Taylor & Francis
Taylor & Francis Group

HUMAN–COMPUTER INTERACTION

Human–Computer Interaction

ISSN: 0737-0024 (Print) 1532-7051 (Online) Journal homepage: http://www.tandfonline.com/loi/hhci20

What Does All This Data Mean for My Future Mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being

Victoria Hollis, Artie Konrad, Aaron Springer, Matthew Antoun, Christopher Antoun, Rob Martin & Steve Whittaker

To cite this article: Victoria Hollis, Artie Konrad, Aaron Springer, Matthew Antoun, Christopher Antoun, Rob Martin & Steve Whittaker (2017) What Does All This Data Mean for My Future Mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being, Human–Computer Interaction, 32:5-6, 208-267, DOI: 10.1080/07370024.2016.1277724

To link to this article: https://doi.org/10.1080/07370024.2016.1277724

Accepted author version posted online: 04 Jan 2017.
Published online: 04 Jan 2017.

HUMAN–COMPUTER INTERACTION, 2017, Volume 32, pp. 208–267

ISSN: 0737-0024 print / 1532-7051 online
DOI: 10.1080/07370024.2016.1277724

What Does All This Data Mean for My Future Mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being

Victoria Hollis, Artie Konrad, Aaron Springer, Matthew Antoun, Christopher Antoun, Rob Martin, and Steve Whittaker
University of California at Santa Cruz

We explore the Examined Life, informing the design of reflective systems to promote emotional well-being, a critical health issue. People now have increasingly rich, digital records of highly personal data about what they said, did, and felt in the past. But social science research shows that people have difficulty in tracking and regulating their emotions. New reflective technologies that promote constructive analysis of rich personal data potentially offer transformative ways that individuals might better understand themselves and improve well-being. However, there are important system design challenges in supporting effective reflection about personal data. We explore *fidelity* in recording and representing past personal mood data, and *forecasting* future actions, feelings, and

Victoria Hollis (hollis@ucsc.edu, www.VictoriaHollis.com) is a human–computer interaction (HCI) researcher with an interest in technologies for quantified self, emotion monitoring, and self-awareness. She is a PhD student of cognitive psychology at the University of California at Santa Cruz. **Artie Konrad** (www.akonrad@ucsc.edu) is an HCI researcher who focuses on the intersection of memory, emotion, and technology. His background is in technology and reminiscence, with other interests in behavior change and computational health. **Aaron Springer** (www.alspring@ucsc.edu) is a computer science PhD student at University of California at Santa Cruz. His research interests include computational systems supporting affect and computer science education. **Matthew Antoun** (www.mantoun@ucsc.edu) is interested in the expressive potential of AI systems. He has a master's degree in computer science from University of California at Santa Cruz. **Christopher Antoun** (www.cantoun@ucsc.edu) is interested in the application of AI to narrative. He has a master's degree in computer science from University of California at Santa Cruz. **Rob Martin** (www.therobmartin@ucsc.edu) is a user experience designer and researcher, currently studying cognitive science and HCI at University of California at Santa Cruz. His interests include human factors, human-centered design, and emerging technologies. **Steve Whittaker** (www.swhittak@ucsc.edu) is a cognitive scientist with an interest in communication, well-being, and memory technologies; he is Professor in the Department of Psychology at University of California at Santa Cruz.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/HHCI.

thoughts. Much prior personal informatics work has been dedicated to past-centric tools for recording and capture. In contrast, forecasting examines how we might use such past data to inform and motivate *our future selves*, providing recommendations about remedial actions to improve future well-being. Fidelity addresses both *how* and *what* reflective systems should show people about their pasts, in particular whether we should filter negative past experiences. To inform reflective system design, we examine forecasting and fidelity in controlled field trial interventions that explore two novel system designs for presenting and reflecting on mood data. We detail findings from 165 participants, 4,693 participant logfiles, 65 surveys, and 15 user interviews. Our novel forecasting system, EmotiCal, uses past mood data to model and visualize future user moods with the goal of encouraging participants to adopt remedial new behaviors to regulate negative moods before they occur. Such forecasting both improved mood and subsequent emotional self-awareness compared with controls who simply monitored their past. Consistent with system goals, interview responses also indicated that participants generated important insights into behaviors that affect their moods. Our second intervention examined filtering; it assessed the impact on well-being of recording and revisiting past experiences containing negative emotions. We compared participants who were encouraged to record and reflect on positive versus negative experiences. Long-term measures of happiness and ruminative behaviors improved by recording and reflecting on positive but not negative experiences, although this depended on the intensity of the negative experience. We discuss general design and theory implications for future systems that support monitoring, reflection, and forecasting to facilitate productive examination of our emotional lives.

CONTENTS

1. INTRODUCTION

People have access to increasingly rich, detailed records of personal data regarding their past emotions and behaviors. This is partly due to the greater use of communications technologies and the records these tools generate about previous activities, conversations, and feelings. We have also seen the recent emergence of dedicated self-tracking tools, such as the Fitbit or Empatica Embrace, and mobile applications to monitor multiple, quantitative aspects of our behavior (e.g., exercise, diet, sleep, mood, locations, and habits). This explosion of available personal data has given rise to a host of *personal informatics* (PI) technologies. PI tools aim to support data reflection to generate insights for self-improvement for health, productivity, or well-being (Choe, Lee, Lee, Pratt, & Kientz, 2014; Li, Dey, & Forlizzi, 2011; Rooksby, Rost, Morrison, & Chalmers, 2014).

Our focus here is on PI tools for emotional well-being. Mental health is a critical societal problem, with 30% of men and 40% of women experiencing a major depressive episode at least once in their life and minorities being even more vulnerable (Kruijshaar et al., 2005; World Health Organization, 2012). In addition, though a considerable portion of the population experiences a mood disorder in their lifetime, reportedly only 36.8% or these individuals seek professional health care (Alonso & Lepine, 2007), further arguing for the importance of low-cost access to interventions. PI technologies allow users to capture and analyze data about personal behaviors that affect mood such as sleep, diet, and exercise. Such data potentially allow users to analyze and modify these behaviors improving mood and promoting well-being (Choe et al., 2014). There is strong public interest in such PI technologies with thousands of emotional well-being apps available in GooglePlay and iTunes. However, there are considerable challenges in designing effective systems for this domain. One key issue is how to support insightful reflection about emotions. How do our past actions and experiences affect the way that we currently feel? How might such experiences affect how we will feel in the future? And more important, how might we engage in remedial behaviors that improve our future moods and well-being?

Findings within social science underscore the need for better systems for emotional reflection. People find it difficult to understand and predict their future emotional state (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 2002; Wilson & Gilbert, 2005). They find it hard to choose activities that will improve long-term mood (Tice, Bratlavsky, & Baumeister, 2001), and in a distressed state they tend to recall more negative information (mood-congruent memory; Watkins, Vache, Verney, & Mathews, 1992). These cognitive biases contribute to the difficulty that many experience in regulating negative emotions, with significant consequences for mental health and well-being (Lyubomirsky, Kasri, Chang, & Chung, 2006; Pennebaker & Chung, 2011; Seligman, Steen, Park, & Peterson, 2005). Despite the promise of PI systems to alleviate these issues, prior research suggests that simple self-monitoring of emotions is not sufficient for improving emotional health outcomes (Depp et al., 2015; Durkin, 2006; Faurholt-Jepsen et al., 2015).

There are various potential explanations for the low success of self-monitoring tools. First, current PI tools place severe cognitive demands on their users, assuming high levels of *data-analytic competence*. One obvious challenge here is the sheer complexity of the data gathered, making it difficult for users to draw clear inferences about which of their habits are affecting target outcomes. For example, a current mood can result from multiple interacting factors including exercise, diet, sleep, social interaction, and so forth. Many PI technologies implicitly assume quite advanced diagnostic abilities on the part of their users in interpreting multivariate streams of time-varying data, despite prior work highlighting low analytic competence (Peters, Hibbard, Slovic, & Dieckmann, 2007). Although recent health systems have begun to provide better support for end-user analytics (Bentley et al., 2013; Epstein, Cordeiro, Bales, Fogarty, & Munson, 2014; McDuff, Karlson, Kapoor, Roseway, & Czerwinski, 2012; Ståhl, Höök, Svensson, Taylor, & Combetti, 2009), many still assume sophisticated data analytic skills. In addition, while these new systems offer exciting potential solutions to promote end-user interpretations of data, most analytic support tools do not address emotional well-being, and most have not been evaluated.

A second, critical problem is converting analytic insights into *actionable behaviors*. Even if users succeed in overcoming the challenges of successfully interpreting their past behaviors, they must also decide what can be done to change these. It is not enough for users to passively understand relations between their behaviors and target well-being goals. Users must also plan and enact practical actions to improve their emotional well-being. There may be a significant gap between a user insight that increased exercise improves mood and developing executable plans that will actually be completed. In support of such planning, PI systems need to incorporate important research findings suggesting that remedial actions need to be simple, achievable, and concrete (Gollwitzer, 1999; Konrad et al., 2015; Locke & Latham, 2002). Proposing to run a half marathon each day could in principle improve mood but is unlikely to be enacted in practice.

A third challenge for emotional well-being systems concerns *mood valence* effects. The mere effect of tracking different types of emotional records could affect subjective well-being. There are well-attested benefits for both reflecting on both prior positive events (Bryant, Smart, & King, 2005; Isaacs et al., 2013; Konrad, Isaacs, & Whittaker, 2016; Konrad, Tucker, Crane, & Whittaker, 2016; Parks, Della Porta, Pierce, Zilca, & Lyubomirsky, 2012) and negative events (Pennebaker & Chung, 2011; Sloan & Marx, 2004). Furthermore, the reflective process itself may be affected by the valence of the events being analyzed. Cognitive processing is affected by mood, with people being more analytic when in a negative mood and more creative when feeling positive (Isen, 2004; Schwarz, 2000). Reflecting on negative events is known to temporarily depress current mood (Konrad, Isaacs, et al., 2016; Sloan & Marx, 2004, 2016), which in turn may alter a participants' ability to analyze information and create remedial plans. This article, therefore, reevaluates a major assumption in PI tools for emotional well-being: that all emotional events should be recorded with equal importance or, at minimum, without any guided suggestions on which *types* of emotional events to prioritize for reflection.

We present two controlled field trial studies that address challenges with current PI systems for emotional well-being. Our first study addresses the dual design challenges of (a) providing improved analytic support for deriving insights about one's emotional patterns and (b) developing actionable recommendations for future remedial actions to boost future mood. We have already noted that much prior work on PI assumes sophisticated data analytic abilities on the part of their users. We present a new technique, *emotional forecasting*, that finesses some of the problems with such analytics and evaluate a novel PI system that supports mood regulation. It tackles the cognitive challenge of analyzing complex personal data by modeling and visualizing the relations between users' past activities and subsequent mood. That visualization also assists users in forecasting the anticipated consequences of not taking action to improve future mood, as well as suggesting remedial actions that improve mood. Our approach also addresses affective components to behavior change motivation (Baumeister, Vohs, Nathan Dewall, & Zhang, 2007) by visualizing the mood-boosting effects of adopting new activities. We show that offering these actionable recommendations and visualizations of future mood increases daily ratings of positive affect, promoting insight as well as increasing users' reported awareness of their emotions.

Our second deployment addresses *emotional filtering*, specifically which events we should record and reflect upon. Most PI systems implicitly assume that users should capture a complete record of their emotional past, including both positive and negative experiences. Although some work shows significant benefits for critically reflecting on past negative events (Pennebaker & Chung, 2007), there is also evidence that reminiscing on positive experiences is critical for well-being (Lyubomirsky & Layous, 2013). Our second study contrasts these two approaches to emotional reflection by testing the effects of tracking strictly positive versus negative events. Our results indicate that recording extremely negative events detracts from well-being, suggesting that designs might encourage users to strategically emphasize positive past experiences to improve well-being.

2. RELATED WORK

2.1. Lifelogging and Personal Informatics

There is a tradition within human–computer interaction of designing systems that potentially support reflection about our everyday lives. One significant early initiative involved lifelogging, an approach that aims to collect a complete record of everything that users say, do, and feel (Bell & Gemmell, 2009). There have been numerous critiques of lifelogging. Theoretical criticisms have challenged the need for exhaustive records of our pasts, instead highlighting the importance of adaptive *forgetting*, identifying situations where a complete record is counterproductive (Bannon, 2006; Mayer-Schönberger, 2009; Sas & Whittaker, 2013; Van House & Churchill, 2008). A second critique is that lifelogging overemphasizes *capture* while failing to identify what benefits might accrue from exhaustive records (Sellen & Whittaker, 2010). Furthermore, the intuition that lifelogging will help us "remember everything" seems overstated; the memory improvements resulting from having rich detailed daily activity records are relatively modest (Kalnikaite, Sellen, Whittaker, & Kirk, 2010; Sellen et al., 2007).

One reaction to critiques of lifelogging has been the emergence of PI (Choe et al., 2014; Li et al., 2011; Rooksby et al., 2014). PI is distinct from lifelogging. Rather than focusing on exhaustive capture, PI seeks to identify precisely how detailed records of our pasts might be exploited to improve important aspects of our lives. In a study of 15 self-trackers, Li et al. (2011) identified key uses of detailed past personal data, showing how such data can help evaluate behavior change goals (weight, exercise, productivity, etc.). PI tools potentially provide detailed personal data to help users analyze causal relations between trigger activities and goals. For example, careful reflective analytics might suggest to a user that he or she should monitor and optimize exercise, as this affects work productivity.

