is motivated by advances in mobile technology as well as the method’s capacity to circumvent retrospective bias and support both longitudinal and ecological validity.
• The EMA design space comprises features of technologies, protocols and the interplay between them.
• Designers of EMA systems must attend to a variety of interconnected and competing values and concerns, including user engagement, reporting burden, data validity and honest disclosure.
• The lines between descriptive, prescriptive, nomothetic and idiographic uses of EMA technologies are often blurred, and designers should consider how such modes of use are inscribed through design.

Keywords: ecological momentary assessment; ecological momentary intervention; experience sampling; ambulatory assessment; mental health; EMA; engagement

Handling Editor: Russell Beale

Received 29 August 2018; Revised 23 April 2019; Accepted 23 June 2020

1. INTRODUCTION

Ecological Momentary Assessment (EMA) describes the methodology of self-report in the moment of daily life (Shiffman et al., 2008). This has been referred to as a form of systematic phenomenology (Hektner et al., 2007), capturing ‘life as it is lived, moment to moment, hour to hour, day to day’ (Shiffman et al., 2008) and permitting the ‘study of the stream of thought or behaviour’ (Hormuth, 1986). The Experience Sampling Method (ESM), Ambulatory Assessment (AA), Diary Methods and Mobile Living Labs all refer to this category of method (Broens et al., 2009, Scollon et al., 2009). EMA has antecedents and applications in the fields of psychology, ethology, sociology, anthropology, human–computer interaction (HCI), design and mental health care (Consolvo & Walker, 2003, Doherty et al., 2018, 2019, Sellen & Whittaker, 2010, Stone et al., 1999). The method’s history features many noteworthy studies and is inextricably linked to the development of technology (Wilhelm & Perrez, 2013). Timed beepers, Personal Digital Assistants (PDAs)
and smartphones each extended the reach, usability, reliability and feasibility of this approach to knowledge.

1.1. Motivations for the adoption of EMA

Mobile devices have long been thought to possess the potential to further our understanding of human experience—transforming such domains as human-centred research, design practice and mental health care (Dufau, 2011, Lazer, 2009, Shiffman et al., 2008). Much of this technological potential relates to the methodological advantages of EMA, which include the circumvention of retrospective bias as well as increased longitudinal (temporal) and ecological validity (Ebner-Priemer & Trull, 2009b, Shiffman et al., 2008).

1.1.1. Circumventing retrospective bias

The practice of self-report in the moment of daily life has often been motivated by the desire to reduce retrospective bias (Stone et al., 1999). Several decades of cognitive psychology and behavioural economics research have revealed striking differences between reports of experience made in the moment and after the fact (Ebner-Priemer & Trull, 2009b, Shiffman et al., 2008). Participants in one study of panic attacks (n = 46), for example, reported experiencing a fear of dying during 70% of panic attacks retrospectively and during 3% in the moment. Faintness was reported during 89% of panic attacks retrospectively and during 10% in the moment (Margraf et al., 1987). Among rheumatology patients (n = 83), increasing the recall period from 1 to 7 days led to heightened retrospective reports of pain despite no increase in momentary ratings over the same period (Broderick et al., 2008).

EMA can also reveal important data less often reported retrospectively. During one study (n = 85) of the side effects of antidepressant use, for example, 37 patients reported experiencing dizziness when prompted in the moment, yet only 7 reported this to their general practitioner (Barge-Schaapveld & Nicolson, 2002). These side effects were linked to a reduction in quality of life and to treatment drop-out. Other studies have found almost no relationship between retrospective and momentary reports of extreme mood changes (n = 100) (Solhan et al., 2009) and low levels of correlation between momentary and retrospective reports of coping behaviours (n = 100) (Stone et al., 1998).

By targeting near-immediate states, EMA protocols aim to avoid the effects of prior beliefs (Shiffman et al., 2008), recent and salient events (the peak-end effect) (Robbins & Kubiak, 2014), the availability heuristic, the affective valence effect, the mood congruent memory effect, duration neglect (Ebner-Priemer & Trull, 2009a), implicit theories and motivations (Barrett & Barrett, 2001, Shiffman et al., 2008), the ‘effort-after-meaning’ effect, ‘telescoping’ bias (Stone et al., 1999), the ‘focusing illusion’ (Kahneman & Krueger, 2006), social desirability effects and cultural norms (Scollon et al., 2009).

The practice of self-report has therefore been aligned with conceptions of multiple selves, which tap distinct sources of self-knowledge (Conner & Barrett, 2012, Doherty & Doherty, 2018a, Kahneman, 2011, Markus & Wurf, 1987, Santangelo et al., 2013, Zirkel et al., 2015). According to this view, momentary reports originate with an experiencing self, ‘a qualitatively different conscious self’ which ‘barely has time to exist’ (Kahneman & Riis, 2005), and which is ‘functionally and neuroanatomically different from the “remembering” and “believing” selves measured through retrospective and trait questionnaires’ (Conner & Barrett, 2012). EMA, by framing questions in the moment of daily life, can therefore produce a very different portrayal of experience—one less subject to the complex set of processes that shape retrospective reports.

1.1.2. Longitudinal validity

The adoption of momentary assessment has also often been motivated by the need to produce a more accurate characterization of experience over time (Stone et al., 1999). Given lightweight, repeated measures, EMA may be employed over significant periods of time without imposing an excessive burden on participants. One longitudinal study of emotional well-being, for example, entailed the practice of self-report five times per day for a period of one week every five years (week 1, n = 184; week 2, n = 191; week 3, n = 178) (Carstensen et al., 2011).

EMA data often facilitates analysis both within and between subjects, also referred to as idiographic or ipsative-normative analysis, where ‘“ipsative” refers to deviations around the individual mean and “normative” refers to deviations around the group mean’ (Conner et al., 2009). Such analyses can provide unique forms of insight. One EMA study (n = 27) found that job-satisfaction varied within individuals as much as average levels varied between individuals (Ilies & Judge, 2002). And another study of smoking (n = 214) found that morning craving reports better predicted relapse risk compared to daily global reports (Shiffman et al., 1997, Stone et al., 1999).

Research further suggests that temporal validity is key to understanding the severity and trajectory of illness, health and wellbeing (Conner et al., 2009, Fournier et al., 2008). The Food and Drug Administration (FDA) guidelines already recommend the use of momentary assessments for patient-reported outcomes (US Department of Health and Human Services, 2006). Diagnoses of mental illness are likewise often based upon characteristic cycles of behaviour. Individuals with bipolar disorder demonstrate significantly greater affective instability than those with depression for example, and longitudinal studies are therefore required to produce a more representative characterization of their experience (Trull et al., 2008).

EMA is therefore often uniquely suited to the study of change over time, whether striving to establish the efficacy of a particular medical intervention (Schwarz, 2012, Stone et al., 1999) or to further our understanding of the evolution of
emotional experience over the course of many years (Carstensen et al., 2011).

### 1.1.3. Ecological validity

EMA also has the potential to support increased ecological validity—with respect to ‘the occurrence and distribution of stimulus variables in the natural or customary habitat of an individual’ (Hormuth, 1986). The ‘white coat hypertension effect’ is a classical example of the importance of ecological validity. Measures of blood pressure conducted in clinical settings consistently produce readings higher than those found in daily life (Ebner-Priemer & Trull, 2009b, Robbins & Kubia, 2014, Smyth & Stone, 2003, Wilhelm & Perez, 2013). A similar effect has been found for endocrine monitoring (Smyth & Stone, 2003), implying the potential misdiagnosis and possible mistreatment of ‘hundreds of thousands of people’ (Ebner-Priemer & Trull, 2009b).

Reports of subjective experience have also often been shown to be strongly influenced by the context, and experience, of self-report (Kahneman et al., 2004). Although both hemodialysis patients and healthy individuals predict substantial differences in their emotional experience of health and illness, for example, both groups report similar levels of mood in the moment of daily life (Riis et al., 2005). In recent years, passive forms of ambulatory monitoring have been increasingly employed to support analysis of the relationship between features of daily life and the natural environment (Raugh et al., 2019).

By gathering data as people go about their daily lives, EMA can therefore support a more accurate causal understanding of events, allowing antecedent conditions, moderating variables and outcome measures to be delineated more effectively than many retrospective methodologies and other experimental designs (Smyth & Stone, 2003).

### 1.2. The potential of EMA

The methodological advantages of EMA apply to contexts as diverse as technology design, mental healthcare and the study of physical activity, addiction, personality, depression, happiness and wellbeing. These fields share challenges of reach, access and disclosure, which it is often claimed mobile technologies have the potential to overcome.

#### 1.2.1. Research

In many fields of research, the predominance of retrospective methodologies, from interviews to survey methods (Brinkman, 2009), has led to calls for a renewed focus on ecological validity—a turn towards the study of lived experience as well as the translation of laboratory findings into ‘real-world emotional, cognitive, or behavioral experiences’ (Hormuth, 1986, Trull & Ebner-Priemer, 2013). Our physical and psychological proximity to mobile devices has already enabled researchers to study actions and events that “take place ‘behind closed doors’ (literally and figuratively)’ (Zirkel et al., 2015), including how children and teenagers spend their time (Larson & Richards, 1991, Strack et al., 2004), children’s and parents’ experiences of Attention Deficit/Hyperactivity Disorder (ADHD) (Whalen et al., 2006), the social reality of young families (Aharony et al., 2011), everyday sexism (Swim et al., 2001), racism (Swim et al., 2003) and stress (Perrez & Reicherts, 1996).

A select number of studies have also demonstrated the potential of mobile applications to engage significant numbers of participants in cognitive psychology (n = 4157) (Dufau, 2011), mind-wandering (n = 5000) (Killingsworth & Gilbert, 2010) and happiness (n = 45 000) (Miller, 2012) research. And many have therefore begun to associate with EMA technologies the potential to revolutionise the practice of human-centred research. As Miller writes, it would ‘border on scientific malpractice,’ if by 2025, ‘when more than five billion people are using smartphones,’ we were ‘still giving paper-and-pencil questionnaires to a few hundred local college students, recruiting a few dozen people to participate in laboratory tasks, or running Internet studies for people just sitting at desks’ (Miller, 2012).

#### 1.2.2. HCI & design

Designers and HCI researchers have also begun to explore the potential of EMA to enhance our understanding of the use of technology and to inform the action of design. Much of our knowledge concerning the design and use of technology currently hinges, as in psychology, ‘on what people tell us’ (Barrett & Barrett, 2001). HCI research therefore also often yields a retrospective framing—rendering design research a process of ‘retrospective reversal’ (Doherty & Doherty, 2018a). Previous reviews of the HCI literature have described a ‘clear bias towards building systems and evaluating them only in laboratory settings, if at all’ (Kjeldskov & Graham, 2003, Oulasvirta, 2009), leading researchers to call for a turn towards time-sensitive studies of lived experience (Broens et al., 2009, Huang & Stotlerman, 2012, Karapanos et al., 2010, Kujala et al., 2013, Pohlmeier et al., 2009, Sun & May, 2013) and the evaluation of systems ‘in the wild’ (Consolvo, 2008, Kjeldskov & Graham, 2003, Rennick-Egglestone et al., 2016, Tamminen et al., 2003).

To date, HCI researchers have employed EMA methodologies to inform the design of a variety of mobile and ubiquitous computing technologies (Consolvo & Walker, 2003, Isaacs et al., 2013, Peesapati et al., 2010, Sellen & Whitaker, 2010), develop personalized mobile phone interruption models (Rosenthal et al., 2011), study player engagement (Fischer & Benford, 2009) and encourage physical activity (Consolvo, 2008). Others have commented that the adoption of EMA may enable designers to better identify unexpected behaviours and adaptations (Carter et al., 2007) in light of an improved understanding of users’ underlying motivations (Pielot et al., 2009). The more technologies come to mediate our experiences, the more difficult it may become to extract meaningful insights.
with respect to their design without attending to the momentary experience of their use (Doherty & Doherty, 2018a).

1.2.3. Mental health
The practice of health and mental healthcare professionals is no less frequently based in reported experience. Knowledge of patients’ wellbeing is often global and retrospective in nature, whether inferred by clinical interviews or validated screening questionnaires. Clinical psychology has faced criticism for neglecting the dynamics of symptoms, given that ‘retrospective self-reports of patients’ symptoms’ remain the primary source of practitioner knowledge (Ebner-Priemer & Trull, 2009a,b), rendering the feedback professionals receive from patients ‘sparse, delayed,’ or ‘too ambiguous to support learning from experience,’ and resulting in what has been described as a ‘low validity environment’ (Kahneman, 2011). Kazdin has remarked that despite ‘decades of psychotherapy research, we cannot provide an evidence-based explanation for how or why even our most well studied interventions produce change’ (Kazdin, 2007). The challenges of mental healthcare provision have been described as ‘so long standing, so vast, and so unresponsive’ that they require a new approach entailing change at multiple levels (Atkins & Frazier, 2011).

Effective mental healthcare rests upon an understanding of what does or does not work, and why, including ‘for whom’ simplified prevention strategies or more intensive forms of intervention are likely to prove most appropriate (Shoham & Insel, 2011, Sloan et al., 2011). Most modern psychotherapies promote awareness of thoughts, emotions and behaviour, and EMA has long been described as promising new insight into the lived experience of health and wellbeing (Atkins & Frazier, 2011, Calvo & Peters, 2014, Rizvi et al., 2011, Shiffman et al., 2008, Wichers et al., 2011). The real-world deployment of such technologies may facilitate the remote screening and monitoring of mental health and illness: extending care to under-served and at-risk groups, enabling timely assessment and intervention, supporting the collection of ecologically valid and longitudinal data, increasing patient participation, supporting honest disclosure and fostering trust between patients and health professionals (Baldassano, 2005, König et al., 2014, Rains & Keating, 2011, Ryan & Deci, 2008, Wichers et al., 2011).

1.3. Challenges for EMA
It has therefore long been claimed that mobile technologies are ‘poised to become the most powerful form of media to influence clinical practice’ (Rizvi et al., 2011), research and the design of technology itself. To date, however, the promise of EMA, and its technologies, stands at odds with the rate of their real-world deployment (Barrett & Barrett, 2001, Ebner-Priemer & Trull, 2009b, Stone et al., 1999, Trull & Ebner-Priemer, 2013). Significant barriers to the design and adoption of these systems remain.

Surprisingly little has been published concerning the design of EMA technologies, and many design guidelines refer solely to the use of PDA devices (Le et al., 2006). Few have refrained, however, from highlighting the many challenges associated with the development and deployment of these systems, including the need to train, monitor, motivate and support participants (Smyth & Stone, 2003), cope with technical difficulties, ethical concerns, low contextual control, substantial study preparation work and large datasets (Miller, 2012), tailor sampling protocols and questionnaires to diverse phenomena of interest, and realize an acceptable burden on users despite repeatedly interrupting their daily lives (Khan et al., 2009). In recent years, the evolving technological affordances of smartphones have introduced a new series of design challenges related to the timing and scheduling of EMA prompts, technical concerns of network connectivity and data loss (Burke, 2017), user attentiveness to mobile notifications, optimal notification delivery, interruptibility (Ghosh et al., 2017) and the acceptability of various forms of data collection (Chang et al., 2017).

In this paper we explore both the challenges facing designers and the strategies we might employ in response, with an eye towards supporting the design and deployment of EMA technologies for research and clinical practice.

2. METHOD
Both HCI and psychology researchers have previously expressed the need for, and conducted, reviews of the EMA literature (Cain et al., 2009, Cohn et al., 2011, Cutmore & James, 2007, Ebner-Priemer & Kubiak, 2007, Fischer, 2009, Kjeldskov & Graham, 2003, König et al., 2014, Trull & Ebner-Priemer, 2013, Wichers et al., 2011). Such prior reviews have, more often than not, focused on analysis of the feasibility of EMA technologies within specific application contexts, often through the lens of participants’ compliance to unique sampling protocols. This paper aims to support designers by contributing an initial and broad characterisation of the EMA technology design space—a grounding that serves to affirm the central role of design in the feasibility of these systems, supports a less reified framing of measures and outcomes and relates design strategies to values of user engagement, reporting burden, data validity and honest disclosure.

This paper employs a narrative literature review methodology, enabling a broad and interdisciplinary perspective as well as insight into the history and development of EMA across research, design and clinical practice, in contrast with the precise and typically quantitative focus of other systematic review methodologies (Green et al., 2006). The review corpus was sourced from databases comprising HCI, psychology and mental health literatures: the ACM Digital Library, PsycINFO, PsycARTICLES, JSTOR and Google Scholar. Each database was searched, by keyword and title where possible, using the following terms: Ecological Momentary Assessment, Ambula-
tory Assessment, Experience Sampling, Ecological Momentary Intervention, Diary Methods, EMA, EMI and ESM. Where dis-
ambiguating terms were required, the terms Design and Mobile Devices were appended to the search string in succession. In the case of Google Scholar, results were parsed until their relevance clearly dissipated. These terms cast a broad net, and searches completed later in the process resulted in the frequent repetition of results. Search results were screened by abstract and full paper, and publications meeting the following inclusion criteria were selected for the final analysis: empirical investigations using an EMA method, user study or implementation of Ecological Momentary Assessment or Intervention systems and review or analysis of the design and use of these systems. Only English language papers were included although no restrictions were placed on the year of publication. As appropriate in the case of a narrative literature review methodology, snowballing was employed to capture pertinent or frequently-cited literature omitted from the search results (Shachak & Reis, 2009). The initial search was conducted in April 2015, producing a corpus of 286 publications, and repeated according to the same criteria in February 2019 in order to identify articles published since January 2015, resulting in an additional 56 publications, for a total of 342 papers.

This corpus was subjected to thematic analysis with a focus on motivations for the adoption of EMA, insightful findings, implications for design and the metrics, features and values of the protocol and technology design space. Particular attention was paid to previous reviews of the EMA literature in order to support the comparison and corroboration of findings (see Table 2). We begin by examining the dials at the hands of the EMA technology designer.

3. RESULTS — DESIGNING EMA TECHNOLOGIES

The feasibility of EMA hinges in large part upon technology. This section examines the characteristics of effective EMA systems, drawing on a review of those described and deployed within the HCI and psychology literatures. Table 1 presents a summary of these results.

3.1. EMA technology features

The first momentary assessments were completed on paper. Timed bleepers, PDAs and smartphones each extended the reach, usability, validity and reliability of these methods. Until recently, however, much of the EMA literature pertained to the design of PDA software (Christensen et al., 2003, Ebner-Priemer & Kubiak, 2007, Le et al., 2006, Stone & Shiffman, 2002). Researchers encouraged the first users of these devices to refrain from concealing their use, in order to support their engagement and therefore the collection of valid data (Trull & Ebner-Priemer, 2013). Problems were rarely reported in this respect, however (Koop, 2002). Short Message Service (SMS) (Carter et al., 2007, Foreman et al., 2011, Hofmann & Patel, 2015) and email messaging (Foreman et al., 2011, Hareva, 2009) have since been used to conduct EMA studies, while WiFi and mobile networks have facilitated synchronous and asynchronous data sharing (Carter et al., 2007).

3.1.1. The shift to smartphones

In 2018, 96% of UK adults reported owning a mobile phone and 78% a smartphone (Ofcom, 2018), a shift reflected in the recent EMA literature’s turn towards smartphone devices. The flexibility, interactivity, intimacy and connectivity of these devices has increasingly facilitated the self-report of lived experience in daily life, leading to a predominant conception of EMA in terms of device-prompted questioning (Boukhechba et al., 2018) and expanding the array and scope of behaviours targeted by EMA studies (Burke, 2017). Regular advancements in technology and the widespread adoption of wearable devices has continued to create new opportunities for the delivery of EMA, as well as new design challenges associated with the relative novelty and powerful affordances of these devices (Hernandez et al., 2016).

Effective EMA technologies permit researchers and practitioners to register participants, send alerts, present questions, accept responses, provide feedback and debrief users. Additional features requested by researchers across the literature include support for varied sampling schemes, multiple alert options (sound, vibration, visual or a combination), external trigger possibilities (physiological or contextual) (Santangelo et al., 2013) and prompt message, reminder and goal setting options (Ramanathan, 2012).

3.1.2. User interaction

Mobile devices increasingly support new means of interaction, and therefore new self-report experiences. Buttons, check boxes, radio buttons, drop-down lists, Likert and visual analogue scales, number wheels and text fields have all been employed within self-report applications (Santangelo et al., 2013). Research suggests that how scales are presented to users matters. The use of frequency scales, different text box sizes, question layouts and scrolling have all been found to affect data validity as well as user engagement (Lanzola et al., 2014, Mavletova & Couper, 2014, Wells et al., 2014). Half of the users of one patient survey application encountered problems with vertical and horizontal scrolling for example (Lanzola et al., 2014). Another recent empirical evaluation of three distinct user interface designs found that users’ experiences were shaped by their perceived understandability, likability (including colour) and enjoyment of the interface (Healey et al., 2018).

Several researchers have also explored unique visual representations of scales, including the paper-based interpersonal grid (Fournier et al., 2008) and Inclusion of Other in the Self (IOS) scale (Aron et al., 1992), interactive digital presentation of the circumplex model of emotion (MoodMap) (Morris et al., 2014).
TABLE 1. Themes & sub-themes

| Motivations for EMA | Technology features | Protocol features |
|---------------------|---------------------|------------------|
| Retrospective bias  | Registering users   | Time contingent sampling |
| Longitudinal validity| Prompting           | Event contingent sampling |
| Ecological validity | Presenting questions| Reporting frequency & duration |
|                     | Accepting responses  | Liveability functions |
|                     | Providing feedback   | Question & questionnaire design |
|                     | Debriefing users     | Establishing shared understanding |
|                     | Passive sensing      | Training programs |
|                     | Machine learning     | Open & closed-ended questions |
|                     | Privacy & security   | Branching |
|                     | Data management      | Reactivity, habituation & attrition |
|                     |                     | Assessment & intervention |
|                     |                     | Validity in daily life |
|                     |                     | Recruitment strategies |

| Design challenges | Engaging users | Discussion |
|-------------------|---------------|------------|
| Reach & access    | Simplicity & efficiency | Design values |
| Honest disclosure | Extrinsic incentives  | Descriptive and prescriptive use |
| User engagement   | Intrinsic incentives   | Nomothetic and idiographic use |
| Reporting burden  | Gamification         | Sufficient & sufficiently valid data |
| Data validity     | Supplementary features| The role of the user |
|                   |                     | Ethics      |

2010) and a minimal stress scale comprising a single question (AndWellness) (Hicks et al., 2010). The enormous creative potential of mobile devices to elicit and express users’ inner experiences remains under-explored, however.

3.1.3. Ancillary data capture

The capacity of mobile devices to capture ancillary data including timestamps has proved critical to the feasibility of EMA. Using paper scales, participants were often found to engage in ‘parking lot compliance,’ completing reports only moments prior to their submission (Smyth & Stone, 2003). One study found an actual compliance rate of 11% for paper diaries compared to a reported rate of 90% (n = 80) (Stone et al., 2002).

Hormuth wrote in 1986 that questions concerning reliability and validity lacked answers: ‘Does the subject respond to the signals on time? Are the subject’s objective circumstances influenced by participation? Is the subject’s subjective perception of a situation influenced by the method? Is the subject capable of reporting and rating situations?’ (Hormuth, 1986). These are questions of compliance, timeliness, context, automaticity, salience and certainty (Bolger et al., 2003, Santangelo et al., 2013) that support a ‘clearer understanding of the ongoing quality of participation’ (Hicks et al., 2010). Mobile devices have allowed researchers to approach answers to many of these questions without increasing the burden on users, through the automated collection of interaction logs, geopositioning, accelerometer and connectivity data. As Wen et al. noted, however, the availability of ancillary data also raises new research and design challenges, such as the need to define an appropriate time-frame for compliance in the context of any one study (Wen et al., 2017).

3.1.4. Passive sensing

Several authors have also turned to the use of passive sensing and sophisticated algorithms to reduce the burden on users, and therefore enhance the feasibility of EMA. One pilot replication study explored the possibility of predicting day-to-day variations in EMA mood ratings using smartphone sensor data and app usage logs (Asselbergs et al., 2016). This application logged call events (time/date, duration and contact information for both incoming and outgoing calls), SMS events (time/date and contact information), screen on/off events (time/date), app usage (which applications were launched, when, and for how long), camera usage (time/date of image capture) and accelerometry data. The study’s findings depict smartphone-based, unobtrusive EMA as a technically feasible and potentially powerful EMA variant.

Improved Prompting The use of passive sensor data to enable ‘smarter’ forms of prompting has also been viewed as a promising means of supporting increased adherence. Boukhechba et al., for example, investigated the relationship
between context features (location, time, phone motion and phone usage), rates of compliance, response latency, and the duration of EMA reporting to support optimal prompting in the context of a time-based sampling protocol entailing the collection of GPS, call and text log data (Boukhechba et al., 2018). Another study, of meaningful smartphone usage, tailored the delivery of EMA prompts according to the date, time and duration of participants’ mobile app usage (Lukoff et al., 2018). Ohmage, a mobile platform designed to facilitate the recording, storage, analysis, and visualization of self-reported and continuous data, permits prompt personalization according to both previous responses and other passive forms of data collection (Ramanathan, 2012). Initial investigations therefore suggest the potential of passive monitoring to support the personalised tailoring of sampling protocol design (Hernandez et al., 2016).

3.1.5. Machine learning
Machine learning approaches have also seen increasingly frequent adoption among authors of EMA studies. One such study employed typing features (speed, duration, mistakes and the use of special characters) and transitions between self-reported emotional states to train a personalised model for automated emotional state classification. Reporting prompts were triggered according to participants’ typing patterns, and the study’s findings suggest a relationship between participants’ emotional state and typing speed (Ghosh et al., 2017b). Another study examined the use of voice commands and natural language processing techniques to support a less obtrusive form of food journaling (Oh et al., 2018). Participants registered food quantity, nutritional values and calorific intake according to an event-contingent sampling protocol which drew on a variety of sensor data sourced from smartphone and wearable devices, including physiological (blood pressure, glucose level, heart rate), semantic context (weather, social media, ambient sound) and daily activity data.