Although a tremendous number of PI products capture and track rich data, relatively few offer end-user analytics or recommend remedial actions to promote behavior change. Instead many systems leave users to conduct complex data analysis to extract insights and determine solutions (see Jawbone UP3, Moves, Tactio Health). Nevertheless, some recent research systems begin to tackle these significant challenges by providing support for end-user analytics. Such support includes interpretable summaries (Bentley et al., 2013; Epstein et al., 2014; Khovanskaya, Baumer, Cosley, Voida, & Gay, 2013) or visualizations that simplify complex information (Bentley et al., 2013; Epstein et al., 2014; McDuff et al., 2012; Ståhl et al., 2009). For example, Health Mashups (Bentley et al., 2013) displays correlations between different streams of data and provides text summaries to explain patterns (e.g., "You feel happier on the weekends"). Similarly, Epstein et al. (2014) designed visualization "cuts" showing trends across multiple data streams to help users identify patterns linking activities and other information (e.g., physical activity and the weather). Despite the promise of these methods, however, many of these systems were not evaluated to determine whether analytic support tools do indeed improve well-being.

These research systems potentially support end user analytics by providing correlations between data streams, but a different approach has been taken in MONARCA (Bardram et al., 2013; Doryab, Frost, Faurholt-Jepsen, Kessing, & Bardram, 2015; Faurholt-Jepsen et al., 2015). MONARCA allows bipolar patients to track activities and mood, to better understand how trigger activities affect manic or depressive components of bipolar disorder. For example, a patient might experience more volatile moods if he or she skips medication, fails to exercise, or sleeps poorly. Unlike many of the prior systems, MONARCA was deployed to a target, clinical population to explicitly test the effects of analytic support. However 78 participants using the monitoring-only version of MONARCA showed no significant improvements and even a tendency for more depressive symptoms compared to a control group (Faurholt-Jepsen et al., 2015). Although an ongoing trial is exploring improved analytic support, this MONARCA evaluation highlights the need for additional work on actionable analytics and a greater exploration of possible benefits for nonclinical users.

The failure of self-monitoring to improve emotional well-being suggests that simply generating insights is not enough. To successfully change their behaviors, users have to convert analytic insights into simple, concrete actionable behaviors (Gollwitzer, 1999). Another important but underresearched factor underlying PI system success is motivation, and far less attention has been paid to affective components of motivation that are critical for adopting new behaviors (Baumeister et al., 2007; Hollis, Konrad, & Whittaker, 2015). Many systems presuppose that behavior change is a purely rational process, assuming that careful analytics will inevitably promote adoption of adaptive new remedial behaviors. However, if users are unmotivated, then behavior change is unlikely (Michie et al., 2011; Prochaska, DiClemente, & Norcross, 1992). New system designs can help users better addressing these affective components to motivate behavior change. Our own work (Hollis et al., 2015) tackled this in a system that helps users change unwanted habits, such as snacking, nail biting, or procrastination. In common with many other PI approaches, our system encourages users to monitor these behaviors. More important, it encourages users to focus on the emotional consequences of indulging in those behaviors. A month-long deployment showed that users who reflected on how they *felt* after engaging in an unwanted behavior were significantly less likely to engage in those behaviors longer term (Hollis et al., 2015).

2.2. Emotions, Memory, and Well-Being

There is extensive psychological research identifying nondigital interventions that promote well-being. Although exact definitions of well-being are debated, there is scientific consensus that it involves two main components—*hedonic*, relating to real-time affect (Kahneman, 1999, 2000; Kahneman, Diener & Schwarz, 1999 Stone, Shiffman, Schwartz, Broderick, & Hufford, 2003), and *eudaimonic*, which concerns progress towards longer term life-goals and values (Deci & Ryan, 2000, 2008; Diener, 1984). Positive activity interventions show a long history of successful outcomes with increased emotional well-being and reduction in depressive symptoms (Cuijpers, Van Straten, & Warmerdam, 2007; Dobson & Joffe, 1986; Ekers, Richards, & Gilbody, 2008; Turner, Ward, & Turner, 1979). This approach first identifies positive activities that can boost mood. Multiple paper-based surveys have created reliable ratings for hundreds of daily activities, determining whether they promote or harm emotional states. For example, seeing old friends is reliably judged as a highly pleasant activity, whereas physical discomfort is judged to be unpleasant (Lewinsohn & Amenson, 1978; Lewinsohn & Libet, 1972). In subsequent interventions, positive activities are recommended as remedial actions to enhance well-being. A meta-analysis of 17 positive activity scheduling interventions for depression (N = 1,109 subjects) found that activity-scheduling interventions improved depressive symptoms relative to waitlist/placebo controls, supportive counselling, and brief psychotherapy (Ekers et al., 2008), with similar success rates to cognitive behavioral therapy.

Other well-being interventions have explored different positive psychology strategies deployed in digital contexts. Seligman et al. (2005) tested five digital interventions, three of which resulted in lasting improvements for emotional well-being: “three good things” (participants make a daily note of three good things that happened to them), exploiting signature strengths, and gratitude exercises. Similar findings were generated by a study using an online app called LiveHappy (Parks et al., 2012), which, like Three Good Things (3GT) (Munson, Lauterbach, Newman, & Resnick, 2010), encouraged participants to implement positive psychology activities (such as expressing gratitude or positive thinking exercises). A meta-analysis of 51 interventions showed that these positive thinking interventions increase well-being and reduce depressive symptoms (Sin & Lyubomirsky, 2009).

Although the preceding research points to the value of a positive outlook and positive thinking, the picture is more complex when we consider negative emotions. Intuitively, it would seem maladaptive to reflect on past negative experiences. Nevertheless, there is overwhelming evidence that people experience significant benefits from reanalyzing past negative events. Pennebaker and Beall (1986) developed the *emotional writing* (EW) paradigm, a paper-based intervention in which people are encouraged to repeatedly write about past negative events, transforming their feelings with positive health benefits. Feelings about past negative events become more positive after writing, as events are reconstrued in *redemption narratives* in which experiencers come to see themselves as more resilient as a result of overcoming adversity (Pennebaker, 2004; Wildschut, Sedikides, Arndt, & Routledge, 2006). More than 200 studies demonstrate EW’s benefits across a wide range of participants, including cancer patients, prisoners, and people suffering from job loss (Pennebaker & Chung, 2011). EW corresponds to significant changes in physical health such as immune system functioning, reduced blood pressure, and fewer doctor visits. EW also corresponds to improved goal outcomes such as higher college grades, greater success in job seeking, and improved mood. Four meta-analyses quantify its effects (Frisina, Borod, & Lepore, 2004; Harris, 2006; Meads, 2003; Smyth, 1998), reporting positive effect sizes of d = 0.15–0.47 depending on population and outcome measures. However EW has limitations. Although there are long-term benefits, participants often experience short-term negative affect (Nolen-Hoeksema et al., 1991; Sloan & Marx, 2004), and EW interventions do not work well for those with ruminative thinking habits or those suffering from PTSD (Gidron, Peri, Connolly, & Shalev, 1996).

EW interventions have largely focused on nondigital contexts. However, there are complex relations between memory, emotions, and well-being that have important possible consequences for digital settings targeted by personal informatics tools. Research in nondigital contexts shows clear adaptive biases in how people remember their pasts, including a bias to remember more positive than negative events, selective editing of past negative events, and faster forgetting of the impact of past negative experiences (Konrad, Isaacs, et al., 2016; Walker, Skowronski, & Thompson, 2003). The result of all these biases is that, for many, memories are skewed overwhelmingly toward the positive and arguments have been made that these biases are adaptive, allowing us to recall the past more positively while selectively forgetting or attenuating our recall of negative events (Walker et al., 2003).

These adaptive positivity biases are important in the digital context because of potential consequences for well-being. Organic unmediated memories are edited over time, excising the negative, but digital recordings are an unchanged rich record of exactly what the user did and felt at the time. We investigated consequences for well-being in a month long field trial of a reflective system finding that digital records showed similar positivity biases to their organic unmediated counterparts (Konrad, Isaacs, et al., 2016). Participants tended to record more positive than negative events, reactions to negative events tended to attenuate faster than to their positive counterparts, and events tended to be remembered more positively over time. All of these were accompanied by improvements in well-being.

2.3. Emotion Tracking and Reflective Systems

Recently we have seen the emergence of new applications that track emotions and memories, some with the goal of promoting sharing of these memories. Often these are very simple emotion tracking systems that allow people to log moods (e.g., MoodPanda, MoodScope). In some cases, simple correlations are generated to allow people to understand relations between event triggers and their mood (InFlow). At the time of writing, none of these systems provides extensive support for analyzing emotional patterns or recommendations about remedial actions that might improve future mood. Newer research systems take a different approach to supporting analysis of our past emotions by encouraging users to actively record and reflect on daily events. They differ from automatic passive lifelogging in two important ways: First, they deal with experiences that are deliberately captured by users themselves. Second, they support active processing of prior recordings, re-presenting intentionally captured experiences back to users who are encouraged to deliberately reprocess them, through either sharing or personal reflection.

Some of these new systems *repurpose* social media content (such as social media posts or photos) by re-presenting these to users for targeted reflection. For example, Facebook's On This Day (Facebook Inc, 2015a) takes status updates from past years and re-presents these to users, encouraging people to personally reflect on or share these posts with others. Photo-based services such as Timehop (Timehop, 2014), Google's Rediscover This Day (Google Inc, 2015), and MorningPics (Mulligan, 2014) do the same for images. Other repurposing services such as Facebook's Year in Review (Facebook Inc, 2015b), or Spotify's Year in Music (Spotify, 2015) have different goals; they recycle past social media behaviors but instead aim to *summarize* intervals from the user's past by combining popular posts or music. Such summaries might be used for personal reflection about the year gone by or be shared with others for social reminiscence. Although these applications are intended to be celebratory, they have created significant discussion over the consequences of unintentionally exposing users to highly negative events, such as the death of a daughter (Meyer, 2014).

A different approach has been taken by other, recent reflection systems. Rather than simply repurposing prior social media posts, these *prospective* reflection systems encourage users to intentionally generate memory-oriented content with the goal of improving future well-being. For example, Echo (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016) and MoodAdaptor (Konrad, Tucker, et al., 2016) encourage users to deliberately record personal experiences for future reflection. Prospective systems also provide structured prompts to reprocess memory content (sometimes repeatedly). Participants log positive experiences they want to revisit and savor, as well as negative experiences they need to reanalyze. Deployments show well-being benefits, sometimes over the long term.

One of the best studied reflection systems is Pensieve, where a systematic research program has explored various aspects of using technology for reminiscence (Cosley et al., 2009; Cosley, Schwanda Sosik, Schultz, Peesapati, & Lee, 2012; Peesapati et al., 2010). Early explorations began with very simple impersonal prompts ("Some of the nicknames that you've had"), many of which were successful in engaging users and promoting reminiscence about past events. A second iteration extended these features by linking to the user's social media, photo, and music sites. These media were then used to prompt reflection, for example, a photo or song from the user's collection might be accompanied by the prompt "Do you remember?" Although social sharing of reflections was not successful (Cosley et al., 2012), other aspects of the system were used extensively. Participants reported that they enjoyed the reflective process and that it improved their mood. Reflections were generally found to be positive, although the nature of the prompt affected this (Peesapati et al., 2010). Long-term deployments found that users valued the tool for reminiscence and that a variety of recommendation prompts for reflection increase engagement (Sosik & Cosley, 2014).

Studies of these research systems offer important lessons for design. First, the nature of reflection is highly dependent on the exact prompts used to elicit users' analysis of their past (Cosley et al., 2012; Peesapati et al., 2010). Second, in prospective reflection, users actively craft recordings of their experiences if they know they will see those recordings again, prospectively editing those experiences to make them more positive (Konrad et al., 2016). Third, the acts of active recording and reflection offer different benefits for well-being that persist for months after using the system (Konrad, Isaacs, et al., 2016). However, we still lack systematic understanding of the exact effects of reflection about negative events on well-being, which we address here.

2.4. Research Questions

In what follows, we present two controlled intervention studies involving 165 participants to explore the design of well-being systems to promote reflection and well-being. In the first study, we explore new designs for emotional forecasting that support end-user analysis of mood over time, recommending and motivating specific future actions to improve emotional well-being. In the second study, we examine emotional filtering: addressing the question of whether prospective reflection systems that promote recording and reviewing of past negative events detracts from well-being.

Forecasting Data to Support Future Actions

The first system design aims to help users analyze complex personal data to encourage adaptive future actions. There are considerable design challenges analyzing past data to motivate helpful future actions. Many PI technologies for emotional well-being offer detailed access to the minutiae of our pasts but much less guidance about how those past experiences might usefully direct our future selves and our future behaviors. We address whether end-user analytics improve emotional well-being compared with monitoring-only systems. Prior work has shown no benefits for tools that support simple emotion tracking (Faurholt-Jepsen et al., 2015). The study also explores which aspects of analytic support provide benefits, probing the effects of visualizations and comparing different types of actionable recommendations.

Reflecting on Positive versus Negative Personal Experiences

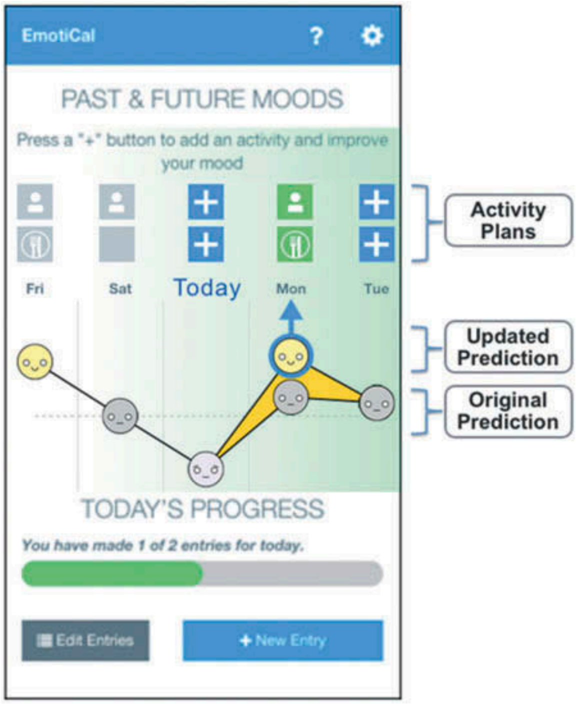
Our second study examines systems that support reflection about past prior experiences to promote well-being. As with forecasting and action recommendation, a critical question is again how we select these prior experiences for reflection. Current commercial systems use simple algorithms such as time or popularity to select prior experiences. However, another important consideration is the *emotional valence* of these reflective events. Should we recommend that people reflect only on positive experiences, or should they also confront more negative aspects of their past (Haimson, Brubaker, Dombrowski, & Hayes, 2015; Sas & Whittaker, 2013; Zhao & Lindley, 2014)? Beyond valence, are there differences in emotional intensity that affect the benefits or risks of monitoring emotions? Nondigital studies of reflection show that benefits are different when participant reflect on emotionally intense versus milder events (Pennebaker & Chung, 2007).

Examining these issues is also important for scientific understanding of online behavior. People are increasingly spending huge parts of their lives using digital technologies. Technologies such as Facebook and Fitbit now make it easy for us to revisit and review many aspects of our past behaviors. It is important that we understand how this affects our emotions and well-being (Burke & Develin, 2016; Gonzales & Hancock, 2011; Kim & Lee, 2011; Kramer, Guillory, & Hancock, 2014). Our work adds to this literature, grounding our research questions in prior social science theory but addressing critical questions in the domain of digital behavior.

3. FORECASTING: ENCOURAGING FUTURE PLANNING FOR WELL-BEING

The first study explores a novel design for well-being that generates emotion forecasts and motivated actionable recommendations for improving mood. Our system is called EmotiCal (*Emoti*onal *Cal*endar), a web and smartphone application. Like many current products, participants first log past moods and events triggering

FIGURE 1. Emotional forecasting and remedial action recommendations.

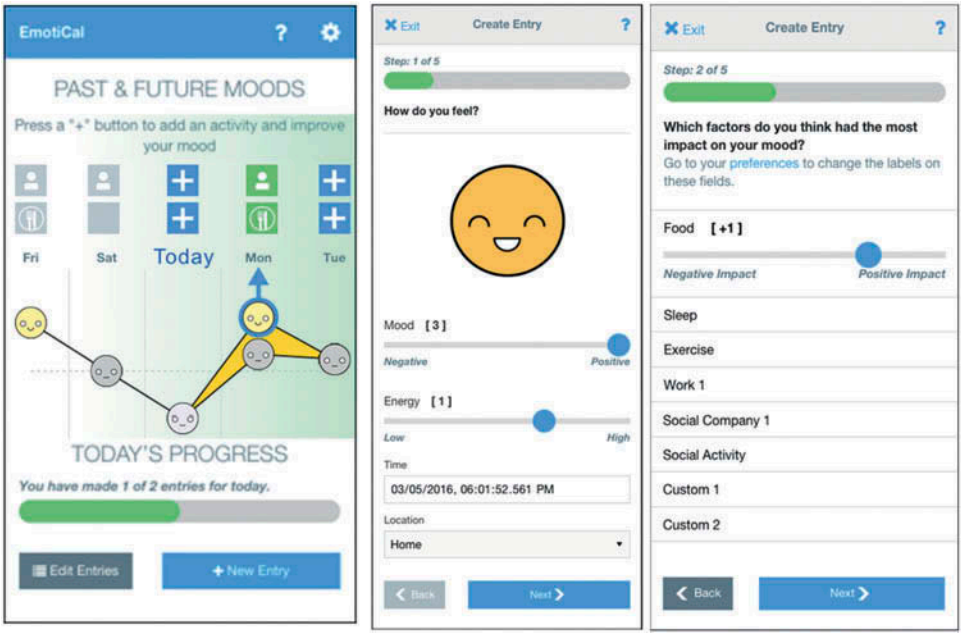


Note. **This image shows the visualization displayed to emotion-forecasting participants in Week 3 of the study. The leftmost two points in the line graph indicate average mood ratings on previous days, and the center point is the average rating for the immediate day. The two rightmost points indicate predicted mood for upcoming days. The + symbol allows participants to explore remedial actions to enhance future mood. This participant added two activity plans for Monday. The visualization displays an updated mood prediction if those activities are enacted.**

those moods. The novel emotional forecasting user interface (UI) is shown in Figure 1. The forecasting visualization highlights potentially problematic future days, encouraging participants to actively plan enjoyable activities to improve their emotion forecast. The figure shows an example of predicted affective states, along with recommended remedial actions to improve that emotional future.

EmotiCal supports mood monitoring and tracking of trigger activities (e.g., sleep, exercise, work, etc.) that affect mood (see Figure 2). EmotiCal analyzes past mood data to generate a 2-day forecast for a user's potential moods for tomorrow and the day after. Most important, the system provides actionable analytics to change these forecasts. Participants can explore the effects of adopting recommended actions to enhance mood. The visualization is updated to show expected changes if the participant enacts activity recommendations, providing motivation to adopt those actions. To improve the likelihood of users actually adopting remedial actions, the

FIGURE 2. EmotiCal system components.



Note. **The leftmost image shows the landing page for participants in the *emotion-forecasting* condition, displayed only in Week 3 of the study. *Monitoring-only* participants did not see the visualization. The center image shows the *mood-monitoring* interface with options to rate mood and energy level, as well as contextual information, for example, time and location. The rightmost image shows the user interface for *choosing trigger activities* that led to current mood (e.g., that food had a positive impact on current mood). There are 14 possible activities the user might select as affecting mood, although not all are shown in this user interface view. If custom labels were specified, these were displayed in addition to the trigger type, for example, "Custom 1 (Teaching class)" or "Social Company 2 (Partner)."**

recommendations are personally tailored, derived from analysis of prior logfile history or profiled to fit the user's Basic Psychological Needs (Deci & Ryan, 2000). For example, past user data might predict that a user would be in a neutral mood tomorrow. However, the same mood modeling would also allow us to recommend that the user meet with a friend or go for a bike ride, because an analysis of the logfiles indicate that both of these activities correspond to higher emotion ratings for that user. Our intervention evaluates this emotional forecasting approach, assessing EmotiCal's effectiveness in improving emotional well-being.

We conducted a 3-week field trial evaluation of EmotiCal with two main objectives. First, we designed and evaluated new methods for end-user analytics leading to remediation. These analytics model past logged emotions data, producing an emotional forecast to motivate actionable future plans to change mood. Second, we assessed whether EmotiCal is more effective for improving well-being than current approaches involving simple tracking of one's past data. We address the following research questions. These questions are framed in the context of emotion forecasting, though they address general personal informatics design issues for using past data to motivate future behaviors and goal achievement.

1. Does emotion forecasting improve overall daily mood ratings compared with simple emotion/action monitoring that is supported by current systems?
2. Does emotion forecasting encourage future directed actions? In particular, does it encourage participants to actively plan new recommended enjoyable activities more than simply monitoring past emotions and actions?
3. Does emotion forecasting improve participant's sense of control over their emotions and self-awareness, again when compared with simply monitoring past emotions and actions?

There are two main contributions. First, we extend personal informatics technologies by designing a novel method that analyzes user-generated data to help users forecast possible moods and to encourage actionable behaviors. Second, we provide evidence for the effectiveness of a new approach to improving well-being.

3.1. The EmotiCal System

EmotiCal has two main system goals: first, to support simple *mood tracking* to collect data about participants' moods and the factors underlying them, and second, to make *past data actionable*, the mood graph visualization motivates participants to analyze and actively plan future enjoyable activities to improve their emotion forecast. The system was developed iteratively using low-fidelity prototyping, extensive user feedback, and a small-scale trial deployment. Figures of early UIs are shown in Appendix A. The next section describes our final system design.

Mood Monitoring

Mood monitoring involved participants logging information about current mood, energy level, and trigger activities contributing to the current mood. Based on early prototype feedback and prior studies (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016), participants were prompted to create at least two mood entries per day. They were prompted via automatic text message if they did not spontaneously submit a minimum of two entries per day. Making a mood entry was lightweight (e.g, low effort) and could typically be completed in about 40 s.

The mood-monitoring component of the UI is shown in Figure 2 (center panel). To create a mood entry, participants first make a simple mood valence decision, choosing a mood ranging from –3 (*very negative*) to +3 (*very positive*) and an energy level ranging from –3 (*low energy*) to +3 (*high energy*). Structured after the circumplex model of emotion (Russell, 1980), participants also logged their energy level so as to supply a more accurate report of their emotional experiences. To provide additional context, participants could also optionally log time and date and set the location (Home, Work, Other). We are aware that there are well-supported emotion models that incorporate multiple complex dimensions (Frijda, 1988; Scherer, 2001), but more complex recording procedures could have compromised our design requirement for lightweight logging and compliance.

After selecting mood, energy level, time, and location, participants were prompted to identify possible trigger activities (Figure 2, right panel) that explained their mood and rate these activities on a scale of –2 (*negatively impacted mood*) to +2 (*positively impacted mood*). So, for example, a user might use the system to attribute their high positive mood to trigger activities such as good sleep, eating well, and so forth. These trigger activities were identified from a combination of literature review, an analysis of logfiles from a previous study including participant free-writes about mood (Konrad, Isaacs, et al., 2016), surveys ($n = 39$), and interviews ($n = 12$) discussing activities that affected mood. Data are shown in Appendix B. Together these data allowed us to identify 14 trigger activities including standard options (food, sleep, exercise, general social activity) and custom options (Work Activity 1, Work Activity 2, Leisure Activity 1, Leisure Activity 2, Leisure Activity 3, Social Company 1, Social Company 2, Social Company 3, Custom 1, Custom 2). Custom options were included because interviews and surveys showed that in addition to *general* trigger activities, there were also more esoteric personal mood triggers. For example, people wanted to record the emotional effects of highly customized trigger activities (meeting a specific friend, or engaging in a particular hobby). After identifying triggers, participants submitted a brief freewrite description about how those trigger activities impacted their mood. Again, recording triggers was lightweight and took around 40 s on average, as participants tended to select a small number of triggers.

Emotion Forecasting

The second, novel part of the EmotiCal system aimed to provide actionable recommendations and to motivate engagement in future activities that directly improve mood. Participants in the EmotiCal condition could interact with a mood graph visualization, updating this by adding or removing activity plans to explore the emotional consequences of future activities. As noted previously, this visualization and UI were the result of extensive prototyping, user feedback, and a prior deployment. Participants in monitoring-only conditions did not see the emotion forecasting UI.

After 2 weeks of data entry, during which participants simply tracked mood and trigger activities, EmotiCal displayed a visualization showing mood over 5 past, present, and future days (Figure 1). Specifically the visualization showed (a) the past 2 days' average mood entries, (b) today's projected mood entry, and (c) the next 2 days' projected mood ratings (right-hand side of Figure 1). Participants were encouraged to actively manipulate their future mood by adding recommended mood-enhancing activities to their schedule in the following way. Two slots (+s) were displayed above today, tomorrow, and the day after tomorrow. Participants could click on a slot and a rank-ordered list displayed 10 recommended activities. There were two sources of recommendations. These could be based on a user's *history* or their *psychological needs* profile. Five *history*-based recommended activities were tailored specifically from the participants' own past data, proposing actions that their own logging data showed had positive past effects on mood. The remaining *needs*-based five activities were generated as follows: Before the study each participant's psychological needs were assessed using the Basic Psychological Needs Scale (BPNS; Deci & Ryan, 2000). Participants next rated a general list of 39 possibly enjoyable activities (MacPhillamy & Lewinsohn, 1982). *Needs*-based recommendations were generated by matching activities to need as identified from the participants' BPNs profile. For example, a participant who scored low on pretest measures of *relatedness* might be recommended a greater number of activities intended to improve this dimension such as more social activities. This additional set of needs-based activities ensured that participants received a greater variety of options if their own logfiles were limited. In addition, *needs* profiling allowed us to recommend actions for participants who made no positive mood entries. This additional set of *needs*-based activities was developed in direct response to user feedback on an initial EmotiCal deployment, which exclusively used *history* for generating recommendations. This led to user complaints about the obviousness and lack of variety in recommended activities.

After selecting a recommended activity, participants were prompted to schedule that activity, as past research shows that concrete implementation intentions improve the likelihood of following through with a plan (Gollwitzer, 1999). Textual feedback then summarized this activity plan (e.g., "At 9 a.m. tomorrow, I will go for a run."). The participant then wrote a brief description of the expected benefits from engaging in that activity and any additional planning information necessary, as prior work also shows this to improve intervention effectiveness (Turner et al., 1979). After finishing activity planning, the visualization would then update to show the predicted changes in mood resulting from adding the new action. Next are example planned activity entries from emotion-forecasting participants (see Figure 3) with sources of activity recommendations coming from either pretest BPNS profiles or their logfile recordings.

Because each recommended activity is chosen to be enjoyable, adding it increases the expected mood for the planned day. For example, 71153's activity plan would increase his estimated mood for the next day from "slightly happy" (+1 on the mood scale) to "happy" (+2 on the mood scale). We now describe how we modeled the impacts of different actions and predicted overall future mood.

Forecasting Algorithms: Determining Planned Activity Impact and Predicted Emotions

From the initial 2 weeks of data entry, we developed a predictive model of each user's future moods using linear regression. This model also allowed us to identify trigger activities tailored to each participant that had an impact on their mood. In other words, we modeled the extent to which exercise, sleep, food, custom factors,

FIGURE 3. Example planned activity entries from emotion-forecasting participants showing source of activity and details.