Other researchers have explored the use of machine learning techniques to reduce the burden of mood, exercise and flu-like symptom logging by tailoring the presentation of on-screen alert dialogues according to measures of context and device usage (Visuri et al., 2017). This study highlighted the value of personalized, rather than generalised, user modelling and typing—clustering users into groups in order to improve classification accuracy and therefore realise a more effective form of EMA. While machine learning techniques therefore display significant promise, their adoption must not divert attention from the highly personal character of subjective experience, the complex nature of reporting itself, nor the ethics of large-scale data collection.

3.1.6. Privacy and security
Many EMA systems give rise to ethical concerns. The need to protect users’ privacy and security, for example, is broadly acknowledged, and often achieved though pragmatic mechanisms such as password protection and the use of secure communications. The majority of studies contained within this corpus, however, typically refrain from discussing such issues directly, despite the fact that many EMA technology features merit more focused ethical reflection, including the use of cameras (Chang et al., 2017, Hernandez et al., 2016), collection of device logs (SMS, phone call and internet browsing information) (Chang et al., 2017), acquisition of sensitive mental health data, location tracking (Lai et al., 2017) and monitoring of personal and stigmatised behaviours (e.g., substance use and sexual behaviour) (Heron et al., 2017).

3.1.7. Data management features
Reporting guidelines for momentary assessment stress the ‘mountain of data’ that these techniques can produce (Stone & Shiffman, 2002). A further advantage of EMA technologies is therefore their capacity to facilitate practices of data management, transfer, analysis and modelling. The mobile app AndWellness, for example, provides a survey authoring tool and a visualization toolkit (Hicks et al., 2010). The core functionality of the extensible system Purple is described in terms of user management, content authorship, content delivery and data management capabilities (Schueller et al., 2014). EMA systems designers are therefore also advised to consider means of reducing the data management burden associated with EMA studies.

3.2. Technological solutions to the challenge of user engagement
The need to engage users has motivated the design of many technologies, but is of central importance to EMA in particular. Firstly, the collection of sufficient and sufficiently valid data hinges upon the engagement of users, often over significant periods of time (Dunn et al., 2011). In many cases, use tends to swiftly decline following an initial, and often novelty-driven, burst of interest (Cherubini & Oliver, 2009). Secondly, many self-report technologies are designed to inspire engagement with processes and outcomes beyond reporting itself (Trull & Ebner-Priemer, 2013). With respect to mental health, for example, EMA is often seen as a means to ‘actively engage patients in the process of recovery’ and increase users’ insight with respect to their own mental health and wellbeing (Wichers et al., 2011). Strategies for user engagement are therefore often a discerning feature of EMA technology designs.

3.2.1. User groups
The characteristics of any particular user group, for example, are likely to influence their engagement. One recent study examined the acceptability of smartphone-based EMA, adherence rates and predictors of non-adherence among older adults with cognitive and emotional difficulties (Ramsey et al., 2016). Technical (e.g., malfunction), logistical (e.g., competing demands), physiological (e.g., hearing difficulties)
3.2.2. User burden
One strategy for engagement in this context is to reduce the burden on users by designing for simplicity and efficiency (Scollon et al., 2009). The EMA literature features calls for ‘brevity and simplicity’ with respect to ‘the user interface and input’ (Koop, 2002), ‘a very simple, clean, and easy to use platform’ (Lanzola et al., 2014) where ‘the number of possible actions should be kept to a minimum’ (Palmblad & Tiplady, 2004). Users’ comments often tend to support these goals: ‘it’s simple and is actually uncomplicated. There’s not much interface necessary’ (Möller et al., 2013). Other strategies in this vein include the use of on-screen dialogues rather than notifications to reduce the number of steps required of users (Visuri et al., 2017) as well as the inclusion of features which grant users’ increased autonomy and control with respect to their engagement, such as ‘nap, suspend and delay’ prompt options (May et al., 2018). Care is required, however, to avoid realizing alternative threats to a study’s validity through such practices, such as problems of non-random missing data (Stone & Shiffman, 2002).

Reducing the burden on users is still viewed by many, however, as the primary challenge of EMA technology design, and strategies to this effect explored in recent years include limiting the number of assessments (Ghosh et al., 2017a), necessitating yes/no responses (Linas et al., 2015), implementing automated symptom-monitoring and remote-intervention systems (Heron et al., 2017), employing machine learning techniques to support optimized sampling (Visuri et al., 2017), as well as combining self-initiated and context-triggered forms of reporting (Chang et al., 2017).

3.2.3. Extrinsic incentives
A more proactive approach is to employ extrinsic incentives, such as monetary remuneration which might vary by rate or frequency and increase over time (Christensen et al., 2003). Presenting participants with a running reward total has been shown to encourage further participation (Aharony et al., 2011). However, too high a reward can also lead to poor data quality, due to selection biases for example (Scollon et al., 2009). A recent review of children’s and adolescents’ usage of EMA also identified raffles and level-up (promotional) strategies as opportunities for engagement (Wen et al., 2017). One intensive study of cigarette smoking among adolescents employed monetary incentives (Mermelstein et al., 2007). The authors write that paying participants “$40 for the baseline EMA weeklong, with escalating payments over waves” led participants to treat the task as a job, and in turn more seriously. The researchers also worked to create a collaborative research relationship, “through one-on-one in-person training, personalized follow-up phone calls, and consistency in contact people” (Mermelstein et al., 2007). Related attempts to motivate reporting by emphasizing the value of participants’ contributions to science may likewise require establishing a ‘viable research alliance’ (Christensen et al., 2003, Scollon et al., 2009).

3.2.4. Intrinsic incentives
Intrinsic incentives have also been used to motivate users’ engagement in reporting. An EMA study of mood, interruptibility and computer use found that providing participants with visualizations of their data resulted in a 23% higher compliance rate (Hsieh et al., 2008). Presenting response rates has alone been shown to boost protocol compliance (Christensen et al., 2003). UbiGreen, a mobile application for tracking transport habits, employed an evolving depiction of an arctic or arboreal scene to reflect the environmental impact of users’ transport behaviours (Froehlich et al., 2009). A similar approach has been used to encourage physical activity (Consolvo, 2008). Research suggests that users who possess learning goals are more likely to persevere in the face of challenges than those who hold performance goals (Dweck & Leggett, 1988), and positive computing researchers have in turn suggested ‘helping each user discover what unique conditions motivate him or her’ (Calvo & Peters, 2014).

3.2.5. Gamification
Gamification has been increasingly employed as an incentive for participation in EMA studies. One recent example found that a gamified version of an experience sampling application resulted in a higher response rate, and increased the quantity of data provided (van Berkel et al., 2017b). The gamified
characteristics of this application included a scoreboard, additional score-related information, animation and the imposition of a time limit on task completion.

3.2.6. Supplementary features

Other applications have also implemented supplementary features for the purposes of user engagement, reflection, analysis and data sharing, including visualizations, information-provision, goal-setting, note-taking and scheduling tools, various interventions and sharing functionalities (Aharony et al., 2011). MoodMap, an application designed to support emotional self-awareness, for example, provides cognitive reappraisal, physical relaxation and breathing visualization exercises in addition to facilitating mood tracking (Morris et al., 2010).

4. RESULTS — DESIGNING EMA PROTOCOLS

In general, optimal EMA protocols have yet to be developed (Shiffman et al., 2008). Methodological and technological considerations are closely interlinked, however, and this section provides insight into their relationship and design. Researchers have variously described interval, signal and event-contingent characteristics of this application included a scoreboard, additional score-related information, animation and the imposition of a time limit on task completion.

4.1. Time contingent sampling

Many EMA protocols are time-contingent. Users are notified to provide reports according to a semi-random or fixed schedule. A typical approach employs a coverage strategy—semi-random sampling within fixed intervals. A user might be prompted randomly within every 3 hour period, for example. This approach aims to minimize adaptation to a predictable schedule (Beal & Weiss, 2003). The choice of reporting frequency is typically context-dependent (Ebner-Priemer & Trull, 2009b). While retrospective, global or trait measures can prove ‘too cold,’ ‘too slow and sluggish to change,’ momentary measures can also be ‘too hot,’ ‘too volatile and overly sensitive to extraneous variables’ (Conner & Barrett, 2012). However, few studies have compared different time-based designs and there are ‘no general conventions,’ a fact which Santangelo et al. write is unsurprising, ‘as the temporal dynamics of emotional and cognitive processes are largely unknown’ (Santangelo et al., 2013).

Many studies have employed intensive sampling protocols. One study prompted 991 Scottish adolescents to report their current activity every 15 minutes outside of school hours for 4 days (Biddle et al., 2009). In another example, almost 200 participants were asked to complete assessments 5 times per day for 1 week every 5 years as part of a longitudinal study of emotional wellbeing (Carstensen et al., 2011). Other researchers have recommended that sampling more than 6 times per day for periods longer than 3 weeks should be avoided unless assessments are particularly short or additional incentives are provided (Delespaul, 1992 as cited in Christensen et al., 2003).

Once again, inter-individual differences appear to have an important effect on reporting protocol compliance. One review of children’s and adolescents’ use of EMA technologies revealed that both prompting frequencies and population characteristics appear to impact compliance rates (Wen et al., 2017). Studies involving clinical participant groups and employing more frequent forms of prompting were found to report significantly higher rates of compliance. Conversely, however, in the case of studies involving non-clinical participants alone, significantly higher compliance rates were encountered among those studies which prompted users least frequently.

There is a need to tailor protocol design to the phenomena of interest. Consider the case of an EMA application designed to screen for anxiety by means of the repeated assessment of worry in daily life. How often, or under what circumstances, should worry be evaluated? We are required to make an initial characterization of the phenomena of interest, possible contingent variables, and appropriate reporting mechanisms. In practice, EMA protocols are often built around context-specific design hypotheses. In this example, this might concern the rate at which worry is likely to fluctuate, shaped by attempts to reduce the bounds of error (such as more frequent sampling) and the burden on users (such as less frequent sampling).

4.2. Event contingent sampling

Event-contingent sampling is more often employed when targeting rare or highly specific events. This entails the choice of a ‘trigger’ (Lathia et al., 2013). Participants may be prompted to provide reports according to an ancillary measure of context, such as a change of location, or asked to complete assessments according to predefined subjective criteria, such as eating a meal or exercise. A total of 562 adolescents participating in one study of smoking were trained to report ‘smoke’ and ‘no smoke’ events in addition to responding to random prompts 5 or 6 times a day for 4 periods of 1 week 6 months apart, for example (Mermelstein et al., 2007). This combination of time and event-contingent sampling has been described as a ‘layering’ of sampling strategies (Lathia et al., 2013). Protocols might also adapt to changes in users’ behaviour over time (Lathia et al., 2013). One system tailored prompting to users’ sleep cycles, according to the times at which morning and evening diaries were completed (Sorbi et al., 2007). The use of objective triggers eliminates reliance on a user’s interpretation of what constitutes an event, and enables the assessment of compliance (Beal & Weiss, 2003). However, signal-contingent protocols can mean that participants are more likely to be prompted at inopportune moments. One study found that users providing reports according to subjective criteria forgot to do so, and those signalled objectively ‘deliberately did not answer
because it was too much effort’ (Möller et al., 2013). Future systems’ capacity to automatically recognise potential events is likely to enable new possibilities for event-contingent sampling (see Sections 3.1.4 & 3.1.5).

4.3. Sampling duration

The length of time for which sampling is conducted is typically dependent upon the time-frame most ‘likely to reveal dynamic processes . . . of interest’ or the duration of the period during which ‘change is likely to occur’ (Bolger et al., 2003). Studies requiring more frequent data entry are also often conducted over shorter periods of time in order to reduce the burden on users (Liu et al., 2018). Burst protocols, entailing contiguous sampling periods separated by extended periods of time, from days to years, have occasionally been employed to support longitudinal reporting (Carstensen et al., 2011, Doherty et al., 2019), and have been described as a means to reduce the burden on participants (Heron et al., 2017). In such instances, it may be necessary to account for routine and weekend effects, by commencing assessment periods on different days for example (Fahrenberg et al., 2001). Theories of behaviour change can also inform sampling protocol design (Schütz et al., 2015). These theories provide insight into the micro-processes of behaviour and can enable assessment based on ‘an understanding of how affect, cognition, and behavior interact and unfold over time’ (Shiffman et al., 2008).

The challenge of imposing an acceptable burden on users is defined by compromise—between the number of reports requested and the time required to complete assessments in particular (Scollon et al., 2009). Some researchers have advised that momentary reports should take no longer than 2 minutes to complete (Christensen et al., 2003, Consolvo & Walker, 2003, Hormuth, 1986). Although, older studies often reported completion times of 5 minutes or longer (Perrez & Reicherts, 1996). A total of 311 participants in one study of stress and coping were prompted to provide reports 6 times per day for 1 week, taking between 4.65 and 7.28 minutes to complete assessments on average. 76% of users found the reporting duration acceptable and 94% judged the experience positively (Perrez et al., 2000).