User ID	Recommendation Source	Activity	Text Entry of Activity Plan
15213	Profile based	Bake or cook	Call my parents to ask them how to cook an African dish. I've been talking about cooking it for over a whole semester, time for action.
80126	History based	Food	It helps me gain more energy and feel happiness. I will go to my favorite restaurant around 6pm tonight.
77777	Profile based	Learn something new	I will go to ASL meetup at Starbucks to practice sign language. It will be fun meeting new people and improving my signing skills in a welcoming environment
13489	Profile based	Invite a friend to the movies	I like being around my friends and watching movies. It makes me happy and it's fun.
42968	History based	Work 2	I feel that I should do some work toward writing daily, not only does it keep up my abilities as a writer while I'm not in school it also feels like what I should be doing.
71153	Profile based	Learn something new	Learning something new is stimulating to me. I found an app to learn Spanish as well as an site to learn coding.

***Note*. Activity sources are based on either *history* (drawn from the participants' personal log of past activities) or *needs* (motivated by the pretest profile and survey of enjoyable activities).**

and so forth, influenced mood for that person. As a result, our predictive model could determine the differential impact of exercise between two participants; whereas one participant's mood may be strongly affected by exercise, exercise may have no effect on another participant, who might be more affected by work.

Determining Impact of Activities on Mood. Individual linear regression models were trained for each user to predict mood using the 14 trigger activities (e.g., sleep, exercise, social activity, etc.) that users recorded when making a mood entry. As users continued to make entries throughout the study, the models were updated on a 12-hr interval to automatically incorporate the new entries. These personalized models determined which triggers most influenced that user's mood. Recommended activities were based on those triggers activities that had a significant positive impact on mood ratings. Participants were presented with 10 possible recommendations each time they planned a new activity; as already described, half were *history* based and the other half were based on *needs* profiles.

Predicting Effect of Added Activities on Mood. To predict the effect of each user-chosen activity on mood, we used the same regression models in a different way. The regression models made two predictions. The first prediction was simply the baseline state of the user if that user engaged in no additional activities that day. The second prediction included nonzero regressors for the added activities that the user scheduled. The exact regressor value was calculated by averaging the user's previous scores for that specific activity. The mood boost, that is, the difference between the baseline and scheduled mood scores, was then displayed in the visualization to show estimated change in future mood resulting from scheduling the activity. The up-arrowed emoticons in Figure 1 depict this boost.

Mood boosts ranged widely from 0.10 to 2.7, on a 7-point scale. We were concerned that especially small boosts risked giving participants the impression that adding activities would do little to improve their mood. Therefore outlier boosts were transformed according to the mood boost distribution across all participants. The transform eliminated the possibility that participants would schedule an activity and receive an extreme prediction, diminishing confidence in the system. This transform also resulted in all mood boosts having a large-enough range that participants could both differentiate between activities when exploring their affects on the visualization and see how adding each activity affected projected future mood. However, we administered a postintervention survey, including questions to assess participants' subjective perception of the forecast accuracy with ratings of perceived accuracy given on a 7-point scale ("Rate the accuracy of the mood predictions") with responses ranging from 1 (*very inaccurate*) to 7 (*very accurate*) and an opportunity to share their thoughts in an open-ended question ("Please explain why you evaluated them as accurate or inaccurate:").

Mood prediction models were statistically highly predictive: Individual linear regression models' average R^2 was .50, with a standard deviation of .14. On average the models were statistically significant at p = .04. The mean absolute error for the individual models averaged .64, with a standard deviation of .16.

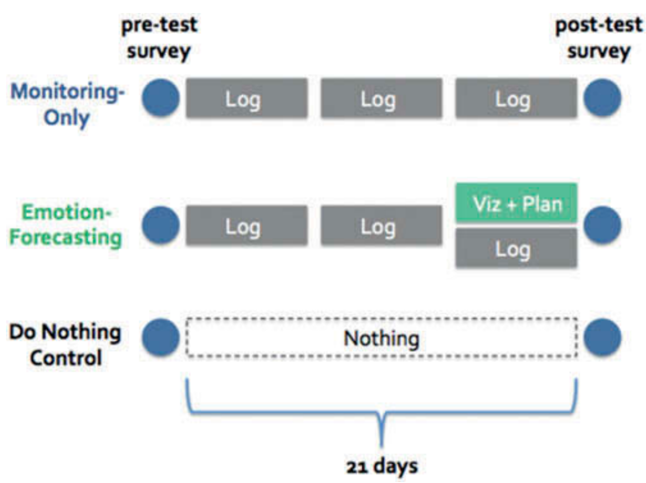
Generating Predictions for Future Moods. Baselines for future moods were predicted from a univariate time series of the previous mood scores. This provides a more dynamic experience than using the baseline linear models, which would simply predict a constant value for future moods. The predictions were made using individualized Autoregressive Integrated Moving Average forecasting models that were trained on t – 1 days to predict days t and t + 1. Autoregressive Integrated Moving Average models have three parameters (p,d,q) indicating the order of the autoregressive model, degree of differencing, and the order of the moving average model. We used a standard automatic method to tune these parameters (Hyndeman & Khandakar, 2007) involving exhaustively searching for the best fitting model according to the Akaike Information Criterion score. The search was completed for all permutations of (p,d,q) with each parameter between (0,2). The best scoring model then made baseline predictions for future days for individual participants. Accuracy for the prediction of future baseline moods varied. Mean absolute error between predicted future moods and the actual mood on the predicted day was 1.13 with a standard deviation of .82.

Hand-Coded Activity Recommendations. In addition to the five history-based activity recommendations derived from the personalized mood models, we also generated five needs-based recommendations. Pretest responses were obtained to the BPNS measuring participants' levels of autonomy, competence, and relatedness, which are a considerable determinant of life satisfaction (Ryan & Deci, 2001). In combination with the pretest ratings of activity enjoyment (from Lewinsohn's Positive Activity Schedule), the researchers hand-selected five activity recommendations tailored toward each user's BPNS ratings(see Appendix C). For example, if a user indicated in the pretest BPNS a low rating for relatedness and competence, then we would choose two activities for relatedness, two activities for competence, and one for autonomy in an effort to optimize personal satisfaction. Comparisons of the effectiveness of history versus needs based recommendations are described in the Engagement and Perceived Accuracy section.

3.2. Intervention

The field trial evaluation of EmotiCal involved three intervention conditions (see Figure 4). First, to serve as a control for state-of-the-art emotion-monitoring systems (such as InFlow or Moodscope) we included a *monitoring-only* condition in

FIGURE 4. EmotiCal study design showing three conditions: monitoring-only, emotion-forecasting, and controls.



Note. **In Week 3, the Emotion-Forecasting group switched from simple monitoring to future-oriented visualizations and mood-enhancing recommendations.**

which participants simply logged their mood and behaviors for 21 days. Second, to test our new system intervention we included an *emotion-forecasting* condition in which participants again logged their mood and triggers each day. However, after 14 days of simple monitoring, emotion-forecasting participants were presented with interactive visualizations and activity recommendations to support future thinking (Figure 4). Third, a *do-nothing* control group simply submitted pre–post surveys and did not participate in any intervention. This final group controlled for simple expectation effects of participating in a study and to control for the general effect of time. The same recruitment materials were used for all conditions to avoid recruitment bias.

We informed all participants in both the monitoring-only and emotion-forecasting conditions that they would eventually receive mood visualizations. We did this to encourage similar logging behavior and avoid the possibility that monitoring-only participants might log differently if they believed their data did not contribute to analytics. However monitoring-only participants were not given their interactive visualizations until *after* study completion. We postponed their exposure to the additional features so that seeing their visualizations had no impact on the monitoring-only participants' survey responses or interviews. We tested the effects of the interventions by measuring changes in positive and negative emotions assessed in pre- and posttest surveys, as well as by analyzing user logfiles for emotional content and mood ratings.

Participants

Eighty-three participants were recruited through Craigslist, Facebook, Quantified Self forums, university classroom announcements, and flyer advertisements. Participants were sent the following: pretest surveys, an instructions document, and daily text reminders to submit at least two mood records per day for a total of 42 entries. Participants were excluded if they did not provide entries in the final week or less than half of the required daily entries (n = 32; M entries = 8.65). The final intervention samples consisted of 36 individuals who were recruited initially and randomized into the two intervention conditions (monitoring-only or emotion forecasting), equalized across gender and pretest well-being scores. These compliance rates are similar to those reported in other studies (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016). A separate group of 24 participants were recruited through the same venues using identical advertisement materials to serve as do-nothing controls. The entire final sample consisted of 60 participants (23 male), with a mean age of 35.42 (SD = 12.02). Participants received compensation per level of involvement, which was advertised as \$5 to submit pre–post surveys and \$5 to participate in daily logging. As a consequence of this incentive structure, participants received either \$10 for the full intervention or \$5 for the do-nothing control. Participants were blind to which group they were in and were not informed that there were different groups. We had previously obtained Institutional Review Board approval for this study.

Procedure

All participants were told that the research goal was to beta test a new technology to help regulate mood and improve well-being. They first completed an online pretest, consisting of a set of surveys to assess baseline emotional well-being and behavior frequencies with enjoyment ratings for those behaviors. We then e-mailed intervention participants a web link to EmotiCal with login information and documentation for how to submit entries and expectations for study participation.

To maintain compliance, researchers individually contacted participants by text and phone within the 1st week to ensure they were consistently submitting entries and to address any technical errors or confusion over the study instructions. Following procedures used in similar studies (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016; Konrad, Tucker, et al., 2016), two researchers additionally called each participant once per week to check that they were continuing to make entries and not experiencing any problems with the application. Participants also received an automatic text message reminder on days they did not make an entry. We also scanned server logs to confirm that participants were indeed making daily entries; correctly following instructions; and, most important, not submitting content that would raise concern (e.g., self-harm or threats of suicide). Fortunately, we had no cases in which researcher intervention was necessary for participant safety.

Three weeks after the start date, participants were contacted by e-mail to answer the posttest survey; they were debriefed, thanked, and given the opportunity to delete or modify any logged data they wished to keep private before data analysis. They were also invited to optionally participate in a 1-hr follow-up audio interview, conducted over conferencing software.

Instructions and Measures

Pretest Materials. All participants completed pre–post intervention surveys and a consent form online. The pretest included demographic questions and surveys to measure their emotional profile (Positive and Negative Affect Scale [PANAS]), psychological needs (BPNS), self-awareness, and perceived choice over behavior (Self-Determination Scale [SDS]). Participants also generated enjoyment ratings for various possible activities by completing the Pleasant Activities Schedule.

Our main goal was to measure differences in the frequency of positive and negative emotions (PANAS), as well as logfile ratings of mood, resulting from our interventions. In addition, we chose the SDS to measure changes in perceived choice over behaviors and self-awareness. The BPNS was included as a pretest measure so that we could better tailor hand-coded activities to supplement the automated activity recommendations.

1. *Positive and Negative Affect Scale*—The PANAS (Watson, Clark, & Tellegen, 1988) is a 20-item scale used to assess separate dimensions of hedonic emotional well-being (positive vs. negative affect). Participants are asked to rate how often over the past week they experienced 10 negative emotions (e.g., distressed, guilty, scared, etc.) and 10 positive emotions (e.g., excited, enthusiastic, proud, etc.). Ratings are given on a 5-point scale ranging from 1 (*very slightly or not at all*) to 5 (*extremely*). This scale was intended to assess participants' general emotional profile as recalled from memory.
2. *Self-Determination Scale*—The SDS (Sheldon, Ryan, & Reis, 1996) consists of two subscales, the Awareness of Self scale and Perceived Choice scale. The Awareness of Self scale measures awareness of one's feelings and sense of self (e.g., "My emotions seem to belong to me."). The Perceived Choice scale measures the extent to which people feel they have control over their own behavior (e.g., "I always feel like I choose the things I do."). Ratings are given on a 1-to-5 scale, with scores averaged for each subscale. This scale was intended to assess the extent to which participants felt control over their actions and felt self-aware.
3. *Pleasant Activities Schedule*—Participants also completed an adaptation of the Pleasant Activities Schedule (MacPhillamy & Lewinsohn, 1982) to estimate how often they engaged in 39 possible behaviors over the previous 2 weeks. Activities include entertainment, socializing, outdoor exercise, and so on. Participants also gave frequency ratings on a 5-point scale ranging from 0 (*not at all*) to 4 (*10 times or more*), and enjoyment ratings of the activity ranging from 0 (*not at all enjoyable*) to 4 (*very enjoyable*).
4. *Basic Psychological Needs Scale*—The BPNS (Deci & Ryan, 2000) is composed of three subscales to measure autonomy, competence, and relatedness. We chose a nine-item BPNS scale with three questions in each of the subscales to obtain a profile of individual user needs. This information motivated *needs*-based recommendations for emotion-forecasting participants.

Posttest Measures. In the posttest survey, we readministered both the PANAS and the SDS to determine the effects of study participation/intervention on emotional well-being, perceived choice over actions, and self-awareness.

As a manipulation check, we also asked participants to estimate the number of activities they engaged in during the prior week to improve their moods (Activities Engaged) and asked how successful these activities were in improving their moods (Activities Success). Participants provided ratings for the following two questions on 7-point scales: (a) Activities Engaged—"Over the past week, how often have you engaged in specific activities to improve your mood? For example, on realizing that you are feeling negative you might decide to exercise or call a friend." Responses ranged from 1 (*not at all*) to 7 (*almost always when I feel negative*). (b) Activities Success—"How often are these activities successful at improving your mood?" Responses ranged from 1 (*never*) to 7 (*always*).

The survey also asked participants to rate mood forecast accuracy and quality of activity recommendations (7-point scales). Participants were also given a final option to provide a freewrite response about their experiences: "Was there anything you learned from this study? Did it change or not change your outlook on your emotions?"