‘Liveability functions,’ including silent, snooze or do-not-disturb modes and repeated prompting strategies, have also been employed as a means to support engagement (Santangelo et al., 2013). Morris et al. allowed participants to choose prompting intervals between 30 minutes and 3 hours in duration, although not the exact time of assessment, and instructed users to ‘ignore prompts that could disrupt their work or personal communication’ (Morris et al., 2010). Hsieh et al. informed participants that ‘while they are not required to respond to all the questionnaires, it would be better for the study if they completed as many as possible’ (Hsieh et al., 2008).

4.4. Questions & questionnaires

A more efficient and engaging reporting process can reduce the burden on participants. While the literature concerning questionnaire development is significant, however, the EMA format remains understudied (Stone & Shiffman, 2002). Schwarz writes that responding to any question requires first interpreting the question, then retrieving relevant information, forming a response, mapping this response onto the options provided, and editing the response for reasons of social desirability and self-presentation (Schwarz, 2012). These are steps amenable to design.

What is a good mood, an average relationship or a bad experience? One of the challenges of question design is establishing a shared interpretation of language, intent and experience. The ‘pragmatic meaning’ of questions posed in research settings is informed by various factors, including ‘the purpose of the study,’ ‘the researcher’s affiliation to the content of adjacent questions’ and ‘the nature of the response alternatives’ (Schwarz, 2012). Adopting concise and targeted language can therefore support face validity.

4.4.1. Training

In the past, researchers often implemented training programs for EMA participants with an eye towards supporting engagement and the collection of valid data. This could include 6 days of ‘video presentations and real-life situations’ (Hormuth, 1986, Perrez et al., 2000). Perrez and Reicherts describe four steps typical of such programs: describing the meaning of underlying constructs, introducing the use of the EMA system, recording fictional episodes and performing a trial of the procedure under real-life conditions (Perrez & Reicherts, 1996). Conducting user testing and providing instructions to participants may still play a role in achieving the compliance and engagement of users (Christensen et al., 2003, Lanzola et al., 2014).

4.4.2. Scales

Presenting a scale enforces a frame of reference. This can lead to bias, imposing ‘reference periods’ or frequencies (a scale which presents low frequencies implies the reporting of major events for example), invoking our tendency to attempt to complete each category of a scale proportionally across multiple questions and to avoid repeating responses or producing anchoring and priming effects (Schwarz, 2012). In the case of

---

1 The most commonly cited behaviour change theories include Ajzen’s Theory of Planned Behaviour (Ajzen, 1991), the Transtheoretical Model of Behavioural Change (Prochaska & Velicer, 1997), the Fogg Behavior Model (FBM) (Fogg, 2009) and the Health Belief Model (HBM) (Janz & Becker, 1986). Other pertinent theories include social cognitive theories, such as Dweck & Leggett’s social cognitive theory (Dweck & Leggett, 1988), Bandura’s theory of self-efficacy (Bandura, 1977) and dual-process theories of cognition including the Heuristic-Systemic Model of Social Information Processing (HSM) (Todorov et al., 2002) and the Elaboration-Likelihood Model of Persuasion (ELM) (Petty & Cacioppo, 1986).
retrospective reporting, participants typically employ counting strategies for infrequent events and estimation strategies for frequent events, recall events as having occurred more recently than was the case (Stone et al., 1999) and provide more accurate responses when asked to make relative rather than frequency assessments (Scollon et al., 2009). The personal use of scales is a particularly difficult bias to counter. This refers to differences of experience and interpretation between respondents: ‘when Tim answers a 4 … maybe that is the equivalent of a 6 for Jim’ (Kahneman & Krueger, 2006). While statistical adjustments may also be possible, the U-index was introduced as one feasible solution to this problem of subjectivity. This momentary measure represents the percentage of time an individual spends in a given (unpleasant) state, in theory reducing the degree of personal sense-making involved in assessment (Kahneman & Krueger, 2006). Others have proposed tailoring ‘mood queries to an individual’s emotional signature, that is, the range and pattern of each person’s emotions’ (Morris et al., 2010).

The scales employed during EMA studies are often designed to capture nominal, ordinal, interval or ratio data as efficiently as possible (Brinkman, 2009). More complex conceptualization of a domain can require the use of coding schemes. Hormuth describes a scheme for a paper booklet which allows participants to describe their activity in the moment using four items (location, interactant, activity and conversation) for each of which 15 to 20 response options are provided (Hormuth, 1986). EMA protocols often employ closed-ended scales and questionnaires. This practice facilitates efficient reporting, simplifies analysis and can avoid ethical dilemmas by limiting response options. However, this requires respondents to characterise their thoughts, emotions and behaviours ‘according to meaningful categories that are often quite abstract’ and can change behaviour by repeatedly presenting lists of categories which influence a subject’s understanding of the domain or serve to remind them of their coping options (Stone et al., 1999).

Open-ended questions support greater freedom of expression and allow users to provide additional forms of information and sources of insight, including pertinent features of context (Runyan et al., 2013). Mobile devices enable not just text but multi-media input: photos, audio, video, drawings and more. Although understudied, these means of self-report have the potential to facilitate less structured, more creative and increasingly conversational modes of interaction which may in turn support insightful self-disclosure and more meaningful forms of communication (Johnston, 2004, Karapanos et al., 2010, Rains & Keating, 2011).

4.4.3. Branching

Applications running on mobile devices can also permit branching: the dynamic adaptation of scales to previous responses (Christensen et al., 2003, Santangelo et al., 2013). Branching can facilitate refined analysis of inter-social and person-environment interactions (Perrez et al., 2000), and support more engaging patterns of self-report akin to a ‘virtual conversation’ (Lanzola et al., 2014). Branching can also affect data validity. The order of questions can introduce priming effects (Christensen et al., 2003, Kahneman & Krueger, 2006, Schwarz, 2012, Schwarz & Hippler, 1995) and several studies have described attempts by users to take advantage of branching to shorten the reporting process (Freedman et al., 2006, Perrez et al., 2000).

Many choices in the design of EMA systems shape both user engagement and data validity. Allowing users to skip questions, for example, might introduce bias and produce less data, while also reducing false reporting rates, revealing the flaws in frequently-skipped questions and improving engagement by supporting users’ autonomy. During one study of the lived experience of ADHD, participants (n = 52) were supplied with a log with which to make note of any erroneous responses provided (Whalen et al., 2006). Allowing users to modify or expand upon their data at a later point in time can evoke similar design tensions of user engagement, honest disclosure and data validity.

4.5. Reactivity, habituation & attrition

Although momentary reporting in daily life is often employed to counter the weaknesses of retrospective assessment, it is not a ‘direct “pipeline” into consciousness’ (Christensen et al., 2003) but presents its own challenges and design trade-offs.

Self-report requires self-reflective awareness: the ability to access relevant information and the willingness to report it (Barrett & Barrett, 2001). Users can, and are likely to, exercise control over their disclosure. Designers can therefore support the collection of valid data by shaping the context and experience of self-report. Researchers conducting a study of smoking behaviours among adolescents (n = 562), for example, included non-smokers to avoid the implication that all participants were smoking (Mermelstein et al., 2007). A demo feature allowed participants to explain the study to family, friends and teachers, and password-protection aspired to prevent other students from entering fake data.

Engaging in reflection is not an inherently positive activity (Hormuth, 1986, Perrez & Reicherts, 1987, Siewert et al., 2011). That self-focus is a known characteristic of depression, for example, requires designers to consider the kinds of reflection users are asked to engage in (Rude et al., 2004)—whether ‘brooding’ or ‘reflective pondering’ (Moberly & Watkins, 2008), analytical rumination or experiential (mindful) self-awareness (Watts & Teasdale, 2004) and maladaptive or adaptive self-focus (Watts & Teasdale, 2004). Calvo

---

2 The HSM posits motivations for the cognition we engage in (whether heuristic or systematic). These potential sources of biased data include striving to present attitudes consistent with reality (an accuracy motive), to preserve one’s self-concept (a defence motive) and to express attitudes that are socially acceptable or conducive to one’s social goals (an impression motive) (Todorov et al., 2002).
& Peters suggest designing to support self-compassion by understanding potential pitfalls, acknowledging the limitations of technology, leaning towards reflective support and allowing for non-absolute categories (Calvo & Peters, 2014). The repeated and time-sensitive nature of EMA can also lead to reactivity: ‘the potential for behavior or experience to be affected by the act of assessing it’ (Shiftman et al., 2008). Reactivity is associated with increased concept or self awareness, enhanced ‘encoding or retrieval of domain-relevant information,’ and gradual entrainment of participants’ conceptualization of a domain to match its assessment (Bolger et al., 2003). EMA studies have revealed both significant (Robbins & Kubik, 2014, Wilhelm & Perrez, 2013) and insignificant (Beal & Weiss, 2003, Hufford et al., 2002) reactivity effects. Wilhelm and Perrez describe a series of experiments by McFall, among others, which demonstrated that cigarette consumption increased among students who counted cigarettes smoked but decreased among those who tracked occasions when they resisted smoking (Wilhelm & Perrez, 2013). Hufford et al. introduced a distinction between behavioural and motivational reactivity during a study of problem drinking (n = 33) although found no significant effects (Hufford et al., 2002). During one EMA study of cocaine addiction, 30% of participants reported that their participation ‘increased their self-awareness about thoughts, feelings, and behaviors and helped them make decisions that supported abstinence.’ Two users also commented that ‘frequent questioning about cocaine cravings and use could be a trigger to relapse’ (Freedman et al., 2006). It is essential to understand whether we are ‘tapping a phenomenon as it exists, or as it has been transformed by measurement’ (Scollon et al., 2009). Methods for the assessment of reactivity include the use of control groups, querying participants directly and analysing changes in mean response time, the meaning assigned to scale ratings and the internal consistency of scales over time (Barrett & Barrett, 2001, Beal & Weiss, 2003). It is frequently difficult to establish the causality of change without qualitative assessment.

The burden of reporting in daily life often leads to high rates of attrition and habituation (Santangelo et al., 2013, Scollon et al., 2009). Beal & Weiss write of their experience of a deterioration in the quality of momentary mood reporting following the second week of data collection (Beal & Weiss, 2003). Participants in another study of time-use (n = 81) were prompted to provide reports 5 to 7 times a day for 3 non-sequential weeks (Runyan et al., 2013). The average number of responses provided decreased from 18.84 in week 3 to 16.06 by week 8 and to 9.15 by week 14. 80.49% of users reported greater awareness and 43.9% expressed ‘changing how they spent their time.’

Reactivity, habituation and attrition pose significant challenges for designers of EMA technologies. These constraints pertain to protocol design, the burden on users, the phenomena of interest, honest disclosure, the nature of reflection, self-awareness and participants’ motivation, sensitivity and desire for change (Perrez et al., 2000, Scollon et al., 2009). Motivated users are more likely to provide data, but also to change their behaviour (Korotitsch & Nelson-Gray, 1999). The methodological concern of reactivity therefore translates into a design constraint for EMA systems.

4.6. Assessment & intervention

Mobile devices have made possible not only assessment of an ecological and momentary nature but intervention also. Ecological Momentary Intervention (EMI) has been characterised as a ‘therapist in your pocket’ approach with ‘the potential to revolutionize clinical treatment’ (Shiftman et al., 2008). Studies of telephone-administered psychotherapies, for example, have revealed low attrition rates (Kazdin & Blase, 2011) and a ‘dose-response relationship’ (Laazr et al., 2004). More calls appear to lead to better outcomes. One study of telephone counselling for smoking-cessation among HIV/AIDS patients (n = 95) found that those who received calls in addition to the usual care over a 2-month period were 3.6 times more likely to have quit smoking 3 months later (Vidrine et al., 2006). Other research suggests that these effects translate to mobile interventions. A video-based mobile intervention for smoking cessation led 9 of 15 participants to quit smoking during the program and half of the remainder to reduce their smoking behaviour (Whittaker et al., 2008). One recent review presents both EMA and EMI methodologies as acceptable and feasible approaches to the treatment of psychotic disorders (Bell et al., 2017). At least one EMA study (n = 130) has also demonstrated positive effects for mindfulness-based cognitive therapy on depression (Geschwind et al., 2011). While others have focused on the use of smartphones sensor data to facilitate ‘just-in-time interventions’ (Klasnja et al., 2015).

The combination of assessment and intervention in daily life has been described as one of the most effective means we possess to ‘reduce misery’ and advance wellbeing (Kahneman & Krueger, 2006). However, the line between assessment and intervention is often blurred. Momentary assessments can also be employed to tailor interventions (Wichers et al., 2011) and motivate or realise change, by producing ‘external memories’ which guard against stress for example (Perrez & Reicherts, 1987, 1996, Perrez et al., 2000). Not all self-report technologies are driven primarily by the need to gather valid data, but also by the desire to support self-insight, meaning, change, reflection, learning, empowerment, self-determination and even enjoyment. Our need to understand the intricate relationship between assessment and change is only heightened by the the growing ubiquity of mobile technologies in our daily lives.

4.7. Validity in daily life

Assessing subjective experience in daily life can support ecological validity but also increases the potential variety of con-
foundering variables. Design for data validity might attempt to account for these effects.

Ancillary data can enable the assessment of validity by facilitating analysis of diurnal patterns of activity such as ‘morningness–eveningness’ effects (Fahrenberg et al., 2001), the time of day or week, features of context such as the weather or location, extraneous circumstances and semantic descriptions of behaviour, location, social context or psychological antecedents (Beal & Weiss, 2003). These practices have been described as integrating ‘satisfaction surveys,’ monitoring ‘participant motivation and quality over time’ (Hicks et al., 2010) and building ‘controls and checks on the subject’ into experience sampling designs (Hormuth, 1986). Other variables might be more appropriately assessed prior to or following the reporting period. Pre-sampling questionnaires can capture personality traits, social-desirability bias, technology acceptance or experience, motivations and implicit theories concerning relationships between reported variables or beliefs pertinent to the phenomena of interest (Beal & Weiss, 2003, McFarland et al., 1989, Perrez & Reicherts, 1996).

Mobile devices provide a variety of unique avenues for the recruitment of participants, including the potential for large scale public engagement though online app stores (Killingworth & Gilbert, 2010, Miller, 2012, Trull & Ebner–Priemer, 2013). However, these strategies can also introduce selection biases, exclude those who do not possess a particular variety of mobile device, or attract only motivated and invested participants. Post-sampling assessments can explore usability and user experience concerns, gather participants’ impressions of the validity of their data and examine other indicators of reliability, validity and reactivity (Fahrenberg et al., 2001, Perrez & Reicherts, 1996). Comparing momentary and retrospective reports might also support a more complete understanding of users’ experiences (Doherty & Doherty, 2018a).

Several researchers have attempted to combine the advantages of momentary and retrospective reporting protocols (Cherubini & Oliver, 2009, Khan et al., 2008, 2007). The Day Reconstruction Method (DRM) is the most prominent of these attempts. Completed once at the end of the day, this measure combines reconstruction of the day’s events in diary form with items related to respondents’ emotions, activities and circumstances (Kahneman et al., 2004). Kahneman et al., however, describe experience sampling as ‘the gold standard to which DRM results must be compared’ (ibid.).

5. RESULTS — CURRENT TRENDS

In recent years, increased interest in the topic of EMA has been reflected in the publication of a large number of papers on the topic across several disciplines, including a number of pertinent reviews. Many of these concern the application of EMA methods to topics related to mental health, including the assessment of anxiety and depression among older adults with cognitive impairments (Ramsey et al., 2016), post-traumatic stress (Chun, 2016), symptoms of psychotic disorders (Bell et al., 2017), ADHD symptomatology and its effects on family relationships (Miguelez-Fernandez et al., 2018), within-person analyses of affective experiences in everyday social environments (Liu et al., 2018) and the provision of information concerning patients’ responses to treatment by psychopharmacologists (Bos et al., 2015). Others concern health more generally, including the intensity of pain among chronic pain patients (May et al., 2018), and nutrition and physical activity among youth demographics (Liao et al., 2016). A final category of reviews pertains to the use of EMA among particular population groups, such as children and adolescents (Heron et al., 2017, Wen et al., 2017).

Table 2 lists recent review papers (post-2015) contained within this corpus. Those reviews focused on the clinical deployment of EMA systems tend to provide an overview of activity within specific application contexts and generally address concerns of feasibility, acceptability and adherence. The majority of reviews highlight difficulties comparing findings across studies, even within specific application areas. Many call for clearer guidelines concerning the design of EMA studies and reporting of results, while several also emphasise the need to communicate the rationale behind any individual study design.

Only van Berkel et al.’s comprehensive review, however, focuses on the current usage of EMA within Computer Science (van Berkel et al., 2017a). The authors describe the recent shift towards the use of personal devices, growing adoption of the method within the HCI literature (as represented by the ACM CHI conference) and increased use of the method overall. The varied, but generally high, response rates contained within their corpus are encouraging, although the authors also note the need to support participants’ intrinsic motivation for engagement in reporting (see Section 3.2). The authors conclude their reflections by commenting on the need for an increasingly cross-disciplinary approach to realizing mobile devices’ potential to support experience sampling research, given their current ‘underuse’ in the computer science domain.

The broader EMA literature may soon come to reflect the significant number of computer science studies that currently incorporate sensor data (one review reports that 25% of EMA studies already employ such data; Heron et al., 2017), and this represents an important topic for future work, with potentially significant implications for design.

6. DISCUSSION

This paper provides an initial characterization of the design space for EMA systems. This framing brings together the technological, methodological and human factors pertinent to design and describes their interrelated contribution to our understanding of experience in its various forms.
### TABLE 2. Recent reviews of EMA

| Reference                        | Review topic                                      | Number of papers |
|----------------------------------|---------------------------------------------------|------------------|
| Wen et al. (2017)                | Children and adolescents                          | 42               |
| Baltasar-Tello et al. (2018)     | Mood disorders in children and adolescents         | 13               |
| Heron et al. (2017)              | Youth                                             | 24               |
| Serre et al. (2015)              | Craving and substance use                          | 91               |
| Liao et al. (2016)               | Diet and physical activity in youth               | 13               |
| Engel et al. (2016)              | Eating disorder and obesity research               | Not reported     |
| Chun (2016)                     | Post-traumatic stress symptoms                     | 15               |
| Raugh et al. (2019)              | Psychiatric populations                            | 128              |
| Lui et al. (2017)                | Psychotherapy                                      | 21               |
| Bos et al. (2015)                | Psychopharmacology                                 | 18               |
| May et al. (2018)                | Chronic pain research                              | 105              |
| Miguelez-Fernandez et al. (2018)| Attention-deficit/hyperactivity disorder           | 23               |
| Liu et al. (2018)                | Social interactions                                | 23               |
| Schembre et al. (2018)           | Behavioral research                                | 20               |
| Rodriguez-Blanco et al. (2018)   | Non-suicidal self-injury                           | 23               |
| Singh & Björling (2019)          | Craving and substance use                          | 56               |
| van Berkel et al. (2017a)        | Mobile devices                                     | 110              |

6.1. Modes of use & design

Design entails choice in the service of use. In the case of self-report technologies, choice often pertains to questions of engagement, burden, validity, change and disclosure: values entangled in tensions of scale, intrinsic and extrinsic motivation, nomothetic and idiographic use, individual, normative and clinical outcomes, self and other. An application for ADHD screening among teenagers which shares users’ data with a healthcare professional is likely to result in very different reporting behaviours than a similar application designed to support self-awareness. Users may be motivated to avoid judgement or enable access to support services for example. These practices are shaped by design. Although introducing a mindfulness exercise to a self-report application may support engagement, and in turn the collection of data, doing so may prove incompatible with the data quality required for research or clinical outcomes. Neither engaged nor unengaged users necessarily provide valid reports of their experiences.

Design values are functional, social, emotional, epistemic, conditional and interpersonal (Fuchsberger et al., 2012). Those expressed by developers of self-report applications include efficiency, ease of use, minimal device performance impact, minimal disruption, data privacy, security, wellbeing, engagement, minimal burden, satisfaction and data quality (Froehlich et al., 2007; Hicks et al., 2010). Enjoyment, empowerment, learning, responsibility, transparency, transformation, autonomy, competence, sociality, belonging, meaning and self-esteem tend to receive a less explicit focus. However, many such values lead to the inscription of distinct modes of use through design.

6.1.1. Descriptive & prescriptive use

The adoption of EMA methods has often been driven by the need to maximise the ecological and temporal validity of self-reported data. In other cases, however, it is the desire to realize change that has most strongly motivated the practice of assessment, whether by supporting personal forms of self-disclosure (Rains & Keating, 2011) and self-awareness (Hormuth, 1986), or by contributing to research outcomes, policy development and the practice of mental healthcare on larger scales (Shiffman et al., 2008). Major threads of research within both HCI and psychology are now centred upon the development of technologies to support positive affect, happiness, health and wellbeing through assessment (Calvo & Peters, 2014, Donker et al., 2013, Kennedy et al., 2012, Revere & Dunbar, 2001). The design of many EMA technologies therefore requires attending not only to the collection of valid data but the concurrent use of that data to bring about change.

6.1.2. Nomothetic & idiographic use

The typically granular nature of EMA data means that it is often leveraged to support both idiographic (within-individual) and nomothetic (between-individual) analyses. Idiographic forms of analysis are increasingly employed not only by researchers but also by individual users of mobile apps motivated by a personal desire to better understand their own experiences. The near-ubiquitous presence of mobile devices in our daily lives has enabled individuals to participate in and conduct self-
motivated forms of inquiry, engaging not only in the practice of self-report but in experimental design, analysis and the dissemination of results (Saeb et al., 2015, Schueller et al., 2014, Wichers et al., 2011). This idiographic mode of use challenges the rigidity of population-level research designs and highlights a potential design tension between the engagement of individual users and the collection of valid data to support nomothetic outcomes.

6.1.3. Sufficient & sufficiently valid data
The success of many EMA sampling protocols can be defined in terms of the collection of sufficient, and sufficiently valid, data. This first criterion requires the engagement of participants over an appropriate period of time. The second hinges upon successful practices of design to mitigate and identify momentary biases, reactivity, habituation, false compliance and contingent variables, support honest disclosure and mutual understanding, impose an appropriate burden on users and tailor the characteristics of protocols, scales and individual questions to diverse and often unique phenomena of interest.

6.2. The role of the user
The design of EMA systems and the validity of the data they capture therefore rests upon an understanding of users’ evolving roles and experiences. This implies, as Miller writes, the need to “take seriously the idea that the people we study are not just passive “subjects” but active “participants,”’ highlighting in turn, the increasingly complex nature of the relationship between participant and collaborator, methodology and technology, use and design (Miller, 2012)

6.2.1. Ethics
While privacy and security concerns have been discussed in Section 3.1.6, more subtle and complex values are also often at play in the design of EMA technologies, particularly in the context of mental healthcare, where perceptions and implications of monitoring can produce adverse consequences and a responsibility for care often exists. Such concerns frequently revolve around values of user autonomy and control and do not always have straightforward design solutions. Responsible monitoring refers not only to the capacity to manage high-risk scenarios in which a risk to a user’s health is uncovered. Self-tracking also has the potential to reinforce obsessive or negative forms of reflection, for example, and the broader implications of engaging in such practices merits increased ethical and critical reflection. As Calvo & Peters write ‘would we encourage children to track how many hugs they get from their parents?’ Logging such experiences is likely to shape how we frame, reflect upon and engage in our daily lives (Calvo & Peters, 2014).

6.3. Future work
While the EMA literature presents enormous scope for future work, particularly in light of recent technological developments, the need to recognize the interlinked and complex nature of the challenges facing designers is of primary importance to supporting both valid interpretation of the data EMA technologies produce and their ethical adoption.

6.3.1. Understanding the data and design of EMA
The development of a coherent understanding of the choices involved in EMA technology design, and their implications for the data these systems generate, is a fundamental future research challenge. This requires grasping the extent to which EMA findings can, and cannot, be generalised across systems, contexts, individuals and time. This in turn implies the need to develop appropriate distinctions between various EMA technology and protocol designs.

Understanding Users Diverse modes of use, design strategies and motivations for the adoption of self-report technologies further implies caution in our interpretation of a variety of EMA metrics. A system designed to prioritize the collection of valid data, by means of a minimal feature set, concise forms of questioning and a rigid sampling protocol for example, can also precipitate the collection of invalid data, and perhaps worse the illusion of validity, by failing to support and sustain autonomy, trust, interest or motivation among participants. Recognizing and mitigating such adverse outcomes demands close collaboration with stakeholders of all kinds including users, not only during the process of technology and protocol design, but also before, during and after use.

Understanding Reflection The highly personal nature of self-report, motivation and change merits more prominent recognition within the literature as well as further research. The vast majority of EMA studies assume participants to possess an innate and equal capacity to reflect upon, organize and articulate their subjective experiences, emotions, thoughts and behaviours. Reflection is not simply a behaviour, however, but a skill which requires ‘the accurate appraisal and expression of emotion in oneself and in others’ (Salovey & Mayer, 1990). Initial attempts by HCI researchers to support users’ acquisition of such skills are highly pertinent to the design of EMA technologies and merit wider consideration (Mamykina et al., 2008, van Gennip et al., 2015).

Understanding Engagement Similarly, engagement is considered a key metric for many EMA technologies and yet there have been few attempts to approach a systematic understanding of differences between users in this respect. Participants’ motivations for engaging in the practice of self-report, for example, have rarely been assessed in advance of system deployment. Narratives of use defined in terms of compliance alone may also neglect users’ autonomy and therefore a large part of the role of design (Doherty & Doherty, 2018b). Given
the complex relationship between the burden on users, their intrinsic motivation and the need to collect valid data, both adaptive protocols and the longitudinal combination of assessment and intervention represent important avenues for future work.