Text Analysis of Logfile Content

During the intervention, for both emotion-forecasting and monitoring-only groups we also collected both freewrite text content and mood ratings for each logfile entry. We used Linguistic Inquiry Word Count (LIWC; Pennebaker, Booth, & Francis, 2007) to automatically analyze the text in people's logfiles. LIWC is a widely used lexical analysis tool that automatically classifies words according to their semantic category. It has good internal reliability and external validity when compared with human judges (Pennebaker et al., 2007; Pennebaker & Francis, 1996; Tausczik & Pennebaker, 2010). Although the LIWC dictionaries are able to measure up to 72 linguistic categories, we focus here only on categories that directly concerned our hypotheses and that have been demonstrated to relate to emotional well-being in previous studies. Specifically, we targeted word categories that provided evidence of changes of emotion, understanding, and insight. We analyzed examples of words expressing both positive ("happy," "joy," "love," etc.) and negative emotion ("hate," "die," "despise," etc.; Campbell & Pennebaker, 2003). Emotion forecasting was also intended to promote understanding and insight, which we measured through usage of insight words ("think," "know," "consider," etc.), causation ("because," "infer," "produce," etc.), and cognitive processes ("cause," "know," "ought," etc.; Klein & Boals, 2001; Petrie, Booth, & Pennebaker, 1998). Prior work (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016; Konrad, Tucker, et al., 2016; Peesapati et al., 2010) shows that use of these words relates to improved emotion regulation and positive changes in well-being.

3.3. Results

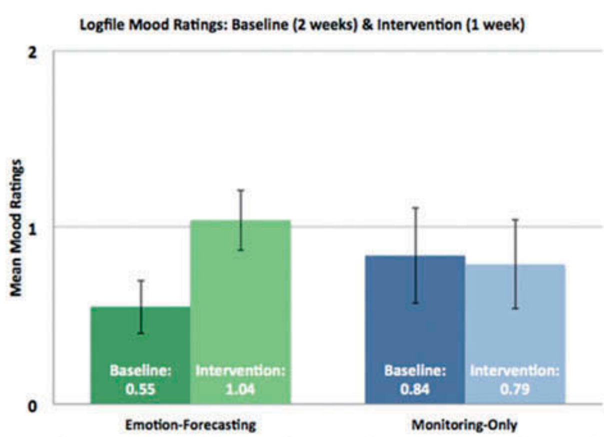
Logfile Content: Emotion-Forecasting Participants Had More Positive Mood Records with Greater Use of Cognitive Mechanism and Insight Terms

Our primary research question concerned the benefits of emotion-forecasting over current monitoring-only approaches. To assess the effects of these planning components we compared the emotion-forecasting with monitoring-only group across the following measures: changes in logfile mood ratings, logfile linguistic content, SDS, PANAS, and Activities Engaged/Perceived Activities Success.

We began by analyzing logfile text and mood ratings. In line with previous findings (Faurholt-Jepsen, Munkholm, Frost, Bardram, & Kessing, 2016; Kahneman, 2000, Tsanas et al., 2016), we expected these to be accurate measures of intervention success, as these are collected twice each day, assessing participants' real-time evaluations of current moods and recently experienced events. We first examined changes in mood ratings and logfile content in the first 2 weeks versus 3rd (final) week of the intervention, as a within-participant comparison. This process was to compare differences in logfile mood before versus after the forecasting group received visualizations and activity recommendations. We then compared these differences with results from the monitoring-only group, which we expected to show fewer changes.

Logfile Mood Ratings. We conducted a *t* test to evaluate changes in logfile mood across conditions (see Figure 5). Changes were calculated as the within-subject difference in mood ratings between baseline (2 weeks) and intervention (1 week) phases. For example, a participant with an average baseline mood of .5 and intervention mood of .75 would have a logfile change rating of .25. These within-subject differences between baseline and intervention period were then compared between the two experimental conditions (monitoring only and emotion forecasting). We found a significant difference in logfile mood change across conditions. Forecasting participants on average increased daily mood ratings by 0.50 (SD = 0.55; baseline: M = .55, SD = 0.72; intervention: M = 1.04, SD = .79).

FIGURE 5. Mean mood ratings and standard error bars for emotion-forecasting and monitoring-only conditions.



Note. **Graph shows that forecasting improves mood. Baseline phase was 2 weeks in which both conditions used the monitoring user interface. The intervention phase was 1 week in which monitoring-only continued to use the same user interface as baseline and emotion-forecasting participants were presented with additional visualization and recommender support.**

In contrast, monitoring-only participants' moods changed on average by –0.06 (SD = .91; baseline: M = .84, SD = 1.01; intervention: M = 0.79, SD = 0.92), t(34) = 2.290, p = .028. In other words, forecasting participants displayed greater improvements in logfile mood ratings, whereas monitoring-only participants dropped slightly during the final week of the study. There was no significant difference in average mood for the baseline period (first 2 weeks) to alternatively explain this difference, t(34) = 1.042, p = .305.

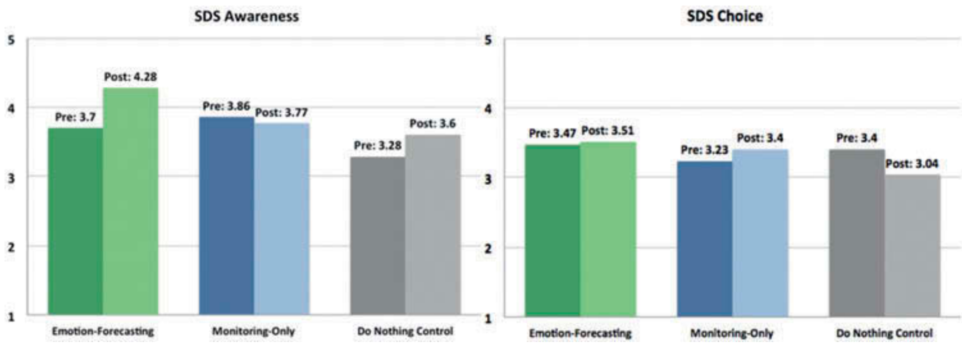
Logfile Text Content. We next used LIWC to analyze differences in the textual content of mood entries. As expected, there was a significant difference between conditions, with forecasting participants using more cognitive mechanism terms, t(34) = 2.855, p = .007, and insight terms, t(34) = –2.589, p = .014. Cognitive mechanism and insight terms indicate that participants are actively processing their mental experiences to analyze which activities explain changes in mood. More frequent insight terms in the emotion-forecasting group also suggests that our interface was supporting key reflective processes. For example, as emotion-forecasting participant 42968 described in a mood entry, "I still feel bad *because* I spent too much money and was out too late." There were no word count differences between conditions that could alternatively explain these results, t(34) = .574, p = .570.

PANAS and SDS Comparisons: Emotion-Forecasting Participants Had Higher Ratings of Self-Awareness, With No Differences in Perceived Choice or PANAS Scores

We next compared pre- versus posttest survey changes in our emotion-forecasting, monitoring-only, and do-nothing controls. We included the do-nothing control group to allow for possible benefits occurring due to simply participating in a study that asks people to think about their moods or to account for possible effects of time. Recall that do-nothing controls simply completed surveys but did not use the system.

Self-Determination Scale. There were significant differences across conditions for SDS awareness and trending differences for SDS choice (see Figure 6). We conducted a 3 (condition: emotion forecasting vs. monitoring only vs. do nothing controls) × 2 (time: pretest vs. posttest) multivariate analysis of variance (MANOVA), with condition as a between-subjects independent variable and time as a within-subjects independent variable, with pre–post SDS Choice and SDS Awareness scales as the dependent. Overall there was an interaction between time and condition, F(4, 112) = 3.870, p = .006 (Pillai's trace, V = .243), indicating differences between SDS scale outcomes by condition. Univariate F tests showed a significant interaction between condition and time for SDS Awareness scales, F(2, 56) = 5.276, p = .008, with the emotion-forecasting condition increasing in

FIGURE 6. Mean scores on SDS Awareness and SDS Choice subscales.

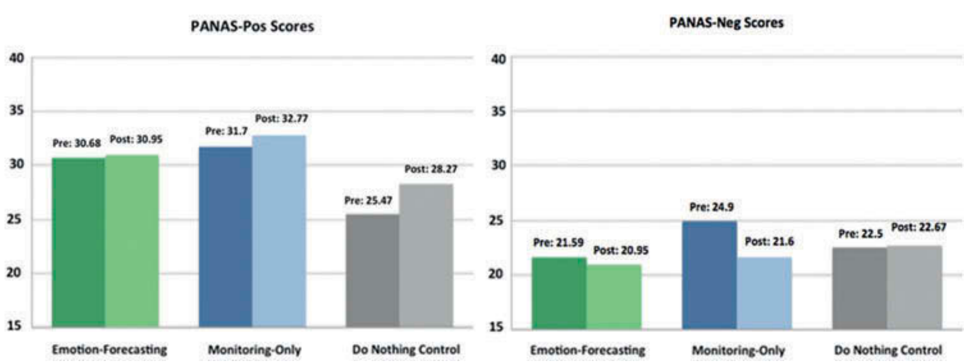


***Note*. Measures were taken immediately before and after the 3-week study period. Awareness increases for forecasting group (left panel). Choice decreases for controls compared with two intervention groups (right panel).**

SDS Awareness scores pre–post. In addition, the test indicated a trend for SDS Choice scales, $F(2, 56) = 3.004$, $p = .058$, with intervention conditions marginally increasing and the do-nothing control decreasing in score (see Figure 6).

Positive and Negative Affect Scale. We conducted a similar MANOVA examining the effects of condition and time on positive and negative PANAS scores (see Figure 7). However, this showed no difference across conditions in pre–post changes in the PANAS scale, $F(4, 94) = 0.633$, $p = .640$ (Pillai's trace, $V = .052$). One possible reason for this might be that the PANAS scale, by probing the *last week*, was not sensitive to very recent changes in emotions. Recall that participants in the forecasting condition had been using the system for just a

FIGURE 7. Mean scores on the Positive and Negative Affect Scale (PANAS).



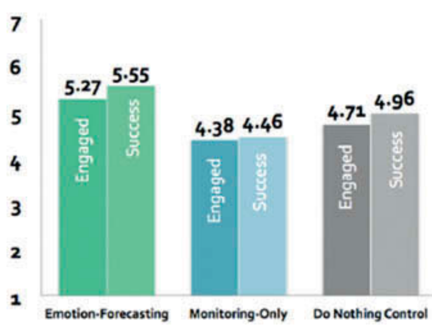
***Note*. There were no differences between conditions.**

week. In addition we chose the validated PANAS-20 scale, which, although including multiple types of positive and negative emotions (e.g., scared, guilty, proud, enthusiastic), does not specifically prompt for ratings of happiness or sadness.

Frequency and Impact of Activities: Emotion-Forecasting Participants Reported Engaging in More Activities and These Being Successful at Improving Their Moods

Activities Engaged and Success at Improving Mood. We also compared responses to the survey questions concerning deliberate activities engaged in and their impact on mood (see Figure 8). These questions were presented after participants had experienced the intervention. A three-way analysis of variance including the do-nothing control showed no significant difference for activities engaged ratings, $F(2, 56) = 2.582$, $p = .085$, there was, however, a significant difference for ratings of activity successfulness, $F(2, 56) = 4.162$, $p = .035$. Emotion-forecasting participants had higher posttest ratings of their activities being more successful at improving their mood ($M = 5.55$, $SD = 1.01$) compared to monitoring-only participants ($M = 4.46$, $SD = 1.66$) or do-nothing controls ($M = 4.96$, $SD = 1.04$). Thus, although forecasting participants were no more likely to choose activities overall, they nevertheless felt that the activities they chose were successful at choosing activities to improve mood.

FIGURE 8. Mean scores for activities engaged and activities success ratings given on 7-point scales.



***Note.* Activities engaged does not differ across conditions, but forecasting participants stated that they were more successful at planning activities to improve mood.**

Potential Confounds

Preintervention Tests and Potential Confounds. We additionally checked for possible differences in preintervention participant characteristics or confounds that could alternatively explain these results. There was no difference in total word count, average word count per entry, or number of mood tracking entries submitted between conditions (ps = .442–.570). Age, gender, pretest ratings of effort, and average word count per entry did not correlate with any dependent variables (all ps = .130–.925). Number of entries did trend to changes in logfile mood, $r(36) = .296$, $p = .080$. An ANCOVA controlling for number of entries showed the mood difference between conditions to still be significant, $F(1, 33) = 4.344$, $p = .045$. A series of t tests examined differences between conditions for preintervention characteristics. Across the two intervention conditions, we found no significant differences for age; gender, $\chi^2(1)$, $p = .729$; effort; or pretest well-being scores (all ps between .297 and .995). However, the do-nothing control had lower pretest well-being scores on the positive emotion PANAS scale ($M = 24.47$, $SD = 4.02$) compared to both the emotion-forecast ($M = 30.7$, $SD = 7.1$) and monitoring-only ($M = 31.7$, $SD = 12.1$) conditions and lower pretest ratings of effort to improve mood (ps = .007–.008).

Engagement and Perceived Accuracy: Emotion-Forecasting and Monitoring-Only Participants Responded Positively to the System Interventions

The previous analyses show positive system effects of emotion forecasting on mood, insight, self-determination, and ability to choose mood-enhancing activities. Nevertheless, we also wanted to explore these effects in more depth to better understand which elements of our design were most effective in promoting new activities to improve mood. We examined relations between how participants used EmotiCal and differences in outcomes. In addition we evaluated user behaviors, including the number of trigger activities reported, differences in what activities were planned, and opinions about the forecasting visualization with respect to user satisfaction.

Activity Planning. On average, emotion-forecasting participants created 8.64 ($SD = 5.77$) activity plans. We were very interested in looking at what types of activities were planned and how these planned activities related to well-being outcomes. Overall, we found that *needs*-based recommendations corresponded to better outcomes for participants than *history*-based recommendations. Making a greater number of needs-based activity plans corresponded to higher postintervention evaluations of activity successfulness, $r(22) = .481$, $p = .023$, as well as how often these planned activities were completed, $r(22) = .498$, $p = .018$. History-based activities did not show either of these correlations (ps = .328–.951). One reason for this may be that history-based activities, being actively recorded by participants, become familiar and so are relatively well understood. However a recommendation to engage in a less obvious but still enjoyable activity, based on psychological *needs*, might generate greater benefits. We return to this point in our discussion.

Trigger Activity Tracking. On average, participants tracked 2.3 triggers per mood entry (SD = 1.21; note that one extreme outlier with 553 factors tracked, averaging 13.8 factors per entry, was not included in this analysis). The number of factors tracked significantly correlated to changes in logfile mood, $r(35) = .395$, $p = .019$; activity helpfulness ratings, $r(21) = .437$, $p = .048$; and activity engagement self-reports in the postintervention survey, $r(21) = .686$, $p = .001$. Overall, it appears that participants who tracked more factors in their records and were more attentive in their mood monitoring had better mood outcomes. This suggests that we should encourage participants to track more specific details about behavior/triggers and mood, and future designs need to provide mechanisms that encourage careful tracking of trigger activities that can promote mood benefits.

Forecasting Accuracy. We also examined participants' judgments of the accuracy of mood predictions, exploring how these judgments related to various mood and usage behaviors. In terms of predicting future moods, emotion-forecasting survey respondents generally found these future predictions accurate (M = 4.95, SD = 0.89, on a 7-point scale). We found that perceived prediction accuracy correlated with (a) the number of activities engaged, $r(22) = .396$, $p = .068$; (b) activities' helpfulness ratings, $r(22) = .419$, $p = .052$; and (c) factor helpfulness ratings, $r(22) = .634$, $p = .002$. In other words, participants who judged the mood models as more accurate were also more likely to act on these and obtain well-being benefits. Consistent with other work showing the importance of motivation (Hollis et al., 2015), if participants believe the system to be accurately tracking measurable mood benefits, they are more likely to engage and show greater compliance in planning activities.

Follow-Up Interviews and Open-Ended Survey Responses

Two days after the study, all system participants were contacted about a voluntary follow-up interview. Fifteen participants (seven monitoring-only, eight emotion-forecasting) volunteered to discuss their experiences. All interviews were conducted individually over audio-only conferencing software and were recorded and transcribed. We additionally collected freewrite responses in the posttest survey to learn more about participants' experiences and better understand their responses to survey ratings about the system.

General Response to the System Interventions. Participants in both intervention conditions could write an open-ended response to the question, "Was there anything you learned from this study? Did it change or not change your outlook on your emotions?" Overall, participants in both conditions were highly positive about the system in their responses to this question. Of the 20 participants in the emotion-forecasting condition who submitted survey responses, 19 described positive changes due to using the system and one user made a negative comment. The response was marginally less positive for the monitoring-only system. Of the 13 participants in the monitoring-only condition who submitted survey responses, eight spoke positively of the system, three provided neutral suggestions, and two spoke negatively. We now explore how people responded to individual system features.

Positive Response to Tracking Activities. The responses to activity tracking were overwhelmingly positive. Participants in both intervention conditions described how the logging process was useful for identifying trigger activities that influence mood. One monitoring-only participant observed this motivated behavior change:

> It enabled me to know what influenced the way I felt. For instance, *knowing that I slept [well] would help improve my mood was motivating to get in bed earlier*. Also, I was able to see if other things influenced my mood to feel worse/better. [75961]

Another monitoring-only participant highlighted the benefits of tracking activities on controlling future mood:

> If you address why you are in a mood and actually think about what causes these things you can avoid them in the future by taking steps to not do the things you've done to end up in a negative mood. … That's beneficial and helpful. [41586]

Emotion-forecasting user 15213 similarly reacted positively to this diagnostic support:

> You could see what put you in a mood, what was responsible for that … maybe a couple things. I was in a good mood today because I got free food and I worked out, but at the gym I hit a new max, so that made me feel a lot better than the food itself.

Again, survey freewrite responses across both conditions suggested similar benefits —"It showed me that what I do normally everyday does affect my mood" [EF: 13489] and "It made me consider the factors and what I people and activities I do throughout the day. It changed my emotions in a positive way" [MO: 12345].

Responses to Activity Planning. Emotion-forecasting participants were asked to rate the helpfulness of planning future activities. Responses on a 7-point scale ranging from 1 (*not helpful at all*) to 7 (*extremely helpful*) were generally positive ($M = 5.1$, $SD = 1.55$). Multiple participants described immediate mood benefits to planning activities: "Optimism. … I put something knowing that it would be positive, I think that as I was doing *it was a good feeling, right when I was adding it*" [21212]. Another participant highlighted mood benefits from adding activities: "It made me make more time for the fun stuff to improve my mood" [dano62]. The following forecasting participant also drew attention to the importance of recommended intervention activities being practical: "As a grad student I need an outlet to relieve stress. These activities are *easy to do and effective*" [15213].

Another emotion-forecasting participant stated the benefit of having a clear objective to improve mood: "It made me realize when I was getting into a negative spot. It helped me to turn that around and be on the lookout to not be on that path again. It gives me goals to achieve to keep me happy" [420]. Some participants additionally described changing specific behaviors to improve mood. This participant described improving his diet to avoid negative mood consequences.

> I have been trying to eat healthier lately, there was a couple of times I went off the wagon. … I would wake up either the next morning or throughout the night, it would cause me to have a crappy night's sleep, and that would definitely affect me the next day. [21212]

Responses to Activity Predictions. We already noted that forecasting participants generally judged mood predictions to be accurate. As we had anticipated, these participants also found it useful and motivating to see the impact of adding activities in the visualization:

> I feel amazing. … I would add that to my planner and then I get to cross it off. And I'm like "Awesome!" Here's a visualization of what I did for myself. … Look at it! It actually did bring up my mood! [smst211]

Emotion-forecasting participant 15213 described how viewing the visualization "made me try to be as positive as possible."

Increased Self-Awareness and Feelings of Control. Consistent with our design goals, forecasting participants also spontaneously discussed changes in their emotional self-awareness, control, and increased perspective, which they attributed to EmotiCal's design features. They described moments of self-awareness and greater control resulting from being able to easily track how different factors influenced their mood: "Makes me *more aware*, opportunity to control" [EF: 15213]. The forecasting visualizations were also critical, as emotion-forecasting user 71153 described being able to *directly see* how different triggers would influence future mood:

> To be *more aware* of the different things that affect my mood. … I think that's very valuable. Perspective for me to see something that gives a source of how things would affect my mood. For instance sleep, [my job and partner] … I don't think I really gave much thought about that until I could *actually see* it in a visual sense.

These planning features increased self-insight, leading some participants to feel that they had genuine control over future moods, as well as their emotions more generally. This emotion-forecasting participant described how planning made him "more self-aware and realize that *you really do have the ability to change your mood.* You know that you do have the power to actually you know, go from a zero to a plus one … *makes you accountable*." The same participant insightfully concluded his interview by pointing to the key role of forecasting in supporting reflection to promote the self-knowledge needed for changing behavior: "That's what I liked about it is that *you know yourself, and you're kind of self assessing and self reporting*, which is important, I think *you've got to know yourself before you can make meaningful changes*" [21212].

Both versions of EmotiCal helped participants feel that they had more control over their moods. As this emotion-forecasting participant stated, "I learned that I am in complete control of the things that can help me feel better. Empowerment" [6107152]. Similarly, this monitoring-only participant explained, "It made me see that I have more control of my emotions when I am aware of what is triggering them" [68950]. Emotion-forecasting participant 9847795 also stated in the survey that she learned "that I'm always in control on how I feel and that I should think positively. Even if something has me down, I have options to make me feel better." Participant EF:15213 described a changed outlook in her interview: "… that I should plan my days better and look out for what things affect my mood the most, and try to control them rather than have them control me."

Future Improvements. Some design issues emerged from the interviews and survey data. Emotion-forecasting improved mood overall and provided insights leading most participants to be positive about the system. However, the effects of our intervention could have been even stronger. First, some participants did not follow through with planned activities, as seen in half of survey respondents, who reported completing less than half of the plans they made. As palo16 described in an interview, "I'd just forget about it and not do the activity. … It would be really cool if there was a reminder system, like with email or texting …"

There were also low reports of actively examining mood predictions (M = 2.09, SD = 0.92 on a 7-point scale), suggesting that more work is needed to engage participants to actively process these visualizations. Although we had worked carefully to personalize recommended activities, some participants still found the number of activity recommendations restrictive and would have appreciated a more expanded list of options. Whereas some obviously enjoyed recommendations for activities they may not otherwise have planned, others also expressed a desire for more open-ended planning. Yet others found predictions initially difficult to understand—"It did not occur to me it was a prediction at first but generally it was right on" [EF: 73737]—or had issues with accuracy —"Not sure if they were really accurate, but fun to look at. Gave me optimism" [EF: 21212]. Although the system produced clear benefits and was generally very positively evaluated, issues of prediction credibility, activity recommendation options, reminders, and visualization clarity are important points of improvement for future iterations of forecasting systems, and we return to these issues in our conclusion.

3.4. Summary

EmotiCal was successful in encouraging future activity planning and improving emotional well-being as evidenced by improved logfile mood ratings. Analysis of logfile content showed that emotion-forecasting participants produced more language indicative of insight and cognitive mechanisms. Open-ended survey responses and interviews showed too that emotion-forecasting participants generally found the visualizations engaging and enjoyed planning activities. Although emotion-forecasting participants didn't plan more activities overall, posttest survey responses showed that nevertheless they believed the activities they chose to be more successful in improving their mood, compared to participants who simply monitored their activities or were controls. As seen in the interviews and surveys, tracking activities also produced benefits across both intervention conditions. We return to these points in our final discussion.

4. FIDELITY: CONSEQUENCES OF SELECTIVE EVENT RECORDING AND REFLECTION FOR WELL-BEING

EmotiCal aims to improve daily well-being by providing end-user analytics and positive actionable recommendations. A complementary approach to improving well-being is reflective memory. Rather than proposing new enjoyable *actions*, memory systems encourage participants to reflect on prior personal *experiences* to improve well-being. Here there is a profusion of commercial systems that gather prior social media posts or photos, repurposing these to participants to reflect on and share. As with EmotiCal, a critical design question is again how we recommend and select these prior experiences. Exactly which past experiences should participants reflect on to improve well-being? Current commercial systems use simple algorithms such as time or popularity to select prior experiences. Time-based systems might therefore recommend that participants reflect on experiences that happened exactly 1 year ago (Timehop, On This Day) or popularity-based systems suggest highly popular events, for example, those posts that received most likes or comments in the last year (Look Back).

In the context of well-being, however, another important consideration is the *emotional valence* of these reflected-upon events. Should we recommend that people reflect only on positive experiences, or should they also confront more negative aspects of their past? We have already reviewed the extensive literature on emotional writing showing that pen-and-paper reflection on negative past experiences has significant well-being benefits (Frisina et al., 2004; Harris, 2006; Meads, 2003; Pennebaker & Chung, 2011; Smyth, 1998). However, other work on *digital* reflection indicates clear design challenges, for example, when participants inadvertently reexperience disturbing aspects of their past (Haimson et al., 2015; Zhao & Lindley, 2014). Sas and Whittaker (2013) documented how participants undergoing a breakup were upset when they accidently encountered posts, photos, texts, or e-mails that reminded them about their ex. Another important consideration here is the *intensity* of the reflective experience. Pen-and-paper studies of reflection show that benefits are different when participant reflect on emotionally intense versus milder events (Gidron et al., 1996).

We therefore addressed the following questions about how we select events for recording and reflection to improve the design of reflective memory systems:

1. Does recording and reflecting on negative experiences improve or detract from well-being? Are these effects different from recording and reflecting on positive events?
2. How does the intensity of the target event affect well-being? Are there differences between recording and reflecting on extremely negative versus moderately negative events?
3. We were also interested in differences between recording and reflection. Reflective systems require that participants record experiences in order to reflect on them later, and it may be that the simple act of recording itself brings well-being benefits, even without reflection.

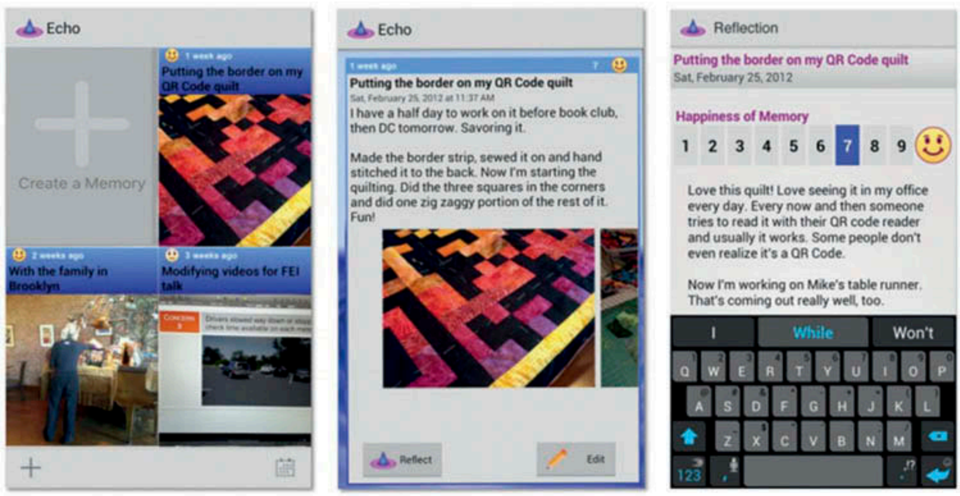
To explore these questions we evaluated four design implementations of a reflective memory application, Echo, in a month-long intervention and assessed well-being effects. The first design simply encouraged participants to record positive experiences. This implementation did not support reflection, and participants never saw these positive recordings again. In a matched design, participants were encouraged to record only negative events. A third and fourth design supported not only recording but also either because the third design focused only on positive reflection and the fourth design only focused on negative reflection. The third design encouraged participants to record and later reflect on exclusively positive events. The fourth matched design encouraged recording and reflection on only negative events. In each case, we wanted to explore effects on well-being, using intervention methods developed and deployed successfully elsewhere (Isaacs et al., 2013; Konrad et al., 2015; Konrad, Isaacs, et al., 2016; Konrad, Tucker, et al., 2016).