6.3.2. Advancing the ethical adoption of EMA

If what we express of our subjective experience is coloured by the experience of self-report itself, then the now near-ubiquitous presence of mobile devices in our daily lives may facilitate a fresh perspective. Mobile devices have long been described as possessing the potential to revolutionize cognitive and behavioral psychology, computational social science, psychotherapy and many other fields (Dufau, 2011, Lazer, 2009, Shiffman et al., 2008). Despite this technological promise, however, the adoption of EMA by designers, HCI and social science researchers, as well as mental health services has proved slow.

Research EMA methods have been employed to support social and psychological research efforts for decades. However, the shift to smartphones has yet, it would appear, to produce many of the advantages increased technological ubiquity was expected to entail. Continuous technological progress, it may even be argued, has rendered the adoption of EMA a more challenging endeavour, given the need to consider an ever-accelerating series of possibilities, concerns and expectations, whose negotiation increasingly requires an approach grounded in the methodology of design research.

While the widespread use of mobile devices has enabled innovative and promising forms of research, their diversity and flexibility has also significantly increased the complexity of the EMA technology design space. Many researchers may feel that they no longer possess the time, technical knowledge or design skills to keep pace with such developments (Ferreira et al., 2015, Schueller et al., 2014). Such perceptions are unlikely to be tempered by the growing adoption of adaptive, machine learning and sensor-driven systems which invite questions of trust, acceptability and the ethical deployment of technology (Thieme et al., 2020).

A more complex eco-system of use and design can also give rise to doubts concerning the relevance and accurate evaluation of the ‘traditional metrics’ of research. The continued adoption and appropriate employment of EMA methods for the purposes of research therefore requires acknowledging the importance of design, developing adaptable and accessible EMA technologies and making such systems available to those without the time or expertise to invest in systems development (Chorpita et al., 2011, Schueller et al., 2014).

HCI & Design As a field, HCI is increasingly focused on the design of self-report technologies not only for the purposes of research but to support self-knowledge, behaviour change and wellbeing, through both the personal use and clinical deployment of mobile devices (Doherty et al., 2019, Saeb et al., 2015, Schueller et al., 2014, Wichers et al., 2011). However, the value of EMA to design research is typically most clear in the context of the longitudinal and real-world deployment of technology, currently a relatively small percentage of the HCI literature. The additional burden on users and implementation costs associated with momentary assessment is likely to be viewed as prohibitive without clear expression of, and argument for, the method’s advantage. Further exploration of the ways in which EMA can contribute to design research would be valuable.

Mental Health When it comes to the use of EMA technologies to support practices of health and mental healthcare, barriers to the adoption of these systems include the lack of standardization and validation of EMA scales, ethical considerations, issues of workload and responsibility, and concerns regarding their potential impact on perceptions of human connectedness (Doherty et al., 2018, Ebner-Priemer & Trull, 2009b, Fahrenberg et al., 2001). The deployment of EMA systems at organizational or national scales and in high-risk scenarios necessitates understanding, more than ever, how design choices shape users’ experiences of self-report, the collection of valid data, and subsequent implications for action within complex eco-systems of care pathways and outcomes. The integration of EMA technologies within healthcare systems therefore presents many unique challenges and opportunities for future research.

7. CONCLUDING REMARKS

The self-report of experience in daily life has long been proclaimed to possess the potential to transform public health, clinical psychology, human-centred research and design practice. The design of technologies to support these practices is uniquely challenging, however, and their adoption has been limited to date. Designers of EMA technologies are tasked with implementing context-appropriate features, navigating divergent modes of use, motivations and outcomes, choosing appropriate questions and sampling protocols, as well as measuring and mitigating bias, reactivity, habituation, attrition and confounding variables, often in pursuit of both the engagement of users and the collection of valid data. While the widespread availability of smartphones with powerful sensing and network capabilities has created many new opportunities for EMA, this shift to mobile devices has also given rise to numerous ethical and design challenges. Realizing the significant potential of EMA therefore requires designers to attend to a complex and interconnected set of methodological, technological and human factors. This review, in support of these aims, has highlighted the features of this design space, the dials at the hands of the EMA technology designer that shape the design and real-world deployment of these systems.
Funding
Science Foundation Ireland (grants 13/RC/2106 and 17/RC-PHD/3488 to the Adapt Centre).

REFERENCES
Aharony, N., Pan, W., Ip, C., Khayal, I. and Pentland, A. (2011) The social fMRI: Measuring, understanding, and designing social mechanisms in the real world. In Proc. 13th Int. Conf. on Ubiquitous Computing. Ubicomp ’11, pp. 445–454. ACM.

Ajzen, I. (1991) The theory of planned behavior. Org. Behav. Hum. Deci. Process., 50, 179–211.

Aron, A., Aron, E. N. and Smollan, D. (1992) Inclusion of others in the self and the structure of interpersonal closeness. J. Pers. Soc. Psychol., 63, 596–612.

Asselbergs, J., Ruwaard, J., Ejdys, M., Schrader, N., Sijbrandij, M., and Riper, H. (2016) Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: An exploratory study. J. Med. Internet Res., 18, e72.

Atkins, M. S. and Frazier, S. L. (2011) Expanding the toolkit or changing the paradigm: Are we ready for a public health approach to mental health? Perspect. Psychol. Sci., 6, 483–487.

Baldassano, C. F. (2005) Assessment tools for screening and monitoring bipolar disorder. Bipolar Disord., 7, 8–15.

Baltasar-Tello, I., Miguez-Fernández, C., Penuelas-Calvo, I. and Carballo, J. J. (2018) Ecological momentary assessment and mood disorders in children and adolescents: A systematic review. Curr. Psychiatry Rep., 20, 66.

Bandura, A. (1977) Self-efficacy: Toward a unifying theory of behavioral change. Psychol. Rev., 84, 191–215.

Barge-Schaapveld, D. Q. and Nicolson, N. A. (2002) Effects of sampling study. J. Clin. Psychiatry, 63, 477–485.

Barrett, L. F. and Barrett, D. J. (2001) An introduction to computerized experience sampling in psychology. Soc. Sci. Comput. Rev., 19, 175–185.

Beal, D. J. and Weiss, H. M. (2003) Methods of ecological momentary assessment in organizational research. Org. Res. Methods, 6, 440–464.

Bell, I. H., Lim, M. H., Rossell, S. L. and Thomas, N. (2017) Ecological momentary assessment and intervention in the treatment of psychotic disorders: A systematic review. Psychiatr. Serv., 68, 1172–1181.

Biddle, S. J., Gorely, T., Marshall, S. J. and Cameron, N. (2009) The prevalence of sedentary behavior and physical activity in leisure time: A study of scottish adolescents using ecological momentary assessment. Prev. Med., 48, 151–155.

Bolger, N., Davis, A. and Rafaeli, E. (2003) Diary methods: Capturing life as it is lived. Ann. Rev. Psychol., 54, 579–616.

Bos, F. M., Schoevers, R. A. and aan het Rot, M. (2015) Experience sampling and ecological momentary assessment studies in psychopharmacology: A systematic review. Eur. Neuropsychopharmacol., 25, 1853–1864.

Boukechba, M., Caï, L., Chow, P. I., Fua, K., Gerber, M. S., Teachman, B. A. and Barnes, L. E. (2018) Contextual analysis to understand compliance with smartphone-based ecological momentary assessment. In Proc. 12th EAI Int. Conf. on Pervasive Computing Technologies for Healthcare, PervasiveHealth ’18, pp. 232–238. ACM, New York, NY, USA.

Brinkman, W. (2009) Design of a Questionnaire Instrument. In Love, S. (ed), Handbook of Mobile Technology Research Methods, pp. 31–53. Nova Science Publishers.

Broderick, J. E., Schwartz, J. E., Vikingstad, G., Pribbenow, M., Grossman, S. and Stone, A. A. (2008) The accuracy of pain and fatigue items across different reporting periods. Pain, 139, 146–157.

Broens, T., ter Hofte, H. and Ebben, P. (2009) A reference architecture for living lab measurement systems. In Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild, pp. 43–45.

Burke, L. E. et al. (2017) Ecological momentary assessment in behavioral research: Addressing technological and human participant challenges. J. Med. Internet Res., 19, 37–53.

Cain, A. E., Dep, C. A. and Jeste, D. V. (2009) Ecological momentary assessment in aging research: A critical review. J. Psychiatr. Res., 43, 987–996.

Calvo, R. A. and Peters, D. (2014) Positive Computing: Technology for Wellbeing and Human Potential. MIT Press.

Carstensen, L. L., Turan, B., Scheibe, S., Ram, N., Ersner-Hershfield, H., Samanez-Larkin, G. R., Brooks, K. P. and Nesselroade, J. R. (2011) Emotional experience improves with age: Evidence based on over 10 years of experience sampling. Psychol. Aging, 26, 21–33.

Carter, S., Mankoff, J. and Heer, J. (2007) Momento: Support for situated ubicomp experimentation. In Proc. SIGCHI Conf. on Human Factors in Computing Systems, CHI ’07, pp. 125–134. ACM.

Chang, Y.-J., Paruthi, G., Wu, H.-Y., Lin, H.-Y. and Newman, M. W. (2017) An investigation of using mobile and situated crowdsourcing to collect annotated travel activity data in real-world settings. Int. J. Hum. Comput. Studies, 102, 81–102.

Cherubini, M. and Oliver, N. (2009) A refined experience sampling method to capture mobile user experience. In Nakhmovsky, Y., Eckles, D., Riegselberger, J. (eds), Presented at the Int. Workshop of Mobile User Experience Research Part of CHI ’09, pp. 1–12.

Chorpita, B. F., Rotheram-Borus, M. J., Daleiden, E. L., Bernstein, A., Cromley, T., Swendeman, D. and Regan, J. (2011) The old solutions are the new problem: How do we better use what we already know about reducing the burden of mental illness? Perspect. Psychol. Sci., 6, 493–497.

Christensen, T. C., Barrett, L. F., Bliss-Moreau, E., Lebo, K. and Kaschub, C. (2003) A practical guide to experience-sampling procedures. J. Happiness Stud., 4, 53–78.

Chun, C. A. (2016) The expression of posttraumatic stress symptoms in daily life: A review of experience sampling methodology and daily diary studies. J. Psychopathol. Behav. Assessment, 38, 406–420.

Cohn, A. M., Hunter-Reel, D., Hagman, B. T. and Mitchell, J. (2011) Promoting behavior change from alcohol use through mobile technology: The future of ecological momentary assessment. Alcohol. Clin. Exp. Res., 35, 2209–2215.

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

Conner, T. S., Tenen, H., Fleson, W. and Barrett, L. F. (2009) Experience sampling methods: A modern idiographic approach
to personality research. Soc. Personal. Psychol. Compass, 3, 292–313.

Consolvo, S. et al. (2008) Activity sensing in the wild: A field trial of ubicomp garden. In Proc. SIGCHI conf. on human factors in computing systems, CHI ’08, pp. 1797–1806. ACM.

Consolvo, S. and Walker, M. (2003) Using the experience sampling method to evaluate ubicomp applications. IEEE Pervas. Comput., 2, 24–31.

Cutmore, T. R. and James, D. A. (2007) Sensors and sensor systems for psychophysiological monitoring: A review of current trends. J. Psychophysiol., 21, 51–71.

Doherty, K., Barry, M., Marcano-Belisario, J., Arnaud, B., Morrison, C., Car, J. and Doherty, G. (2018) A mobile app for the self-report of psychological well-being during pregnancy (BrightSelf): Qualitative design study. JMIR Ment. Health, 5, e10007.

Doherty, K. and Doherty, G. (2018a) The construal of experience in HCI: Understanding self-reports. Int. J. Hum. Comput. Studies, 110, 63–74.

Doherty, K. and Doherty, G. (2018b) Engagement in HCI: Conception, theory and measurement. ACM Comput. Surv., 51, 99.

Doherty, K., Marcano-Belisario, J., Cohn, M., Mastellos, N., Morrison, C., Car, J. and Doherty, G. (2019) Engagement with mental health screening on mobile devices: Results from an antenatal feasibility study. In Proc. 2019 CHI conference on human factors in computing systems, p. 186. ACM.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.-R. and Christensen, H. (2013) Smartphones for smarter delivery of mental health programs: A systematic review. J. Med. Internet Res., 15, 239–251.

Dufau, S. et al. (2011) Smart phone, smart science: How the use of smartphones can revolutionize research in cognitive science. PLoS One, 6, e24974.

Dunn, T., Casey, L., Sheffield, J., Newcombe, P. and Chang, A. (2011) Dropout from computer-based interventions for children and adolescents with chronic health conditions. J. Health Psychol., 17, 429–442.

Dweck, C. S. and Leggett, E. L. (1988) A social-cognitive approach to motivation and personality. Psychol. Rev., 95, 256–273.

Ebner-Priemer, U. W. and Kubiak, T. (2007) Psychological and psychophysiological ambulatory monitoring: A review of hardware and software solutions. Eur. J. Psychol. Assessment, 23, 214–226.