4.1. Intervention

The Echo Application

Echo (see Figures 9–11) is a smartphone application that allows participants to create rich event records of their choosing; rate their emotional reaction to those events; and, in the case of the reflection conditions, revisit these records later for subsequent reevaluation (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016). An event record consists of a label, a short description of the event, and an emotional rating of that event (from 1 [*highly negative*] to 9 [*highly positive*]). Users also have the option to append pictures, audio, or video of the event (see Figure 10). To ensure consistency between an individual's emotional ratings over time, all participants create a Personal

FIGURE 9. The Echo Smartphone Application, showing record+reflect process.



Note. Left panel **home screen: Participants record a new experience by clicking on the large + in the upper left. *Middle panel* shows a completed Echo event record, which consists of a header, textual entry, emotion rating (☺ = 7) and image. *Right panel* shows participant reflecting, by rating their current emotional reaction to the initial record (again a 7) and providing a new textual reappraisal.**

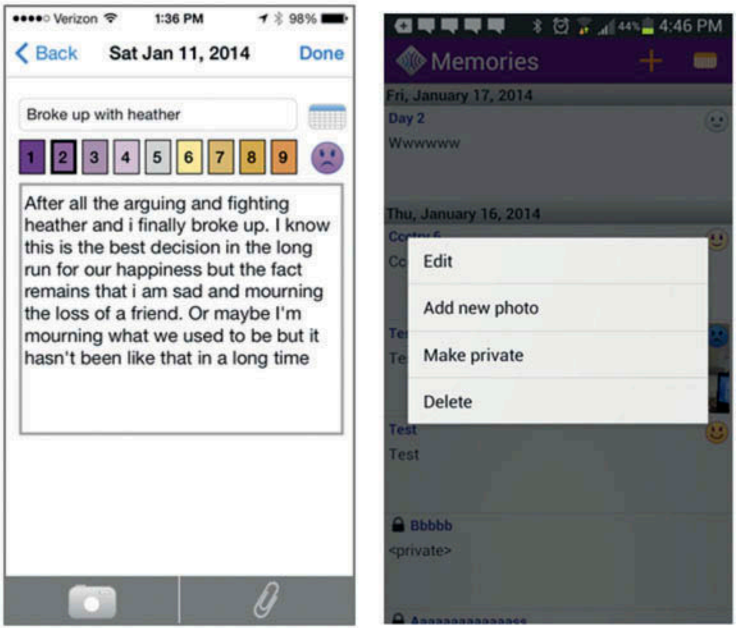
Emotional Scale at the beginning of the study. The basic recording interface is shown in Figure 10 and the actual reflection interface in Figure 9.

For this research, we designed and deployed four versions of Echo: (a) a positive record-only version, (b) a negative record-only version, (c) a positive record+reflect version, and (d) a negative record+reflect version. Each of these options was supported on iPhone and Android platforms.

Participants

There were 105 participants recruited through Facebook and a university mailing list, using a snowball recruitment strategy. The average age was 22.9 (SD = 5.9), with 29 men and 76 women. Participants were randomly assigned to one of four conditions, balanced across gender, age, and pretest well-being scores, and were not informed there were different groups. Final condition allocations resulted with 28 in the positive-record-only group, 23 in the negative-record-only group, 27 in the positive-record+reflect group, and 27 in the negative-record+reflect group. There were no demographic or pretest well-being differences across these groups (t tests showed all ps between .390 and .997). We had previously obtained Institutional Review Board approval for this study.

FIGURE 10. The left image is an example of the event-recording interface, which supports adding photos and other media files to an event record.



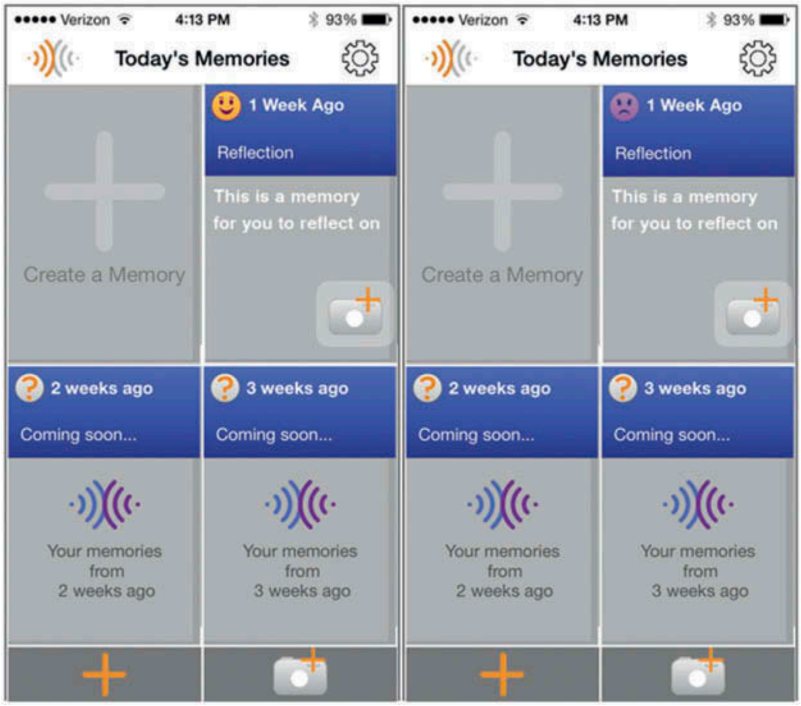
Note. **Participants could also enter an event title for the recording, supply body text, and apply an emotion rating ranging from 1 (*highly negative*) to 9 (*highly positive*). The right image displays options participants had to edit, delete, or privatize past records.**

Assessment

Well-being was assessed at pre- and posttest using three standard well-being scales. Three complementary surveys were chosen to triangulate multiple dimensions of well-being: positive affect/hedonic happiness (Subjective Happiness Scale [SHS]), eudaimonic happiness (Ryff's Scales of Psychological Well-Being [RPWB]), and the presence of rumination (Ruminative Responses Scale [RRS]). All scales are widely used and have high discriminant and convergent validity and test–retest reliability in multiple populations. We use different scales from the previous study, as our interest here was also in longer term well-being rather than exclusively focusing on current mood and emotion regulation.

Subjective Happiness Scale. The SHS consists of four items that assess global subjective happiness using absolute ratings, as well as ratings of self relative to perception of others (Lyubomirsky & Lepper, 1999). Participants evaluate their general happiness levels rather than how happy they have been across any specific period. An example item is, "Compared to most of my peers, I consider myself …", which has response categories ranging from 1 (*less happy*) to 7 (*more happy*).

FIGURE 11. The Echo Smartphone Application reflection interface.



***Note*. The left panel is an example of the reflection user interface for positive record+reflect participants. The right panel is an example of the reflection user interface for negative record +reflect participants.**

Ryff's Scales of Psychological Well-Being. The RPWB reflects six facets of eudaimonic well-being: (a) autonomy, (b) environmental mastery, (c) personal growth, (d) positive relation with others, (e) purpose in life, and (f) self-acceptance. Responses are totaled for each of the six subscales (higher scores representing more mastery in that area), and a total score is formed by either summing or averaging these scores. An example item is, "When I look at the story of my life, I am pleased with how things have turned out," which has response categories ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). The RPWB scale can be 18, 54, 84, or 120 items long. For this research, the 54-item version was used to reduce participant burden and because it is considered to have greater reliability than the 18-item version (Ryff & Keyes, 1995; Van Dierendonck, 2004).

Ruminative Responses Scale. The RRS consists of 22 items designed to assess individual differences in rumination. Rumination is defined as a self-focused method for coping with negative mood that involves repetitive and passive focus on one's negative emotions (Treynor, Gonzalez, & Nolen-Hoeksema, 2003). A total rumination score is formed by summing the scores on each item. An example item is, "How often do you think about all your shortcomings, failings, faults, mistakes?" which has response categories ranging from 1 (*almost never*) to 4 (*almost always*).

In addition to these surveys deployed before and after our intervention, as in Study 1 we also directly assessed participant mood twice a day, using a Personal Emotional Scale.

Personal Emotion Scale. We relied on a simple single (1–9) scale method of recording emotional responses to events (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016; Konrad, Tucker, et al., 2016). This was intended to be lightweight, as we did not want to overburden participants with more complex emotion metrics. Each participant would make multiple emotional entries a day, sometimes using the application while on the move, and we wanted to make entry creation straightforward. Nevertheless it was important that participants were consistent in their emotion ratings of events. We therefore used a normative rating method to improve within-participant reliability of the 9-point mood scale, an approach that has been used successfully in multiple prior studies (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016; Konrad, Tucker, et al., 2016)). Prior to participating in the study, participants supplied normative event examples, specifying a concrete personally experienced event for each of the 1–9 emotion ratings on the scale. For example, a rating of 1 might correspond to a very negative experience such as a relationship breakup, a 5 might correspond to a neutral experience such as a work meeting, and a 9 might correspond to a positive experience such as a job promotion. Participants were instructed to use this normative personal scale throughout the study when rating emotional reactions to event records and reflections.

Logfiles Containing Text of Recordings and Reflections. As in Study 1, we also collected the text of participants' recordings and, where relevant, their reflections. To preserve privacy, participants also viewed and edited their own posts before sharing these with the researchers.

Procedure

The experiment was a randomized pretest–posttest field study with each of the designs (positive record-only, negative record-only, positive record+reflect, negative record+reflect) as the manipulation and three well-being scales as the dependent variables (SHS, RPWB, RRS). Participants completed the pretest (Time 1) survey online remotely (through www.surveymonkey.com) and the same survey at posttest (Time 2) after using Echo for 1 month.

After completing the pretest survey, participants were randomized to a design condition and then sent an instruction document that explained in detail their responsibilities for the study. Participants were instructed to create their Personal Emotional Scale and save this for future use. In addition, the researchers called each participant to go over the instructions verbally. To further encourage compliance, participants were called weekly to briefly check that they were using the scale and understood the intervention protocol (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016; Konrad, Tucker, et al., 2016). After installing Echo and completing an initial phone call, participants began the 28-day intervention. Instructions depended on what condition they were assigned to.

The negative record-only condition was instructed to make recordings only between 1 (*extremely negative*) to 4 (*mildly negative*) on the Echo Emotion scale. Whereas emotion ratings of 5 and higher were still visible in the interface, these options were not selectable.

Similarly, the positive record-only condition was instructed to record only entries rated 5 (*mildly positive*) to 9 (*extremely positive*). Rating options 4 and lower were viewable but not selectable.

Regardless of whether they were allocated to the negative versus positive condition, recorders were instructed to make two or more recordings per day, and all reflection capabilities were disabled.

The *negative record+reflect* condition was instructed to make recordings only between 1 (*extremely negative*) and 4 (*mildly negative*) on the Echo Emotion scale. Although emotion ratings 5 and higher were still visible in the interface, these options were not selectable. However unlike the negative record-only group, participants in the negative record+reflection condition were re-presented with one of their negative recordings later in the intervention and asked to reflect on how they now felt about that recording. Reflection involved generating an emotion rating evaluating their current feelings about the experience along with a new textual reappraisal (see Figure 9, right-hand panel). During reflection, unlike recording, participants could use the entire emotional scale.

Similarly, the *positive record+reflect* condition was instructed to record only entries rated 5 (*mildly positive*) to 9 (*extremely positive*). Rating options 4 and lower were viewable but disabled. Again, positive record+reflectors were re-presented with these positive recordings later and were asked to write a new text entry to reflect on how they now felt (Figure 9, right-hand panel). Participants could use the entire emotion scale during reflection.

We implemented a selection algorithm so that reflections would be evenly spaced throughout the whole study and reflection intervals balanced so that overall reflections weren't disproportionately recent or old. These decisions were informed by prior deployments (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016; Konrad, Tucker, et al., 2016). The Echo interface displayed four boxes for reflection capabilities (see Figure 9, left-hand panel). Daily recordings were stored in Box 1 (upper left, left-hand panel) and disappeared after the day was over. Box 2 (upper right, left-hand panel) is for memories that are 1 week old, Box 3 (lower left, left-hand panel) is for 2 weeks ago, and Box 4 (lower right, left-hand panel) is for 3 weeks ago. Participants were unable to review past reflections freely and could reflect only when the system selected reflections for them. Reflections began after 1 week of logging to give participants sufficient time to make recordings. When reflections were available, participants would receive a smartphone notification. Once viewed, a reflection was removed from the pool and never again sent back. Reflections included pictures if these were present in the original recording.

When recording events, we instructed participants that "an event can be anything ranging from a social gathering, conversation, or lecture to just watching TV, getting good or bad news, having coffee with a friend etc." Actual recorded events ranged from highly positive (e.g., beginning a dream job) to very negative (e.g., separation from a long-term partner). To improve accuracy of recordings, participants were asked to record the event while they were experiencing it, or as close to the event as was practically possible. Of course this was impractical in certain circumstances, for example, in certain social settings or when driving.

After 28 days of using Echo, participants completed the posttest surveys and were issued passcodes that allowed them to see all posts they had submitted, with the option to delete any entries. After reviewing prior entries, participants e-mailed their Echo logfiles to the researchers. The content of private posts was excluded from the shared logfiles, but their emotion ratings were still viewable, as participants had been informed in the consent form.