Ebner-Priemer, U. W. and Trull, T. J. (2009a) Ambulatory assessment: An innovative and promising approach for clinical psychology. Eur. Psychol., 14, 109–119.

Ebner-Priemer, U. W. and Trull, T. J. (2009b) Ecological momentary assessment of mood disorders and mood dysregulation. Psychol. Assessment, 21, 463–475.

Engel, S. G., Crosby, R. D., Thomas, G., Bond, D., Lavender, J. M., Mason, T., Steffen, K. J., Green, D. D. and Wonderlich, S. A. (2016) Ecological momentary assessment in eating disorder and obesity research: A review of the recent literature. Curr. Psychiatry Rep., 18, 37.

Fahrenberg, J., Hüttner, P. and Leonhart, R. (2001) Monitor: Acquisition of Psychological Data by a Hand-Held PC. In Fahrenberg, J., Myrtek, M. (eds), Progress in Ambulatory Assessment: Computer Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies, pp. 93–112. Hogrefe & Huber Publishers.

Ferreira, D., Kostakos, V. and Dey, A. K. (2015) AWARE: Mobile context instrumentation framework. Front ICT, 2, 1–9.

Fischer, J. E. (2009) Experience-Sampling Tools: A Critical Review. In Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild, pages 35–38.

Fischer, J. E. and Benford, S. (2009) Inferring Player Engagement in a Pervasive Experience. In Proc. SIGCHI conf. on human factors in computing systems, CHI ’09, pp. 1903–1906. ACM.

Fogg, B. J. (2009) A behavior model for persuasive design. In Proc. 4th int. conf. on persuasive technology, Persuasive ’09, pp. 40:1–40:7. ACM.

Foreman, A., Hall, C., Bone, K., Cheng, J. and Kaplin, A. (2011) Just text me: Using SMS technology for collaborative patient mood charting. J. Participat. Med., 3, e45.

Fournier, M. A., Moskowitz, D. and Zuroff, D. C. (2008) Integrating dispositions, signatures, and the interpersonal domain. J. Pers. Soc. Psychol., 94, 531–545.

Freedman, M. J., Lester, K. M., McNamara, C., Milby, J. B. and Schumacher, J. E. (2006) Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. J. Subst. Abuse Treat., 30, 105–111.

Froelich, J., Chen, M. Y., Consolvo, S., Harrison, B. and Landay, J. A. (2007) MyExperience: A system for situ tracing and capturing of user feedback on mobile phones. In Proc. 5th int. conf. on mobile systems, applications and services, MobiSys ’07, pp. 57–70. ACM.

Froelich, J., Dillahunt, T., Klasnja, P., Mankoff, J., Consolvo, S., Harrison, B. and Landay, J. A. (2009) Ubigreen: Investigating a mobile tool for tracking and supporting green transportation habits. In Proc. SIGCHI conf. on human factors in computing systems, CHI ’09, pp. 1043–1052. ACM.

Fuchsberger, V., Moser, C. and Tscheligi, M. (2012) Values in Action (VIA): Combining Usability, User Experience and User Acceptance. In Extended Abstracts on Human Factors in Computing Systems – CHI EA ’12, pp. 1793–1798. ACM.

Geschwind, N., Peeters, F., Drukker, M., van Os, J. and Wichers, M. (2011) Mindfulness training increases momentary positive emotions and reward experience in adults vulnerable to depression: A randomized controlled trial. J. Consult. Clin. Psychol., 79, 618–628.

Ghosh, S., Ganguly, N., Mitra, B. and De, P. (2017a) TapSense: Combining Self-report Patterns and Typing Characteristics for Smartphone Based Emotion Detection. In Proc. 19th int. conf. human-computer interaction with mobile devices and services, MobileHCI ’17, pp. 2:1–2:12. ACM, New York, NY, USA.

Ghosh, S., Ganguly, N., Mitra, B. and De, P. (2017b) TapSense: Combining Self-report Patterns and Typing Characteristics for Smartphone Based Emotion Detection. In Proc. 19th int. conf. on human-computer interaction with mobile devices and services, MobileHCI ’17, pp. 2:1–2:12. ACM, New York, NY, USA.

Green, B. N., Johnson, C. D. and Adams, A. (2006) Writing narrative method to evaluate ubicomp applications. In Proc. SIGCHI conf. on human factors in computing systems, CHI ’09, pp. 1903–1906. ACM.

Green, B. N., Johnson, C. and Doherty, G. (2018) A mobile app for the self-report of psychological well-being during pregnancy (BrightSelf): Qualitative design study. JMIR Ment. Health, 5, e10007.

Hareva, D. H. (2009) Supportive intervention using a mobile phone in behavior modification. Acta Medica Okayama, 63, 113–120.

Healey, J., Denman, P., Syed, H., Nachman, L. and Raj, S. (2018) Circles vs. Scales: An Empirical Evaluation of Emotional Assessment GUIs for Mobile Phones. In Proc. 20th int. conf. on human-
Lathia, N., Rachuri, K. K., Mascolo, C. and Rentfrow, P. J. (2013) Contextual dissonance: Design bias in sensor-based experience sampling methods. In Proc. 2013 ACM int. joint conf. on pervasive and ubiquitous computing, Ubicomp '13, pp. 183–192. ACM.

Lazer, D. et al. (2009) Life in the network: The coming age of computational social science. Science, 323, 721–723.

Lazev, A. B., Vidrine, D. J., Arduino, R. C. and Gritz, E. R. (2004) Increasing access to smoking cessation treatment in a low-income, HIV-positive population: The feasibility of using cellular telephones. Nicotine Tob. Res., 6, 281–286.

Le, B., Choi, H. N. and DeaL, J. (2006) Pocket-sized psychology studies: Exploring daily diary software for palm pilots. Behav. Res. Methods, 38, 325–332.

Liao, Y., Skelton, K., Dunton, G. and Bruening, M. (2016) A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: An adopted STROBE checklist for reporting EMA studies (CREMAS). J. Med. Internet Res., 18.

Linas, B. S., Latkin, C., Westergaard, R. P., Chang, L. W., Bollinger, R. C., Genz, A. and Kirk, G. D. (2015) Capturing illicit drug use where and when it happens: An ecological momentary assessment of the social, physical and activity environment of using versus craving illicit drugs. Addiction, 110, 315–325.

Liu, H., Xie, Q. W. and Lou, V. W. (2018) Everyday social interactions and intra-individual variability in affect: A systematic review and meta-analysis of ecological momentary assessment studies. Motivation and Emotion, pp. 1–15.

Lui, J. H. L., Marcus, D. K. and Barry, C. T. (2017) Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Prof. Psychol. Res. Prac., 48, 199–210.

Lukoff, K., Yu, C., Kientz, J. and Hiniker, A. (2018) What makes smartphone use meaningful or meaningless? Proc. ACM Interact. mob. wearable ubiquitous technol., 2, 22:1–22:26.

Manykina, L., Mynatt, E., Davidson, P. and Greenblatt, D. (2008) MAHI: Investigation of social scaffolding for reflective thinking in diabetes management. In Proc. SIGCHI conf. on human factors in computing systems, CHI ’08, pages 477–486. ACM.

Margraf, J., Taylor, C. B., Ehlers, A., Roth, W. T. and Agras, W. S. (1987) Panic attacks in the natural environment. J. Nerv. Ment. Dis., 175, 558–565.

Markus, H. and Wurf, E. (1987) The dynamic self-concept: A social psychological perspective. Ann. Rev. Psychol., 38, 299–337.

Mavletova, A. and Cooper, M. P. (2014) Mobile web survey design: Scrolling versus paging, SMS versus e-mail invitations. J. Surv. Stat. Methodol., 2, 498–518.

May, M., Junghaenel, D. U., Ono, M., Stone, A. A. and Schneider, S. (2018) Ecological momentary assessment methodology in chronic pain research: A systematic review. J. Pain, 19, 699–716.

McFarland, C., Ross, M. and DeCourville, N. (1989) Women’s theories of menstruation and biases in recall of menstrual symptoms. J. Personal. Soc. Psychol., 57, 522–531.

Mermerstein, R., Hedecker, D., Flay, B. and Shiffman, S. (2007) Real-Time Data Capture and Adolescent Cigarette Smoking: Moods and Smoking. In Stone, A., Shiffman, S., Atienza, A., Nebeling, L. (eds), The Science of Real-Time Data Capture. Self Reports in Health Research, pp. 117–135. Oxford University Press.

Migueléz-Fernandez, C., de Leon, S. J., Baltasar-Tello, I., Peñuelas-Calvo, I., Barrigon, M. L., Capdevila, A. S., Delgado-Gómez, D., Baca-García, E. and Carballo, J. J. (2018) Evaluating attention-deficit/hyperactivity disorder using ecological momentary assessment: A systematic review. ADHD Attention Deficit and Hyperactivity Disorders, pp. 1–19.

Miller, G. (2012) The smartphone psychology manifesto. Perspect. Psychol. Sci., 7, 221–237.

Moberly, N. J. and Watkins, E. R. (2008) Ruminative self-focus and negative affect: An experience sampling study. J. Abnorm. Psychol., 117, 314–323.

Möller, A., Kranz, M., Schmid, B., Roalter, L. and Diewald, S. (2013) Investigating self-reporting behavior in long-term studies. In Proc. SIGCHI conf. on human factors in computing systems, CHI ’13, pp. 2931–2940. ACM.

Morris, M. E., Kathawala, Q., Leen, T. K., Gorenstein, E. E., Guilak, F., Labhard, M. and Deleeuw, W. (2010) Mobile therapy: Case study evaluations of a cell phone application for emotional self-awareness. J. Med. Internet Res., 12, e10.

Ofcom (2018). The Communications Market Report.

Oh, H., Nguyen, J., Soundararajan, S. and Jain, R. (2018) Multimodal food journaling. In Proc. 3rd int. workshop on multimedia for personal health and health care, HealthMedia’18, pp. 39–47. ACM, New York, NY, USA.

Oulasvirta, A. (2009) Field Experiments in HCI: Promises and Challenges. In Saariluoma, P., Isomäki, H. (eds), Future Interaction Design II, pp. 87–116. Springer.

Palmblad, M. and Tiplady, B. (2004) Electronic diaries and questionnaires: Designing user interfaces that are easy for all patients to use. Qual. Life Res., 13, 1199–1207.

Peesapati, S. T., Schwanda, V., Schultz, J., Lepage, M. and Jeong, S.-y., and Cosley, D. (2010) Penseive: Supporting everyday reminiscence. In Proc. SIGCHI conf. on human factors in computing systems, CHI ’10, pp. 2027–2036. ACM.

Perrez, M. and Reicherts, M. (1987) Coping behavior in the natural setting: A method of computer-aided self-observation. In Controversial Issues in Behavior Modification. Annual Series of European Research in Behavior Therapy, vol. 2, pp. 127–137. Swets & Zeitlinger.

Perrez, M. and Reicherts, M. (1996) A computer-assisted self-monitoring procedure for assessing stress-related behavior under real life conditions. In Fahrenberg, J., Myrtek, M. (eds), Progress in Ambulatory Assessment: Computer Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies, pp. 51–67. Hogrefe & Huber Publishers.

Perrez, M., Schoebi, D. and Wilhelm, P. (2000) How to assess social regulation of stress and emotions in daily family life? A computer-assisted family self-monitoring-system (FASEM-C). Clin. Psychol. Psychother., 7, 326–339.

Petty, R. E. and Cacioppo, J. T. (1986) The elaboration likelihood model of persuasion. Adv. Exp. Soc. Psychol., 19, 123–205.

Pielot, M., Poppinga, B. and Boll, S. (2009) Understanding Tourists on a Bicycle Trip in the Wild. In Mobile Living Labs 09: Methods and Tools for Evaluation in the Wild, pp. 19–22.

Pohlmeier, A. E., Hecht, M. and Blessing, L. (2009) User experience lifecycle model ContinUE[continuous user experience]. Der Mensch Im Mittepunkt Technischer Systeme. Fortschr-Berichte VDI Reihe, 22, 314–317.

Prochaska, J. O. and Velicer, W. F. (1997) The transtheoretical model of health behavior change. Amer. J. Health Promotion, 12, 38–48.
Rains, S. A. and Keating, D. M. (2011) The social dimension of blogging about health: Health blogging, social support, and well-being. *Commun. Monographs*, 78, 511–534.

Ramanathan, N. et al. (2012) Ohmage: An open mobile system for activity and experience sampling. In *Proc. 6th int. conf. on pervasive computing technologies for healthcare, PervasiveHealth ’12*, pp. 203–204. IEEE.

Ramsey, A. T., Wetherell, J. L., Depp, C., Dixon, D. and Lenze, E. (2016) Feasibility and acceptability of smartphone assessment in older adults with cognitive and emotional difficulties. *J. Technol. Hum. Serv.*, 34, 209–223.

Raugh, I. M., Chapman, H. C., Bartolomeo, L. A., Gonzalez, C. and Strauss, G. P. (2019) A comprehensive review of psychophysiological applications for ecological momentary assessment in psychiatric populations. *Psychol. Assessment*, 31, 304.

Rennick-Egglestone, S., Knowles, S., Toms, G., Bee, P., Lovell, K. and Bower, P. (2016) Health technologies ‘in the wild’: Experiences of engagement with computerised CBT. In *Proc. 2016 CHI conference on human factors in computing systems, CHI ’16*, pp. 2124–2135. ACM.

Revere, D. and Dunbar, P. J. (2001) Review of computer-generated outpatient health behavior interventions clinical encounters “in absentia”. *J. Amer. Med. Inform. Assoc.*, 8, 62–79.

Riis, J., Loewenstein, G., Baron, J., Jepson, C., Fagerlin, A. and Ubel, P. A. (2005) Ignorance of hedonic adaptation to hemodialysis: A study using ecological momentary assessment. *J. Exp. Psychol. Gen.*, 134, 3–9.

Rizvi, S. L., Dimeff, L. A., Skutch, J., Carroll, D. and Linehan, M. M. (2011) A pilot study of the DBT coach: An interactive mobile phone application for individuals with borderline personality disorder and substance use disorder. *Behav. Therapy*, 42, 589–600.

Robbins, M. L. and Kubiak, T. (2014) Ecological Momentary Assessment in Behavioral Medicine: Research and Practice. In Mostofsky, D. I. (ed), *The Handbook of Behavioral Medicine*, pp. 429–446. Wiley Blackwell.

Rodriguez-Blanco, L., Carballe, J. J. and Baca-Garcia, E. (2018) Use of ecological momentary assessment (EMA) in non-suicidal self-injury (NSSI): A systematic review. *Psychiatry Res.*, 263, 212–219.

Rosenthal, S., Dey, A. K. and Veloso, M. (2011) Using decision-theoretic experience sampling to build personalized mobile phone interruption models. In *Proc. 9th int. conf. on pervasive computing, Pervasive ’11*, pp. 170–187. Springer.

Rude, S., Gottner, E.-M. and Pennebaker, J. (2004) Language use of depressed and depression-vulnerable college students. *Cogn. Emot.*, 18, 1121–1133.

Runyan, J. D., Steenbergh, T. A., Bainbridge, C., Daugherty, D. A., Öke, L. and Fry, B. N. (2013) A smartphone ecological momentary assessment/intervention “app” for collecting real-time data and promoting self-awareness. *PLoS One*, 8, e71325.

Ryan, R. M. and Deci, E. L. (2008) A self-determination theory approach to psychotherapy: The motivational basis for effective change. *Can. Psychol.*, 49, 186–193.

Saeb, S., Zhang, M., Kwansy, M., Karr, C. J., Kording, K. and Mohr, D. C. (2015) The relationship between clinical, momentary, and sensor-based assessment of depression. In *Proc. 9th int. conf. on pervasive computing technologies for healthcare (Pervasive-Health)*, pp. 229–232.

Salovey, P. and Mayer, J. D. (1990) Emotional intelligence. *Imagin. Cogn. Personal.*, 9, 185–211.

Santangelo, P. S., Ebner-Priemer, U. W. and Trull, T. J. (2013) Experience Sampling Methods in Clinical Psychology. In Comer, J., Kendall, P. (eds), *The Oxford Handbook of Research Strategies for Clinical Psychology*, pp. 188–210. Oxford University Press.

Schembre, S. M., Liao, Y., O’Connor, S. G., Hingle, M. D., Shen, S.-E., Hamoy, K. G., Huh, J., Dunton, G. F., Weiss, R. and Thomson, C. A. (2018) Mobile ecological momentary diet assessment methods for behavioral research: Systematic review. *JMIR Mhealth Uhealth*, 6, e11170.

Schueller, S. M., Begale, M., Penedo, F. J. and Mohr, D. C. (2014) Purple: A modular system for developing and deploying behavioral intervention technologies. *Journal of Medical Internet Research*, 16.

Schüz, N., Cianchi, J., Shiffman, S. and Ferguson, S. G. (2015) Novel technologies to study smoking behavior: Current developments in ecological momentary assessment. *Curr. Addict. Rep.*, 2, 8–14.

Schwarz, N. (2012) Why Researchers Should Think “Real-Time”: A Cognitive Rationale. In Mehl, M., Conner, T. (eds), *Handbook of Research Methods for Studying Daily Life*, pp. 22–42. Guilford Press.

Schwarz, N. and Hippler, H.-J. (1995) Subsequent questions may influence answers to preceding questions in mail surveys. *Public Opin. Quarterly*, 59, 93–97.

Scollon, C. N., Prieto, C.-K. and Diener, E. (2009) Experience Sampling: Promises and Pitfalls, Strength and Weaknesses. In Diener, E. (ed), *Assessing Well-Being: The Collected Works of Ed Diener*, vol. 39, pp. 157–180. Springer.

Sellen, A. J. and Whitaker, S. (2010) Beyond total capture: A constructive critique of lifelogging. *Commun. ACM*, 53, 70–77.

Serge, F., Fatseas, M., Swendsen, J. and Auricacme, M. (2015) Ecological momentary assessment in the investigation of craving and substance use in daily life: A systematic review. *Drug Alcohol Depend.*, 148, 1–20.

Shachak, A. and Reis, S. (2009) The impact of electronic medical records on patient–doctor communication during consultation: A narrative literature review. *J. Eval. Clin. Pract.*, 15, 641–649.

Shiffman, S., Engberg, J. B., Faty, J. A., Perz, W. G., Gny, M., Kassel, J. D. and Hickcox, M. (1997) A day at a time: Predicting smoking lapse from daily urge. *J. Abnorm. Psychol.*, 106, 104–116.

Shiffman, S., Stone, A. A. and Hufford, M. R. (2008) Ecological momentary assessment. *Amn. Rev. Clin. Psychol.*, 4, 1–32.

Shoahm, V. and Insel, T. R. (2011) Rebooting for whom? Portfolios, technology, and personalized intervention. *Perspect. Psychol. Sci.*, 6, 478–482.

Siewert, K., Kubiak, T., Jonas, C. and Weber, H. (2011) Trait anger moderates the impact of anger-associated rumination on social well-being. *Personal. Individ. Differ.*, 51, 769–774.

Singh, N. B. and Björling, E. A. (2019) A review of EMA assessment period reporting for mood variables in substance use research: Expanding existing EMA guidelines. *Addict. Behav.*, 94, 133–146.

Sloan, D. M., Marx, B. P. and Keane, T. M. (2011) Reducing the burden of mental illness in military veterans: Commentary on Kazdin and Blase (2011). *Perspect. Psychol. Sci.*, 6, 503–506.

Smyth, J. M. and Stone, A. A. (2003) Ecological momentary assessment research in behavioral medicine. *J. Happiness Stud.*, 4, 35–52.
Solhan, M. B., Trull, T. J., Jahng, S. and Wood, P. K. (2009) Clinical assessment of affective instability: Comparing EMA indices, questionnaire reports, and retrospective recall. Psychol. Assessment, 21, 425–436.

Sorbi, M., Mak, S., Houtveen, J., Kleiboer, A. and van Doornen, L. (2007) Mobile web-based monitoring and coaching: Feasibility in chronic migraine. J. Med. Internet Res., 9.

Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., Paty, J., Porter, L. S. and Cruise, L. J. (1998) A comparison of coping assessed by ecological momentary assessment and retrospective recall. J. Personal. Soc. Psychol., 74, 1670–1680.

Stone, A. A. and Shiffman, S. (2002) Capturing momentary, self-report data: A proposal for reporting guidelines. Ann. Behav. Med., 24, 236–243.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E. and Hufford, M. R. (2002) Patient non-compliance with paper diaries. Br. Med. J., 324, 1193–1194.

Stone, A. A., Shiffman, S. S. and DeVries, M. W. (1999) Ecological Momentary Assessment. In Kahneman, D., Diener, E., Schwarz, N. (eds), Well-Being: The Foundations of Hedonic Psychology, pp. 26–39. Russell Sage Foundation.

Strack, R. W., Magill, C. and McDonagh, K. (2004) Engaging youth through photovoice. Health Promot. Pract., 5, 49–58.

Sun, X. and May, A. (2013) A comparison of field-based and lab-based experiments to evaluate user experience of personalised mobile devices. Adv. Hum. Comput. Interact., 2013, 14–23.

Swim, J. K., Hyers, L. L., Cohen, L. L. and Ferguson, M. J. (2001) Everyday sexism: Evidence for its incidence, nature, and psychological impact from three daily diary studies. J. Soc. Issues, 57, 31–53.

Swim, J. K., Hyers, L. L., Cohen, L. L., Fitzgerald, D. C. and Bylsma, W. H. (2003) African american college students’ experiences with everyday racism: Characteristics of and responses to these incidents. J. Black Psychol., 29, 38–67.

Tamminen, S., Oulasvirta, A., Toiskallio, K. and Kankainen, A. (2003) Understanding Mobile Contexts. In Chittaro, L. (ed), Human-Computer Interaction with Mobile Devices and Services, vol. 2795, pp. 17–31. Springer.

Thieme, A., Belgrave, D. and Doherty, G. (2020) Machine learning in mental health: A systematic review of the HCI literature to support the development of effective and implementable ML systems. ACM Trans. Comput. Hum. Interact., 27, Article 34.

Todorov, A., Chaiken, S. and Henderson, M. D. (2002) The Heuristic-Systematic Model of Social Information Processing. In Dillard, J., Pfau, M. (eds), The Persuasion Handbook: Developments in Theory and Practice, pp. 195–211. Sage.

Trull, T. J. and Ebner-Priemer, U. (2013) Ambulatory assessment. Ann Rev. Clin. Psychol., 9, 151–176.

Trull, T. J., Solhan, M. B., Tragesser, S. L., Jahng, S., Wood, P. K., Piatecki, T. M. and Watson, D. (2008) Affective instability: Measuring a core feature of borderline personality disorder with ecological momentary assessment. J. Abnorm. Psychol., 117, 647–661.

US Department of Health and Human Services (2006) Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. Health Qual. Life Outcomes, 4, 1–79.

van Berkel, N., Ferreira, D. and Kostakos, V. (2017a) The experience sampling method on mobile devices. ACM Comput. Surv., 50, 93.

van Berkel, N., Goncalves, J., Hosio, S. and Kostakos, V. (2017b) Gamification of mobile experience sampling improves data quality and quantity. Proc. ACM on interactive, mobile, wearable and ubiquitous technologies, 1, 1–21.

van Gennip, D., van den Hoven, E. and Markopoulos, P. (2015) Things that make us reminisce: Everyday memory cues as opportunities for interaction design. In Proc. 33rd annual ACM conf. on human factors in computing systems, CHI ’15, pp. 3443–3452. ACM.

Vidrine, D. J., Ardovino, R. C., Lave, A. B. and Gritz, E. R. (2006) A randomized trial of a proactive cellular telephone intervention for smokers living with HIV/AIDS. Aids, 20, 253–260.

Visuri, A., van Berkel, N., Luo, C., Goncalves, J., Ferreira, D. and Kostakos, V. (2017) Predicting interruptibility for manual data collection: A cluster-based user model. In Proc. 19th int. conf. on human-computer interaction with mobile devices and services, MobileHCI ‘17, pp. 1–14. ACM Press, Vienna, Austria.

Watkins, E. and Teasdale, J. D. (2004) Adaptive and maladaptive self-focus in depression. J. Affect. Disord., 82, 1–8.

Wells, T., Bailey, J. T. and Link, M. W. (2014) Comparison of smartphone and online computer survey administration. Soc. Sci. Comput. Rev., 32, 238–255.

Wen, C. K. F., Schneider, S., Stone, A. A. and Spruijt-Metz, D. (2017) Compliance with mobile ecological momentary assessment protocols in children and adolescents: A systematic review and meta-analysis. J. Med. Internet Res., 19.

Whalen, C. K., Henker, B., Jamner, L. D., Ishikawa, S. S., Floro, J. N., Swindle, R., Perwien, A. R. and Johnston, J. A. (2006) Toward mapping daily challenges of living with ADHD: Maternal and child perspectives using electronic diaries. J. Abnorm. Child Psychol., 34, 111–126.

Whittaker, R., Maddison, R., McRobbie, H., Bullen, C., Dunn, S., Dorey, E., Ellis-Pegler, M., van Rooyen, J. and Rodgers, A. (2008) A multimedia mobile phone–based youth smoking cessation intervention: Findings from content development and piloting studies. J. Med. Internet Res., 10.

Wichers, M., Simons, C., Kramer, I., Hartmann, J., Myin-Germeys, I., van Emmel, A., Peeters, P., de Leeuwaan, P. and van Os, J. (2011) Momentary assessment technology as a tool to help patients with depression help themselves. Acta Psychiatr. Scand., 124, 262–272.

Wilhelm, P. and Perez, M. (2013) A History of Research Conducted in Daily Life. Scientific Report No. 170. Department of Psychology, University of Fribourg, Switzerland.

Zirkel, S., Garcia, J. A. and Murphy, M. C. (2015) Experience-sampling research methods and their potential for education research. Educ. Res., 44, 7–16.