4.2. Results

Positive Recording and Reflection Increase Long-Term Well-Being

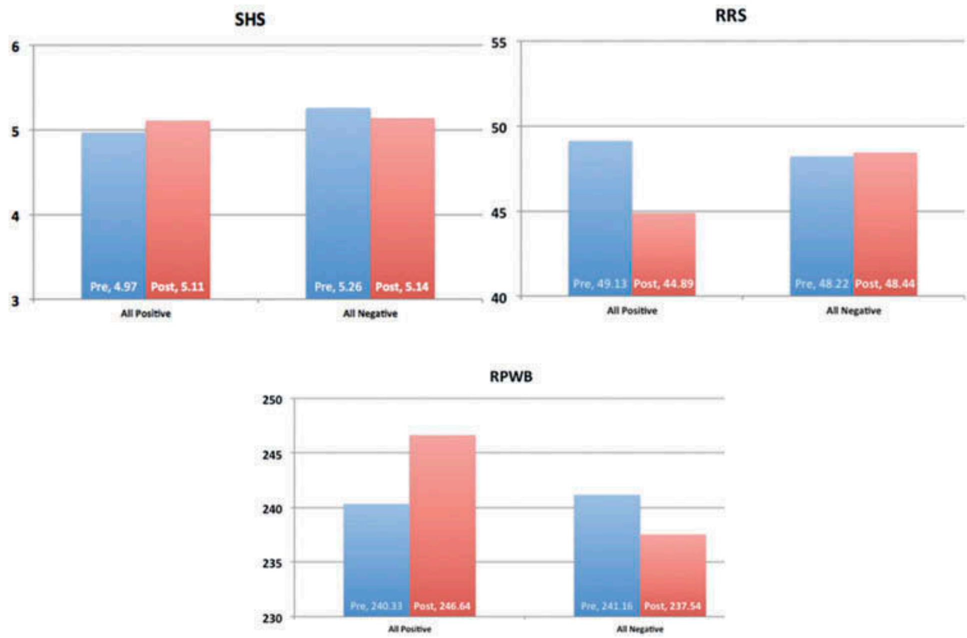
We first explored well-being differences between the four design conditions on the pre- and posttest surveys. Survey data were analyzed using a mixed-design MANOVA with two between-subjects factors: system (record-only vs. record+reflect) and emotional valence (positive vs. negative). There was also one within-subjects factor Time (Time 1: Pretest vs. Time 2: Posttest). The dependent variables were the three well-being scales (SHS, RPWB, RRS).

Overall there was an interaction between time and valence (Pillai's trace, V = 0.096), $F(3, 99) = 3.689$, $p = .018$, partial $\eta^2 = 0.096$), with participants in positive conditions improved well-being over time. Univariate analyses of the interaction revealed a significant difference between the three scales (see Figure 12). Confirming our expectations, positive valence conditions went up significantly in RPWB, $F(1, 101) = 8.37$, $p = .005$, partial $\eta^2 = .08$, and decreased in rumination scores, $F(1, 101) = 4.08$, $p = .046$, partial $\eta^2 = .04$, compared to negative conditions, though we saw no significant change in SHS, $F(1, 101) = 2.57$, $p = .112$, partial $\eta^2 = 0.025$). There were no interactions between record-only and record+reflect conditions and time (Pillai's trace, V = .019), $F(3, 99) = 0.642$, $p = .589$, partial $\eta^2 = 0.019$, indicating that record+reflect had no additional effects over record-only. There were also no significant higher order interactions.

Negative Recording and Reflection Induces Greater Use of Analytic Language

To better understand the underlying reasons for changes in well-being we again analyzed the content of recordings and reflections from participants' logfiles. As with

FIGURE 12. Mean scores on the Subjective Happiness Scale (SHS), Ruminative Responses Scale (RRS), and the Ryff's Scales of Psychological Well-Being (RPWB) for the combined positive conditions (*n* = 55) and negative conditions (*n* = 50).



***Note*. Positive recording and reflection increase perceived well-being and decrease rumination.**

Study 1, we used LIWC (Pennebaker et al., 2007). Following prior work (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016; Konrad, Tucker, et al., 2016; Peesapati et al., 2010), we again targeted word categories that provided evidence of changes of emotion, understanding, and insight. As there were no differences between both record-only conditions and record+reflect conditions combined, we combined these in our analyses of post content. We analyzed differences using *t* tests to compare language use across positive and negative conditions.

We first examined differences in overall use of emotional language independently of whether this referred to positive or negative emotions. As we note next, our manipulation check revealed that participants in positive conditions unsurprisingly generated more positive emotion terms than those in negative conditions who expressed more negative emotions. However, we also found that the combined positive conditions (i.e., including both record-only and record+reflect) had a greater use of affective language (M = 8.83, SD = 2.2) than combined negative conditions (M = 6.78, SD = 2.01), $t(103)$ = 4.898, p < .0005.

We also found that, consistent with previous findings (Pennebaker & Chung, 2011), negative conditions overall wrote more (M = 1,515.78, SD = 1,301) than positive conditions (M = 982.1, SD = 654.7), $t(103)$ = –2.691, p = .008. In our subsequent analyses we therefore controlled for word count differences between positive and negative conditions. With these controls, negative conditions showed a greater use of cognitive mechanism terms (*M* = 16.86, *SD* = 2.75, $F(1, 102) = 13.165$, $p < .0005$, and causal terms (*M* = 1.68, *SD* = .59), $F(1, 102) = 9.462$, $p = .003$, although there were no differences in insight terms (*M* = 2.16, *SD* = 0.76), $F(1, 102) = 2.820$, $p = .096$. Negative conditions were also more self-referential as indicated by an increased use in "I" (*M* = 9.82, *SD* = 2.5) compared to positive conditions (*M* = 8.02, *SD* = 3.04), $t(103) = -3.287$, $p = .008$. Overall these results are consistent with prior work on expressive writing (Pennebaker & Chung, 2007), suggesting that negative experiences elicit active cognitive reappraisal; however, in contrast to that prior work, reappraisal did not elicit well-being benefits. We return to this point in our conclusions.

Recording and Reflection of Extremely Negative Events Reduces Well-Being

We next went on to explore the effects of emotional extremity. There may be important differences between reflecting on extremely negative and more mildly negative events (Pennebaker & Chung, 2011), with reflection on extremely negative experiences not improving well-being. We therefore examined effects of extremity. Recall that the emotion scale ranged from 1 to 9. For each participant we computed the average emotion rating. We defined as extreme those participants whose average emotion ratings were 3 or less (extreme negative), and 7 or above (extreme positive). Those with ratings of 4 or 5 in the negative condition were defined as mild negative and with ratings of 5 or 6 in the positive condition were defined as mild positive.

Survey data were again analyzed using a mixed-design MANOVA with two between-subjects factors: extremity (extreme vs. mild) and emotional valence (positive vs. negative). There was also one within-subjects factor, time (Time 1: pretest vs. Time 2: posttest). The dependent variables were the three well-being scales (SHS, RPWB, RRS). As before, we found an overall effect of valence with positive conditions improving more than negative ones (Pillai's trace, V = 0.126), $F(3, 99) = 4.755$, $p = .004$. We also found a significant interaction between time and extremity (Pillai's trace, V = .126), $F(3, 99) = 3.90$, $p = .011$, with mild posters showing more improvement overall than extreme posters. Furthermore there was a three-way interaction between time, valence, and extremity (Pillai's trace, V = .131), $F(3, 99) = 4.99$, $p = .003$. Univariate *F* tests showed a significant difference between groups for SHS, $F(1, 101) = 12.623$, $p = .001$, and a trend for RRS, $F(1, 101) = 2.943$, $p = .089$. There was no significant difference for RPWB, $F(1, 101) = 1.12$, $p = .292$. Analysis of the interaction showed that SHS scores improved for both extreme and mild positive groups, as well as mild negative. However SHS scores decreased for extreme negative groups.

An alternative explanation for these results is that extreme negative participants have a different emotional disposition from others. We therefore profiled this group, comparing them with remaining participants. They were no different for any pretest measure. There were no differences in SHS pre, $t(103) = -1.526$, $p = .130$; RPWB pre, $t(103) = -.292$, $p = .771$; or RRS pre, $t(103) = -.074$, $p = .941$. There were also no differences in gender, age, or highest level of education completed (*p*s = .481–.973). It therefore seems that our results cannot be explained by prior participant profiles and instead were due to what transpired during the intervention itself.

Exploring this effect further, a LIWC analysis comparing the post contents of extreme negative participants versus other participants showed a significant difference in cognitive mechanism, $t(101) = -2.358$, $p = .020$, and insight terms, $t(101) = -2.006$, $p = .048$, with extreme negative participants providing higher rates of each (insight = 0.6 greater, cognitive mechanism = 2.21 greater). This is an unexpected finding that we return to in the Discussion and Conclusions section.

Preintervention Differences, Compliance, and Manipulation Checks

Finally we conducted a series of compliance and manipulation checks. We assessed whether there were pretest differences between the four groups on the three initial surveys. There were no differences across four conditions in pretest RPWB ($p = .997$), SHS ($p = .476$), RRS ($p = .958$), or age ($p = .588$). To test compliance we calculated overall word count, word count per post, and number of recording+reflections. There was no difference in number of recordings or number of reflections across conditions (*p*s = .358–.732). LIWC was used to compare the frequency of negative emotion terms (e.g., sad, unhappy) against the frequency of positive emotion terms (e.g., happy, bliss) across the combined negative ($n = 50$) and positive ($n = 55$) conditions. Consistent with our manipulation, positive conditions had a greater use of positive emotion terms ($M = 7.79$, $SD = 2.24$) than negative conditions ($M = 2.8$, $SD = 1.2$), $t(103) = 13.928$, $p < .0005$. Negative conditions had a greater use of negative emotion terms ($M = 3.89$, $SD = 1.43$) than positive conditions ($M = 1.00$, $SD = 0.50$), $t(103) = -13.985$, $p < .0005$.

4.3. Summary

Overall, we found that recording positive experiences boosts emotional well-being, whereas posting about negative experiences reduces it. We did not find differences between record-only and record+reflect, confirming (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016). This suggests that reanalyzing prior experiences does not induce additional well-being benefits over recording-only. Post intensity was also important, however. Participants who habitually posted extremely negative posts did not experience well-being benefits. Although mildly negative posters improved in SHS scores, extremely negative posters decreased in SHS scores. This observation confirms prior psychological work showing that extreme ruminators do not improve in expressive writing therapy (Pennebaker & Chung, 2011).

Analysis of the underlying content of posts indicates that event valence had strong effects on participants' thoughts while recording and reflecting. Those who recorded and reflected on extremely negative experiences wrote more; in addition they showed more self-focus, which is consistent with prior work (Campbell & Pennebaker, 2003). As expected from prior work, negative experiences also led participants to be more analytic, engaging in more causal analysis and analysis of cognitive mechanisms (Klein & Boals, 2001; Petrie et al., 1998); however, to our surprise this did not induce changes in well-being.

5. DISCUSSION AND CONCLUSIONS

We begin by discussing the results of the two studies and conclude with more general lessons for designing systems that support emotional well-being. We describe two field studies: one with EmotiCal, a system for goal-driven mood tracking, and another with Echo, for reflecting on life events. Both goal-driven tracking and documenting life events are common motivations for adopting self-tracking technologies (Rooksby et al., 2014).

5.1. Emotion Forecasting

Our research explored two design issues concerning actionable insights and choice of events to reflect on in the context of the examined emotional life. Consistent with recent research systems (Bardram et al., 2013; Rabbi, Pfammatter, Zhang, Spring, & Choudhury, 2015), our forecasting results are promising, addressing a critical challenge with personal informatics systems: how analytics on past data can inform and motivate future actions. EmotiCal introduces a successful new technique that supports emotion forecasting to promote well-being. It uses past mood data to forecast and visualize future user moods, encouraging users to adopt new behaviors to improve their future happiness. To increase user compliance, these new behaviors are chosen to be actionable and personalized to the user. Our intervention results showed that forecasting improved daily mood, ratings of self-awareness, and reported activity effectiveness to improve mood when compared with two control groups whose members simply monitored their mood or provided only pre–post surveys. A majority of users were also highly positive about emotion-forecasting system features.

Source of Activity Recommendations

An important determinant of improved well-being was the nature of the recommended activities. Recall that recommended activities could be *history* based, that is, drawn from logged activities that participants actively tracked, or *profile* based, that is, generated from a set of activities, derived from pretest survey responses to a basic psychological needs survey. Participants reported higher ratings of activity engagement and showed greater benefits for *profile*- rather than *history*-based recommendations. Why is this? It may be that history-based recommendations are not insightful; participants actively track frequent familiar activities (e.g., concerning health, social, and work life), in the monitoring part of the intervention leading them to become aware of exactly how these activities affect mood. They may already be deploying these activities to strategically influence mood. Participants may not therefore need to have these *history* actions recommended to them during emotional forecasting. Interviews and open-ended responses in the final survey suggest that monitoring-only participants similarly became aware of how regular daily activities impact their moods. In contrast, profile-based methods to recommend relevant activities drawn from outside this familiar pool may be nonobvious. They may also be unusual for typical lifestyles, more motivating and consequentially more influential on mood.

Perceived Accuracy and Compliance

Although our results are positive, they give rise to several important challenges around the design of future forecasting systems. The first issue concerns model accuracy. Objective model accuracy was good (explaining 50% of mood variance), but *perceived accuracy* was critical for both motivation and user engagement. Overall, users felt that the models were accurate, with an average accuracy rating of 4.95 (SD = 0.89) on a 7-point scale. As we expected, participants who rated models as accurate were more likely to engage in planning and adopt new activities. In contrast, participants who did not believe model predictions were less likely to engage with the system, indicated by a lower likelihood to plan new activities or adopt system activity recommendations. These issues of perceived accuracy relate to compliance, a general challenge for behavior change systems. Future work should examine other reasons for skepticism including when this arises from a conflict with the participant's self-image.

How then might we further improve perceived model accuracy, compliance, and quality of recommendations? At a basic level, we could improve forecasting model accuracy and consequent perceptions by simply having more data. Current mood models were derived from 2 to 3 weeks' data, and longer term deployments with more participant data would clearly improve this. Another challenging opportunity to improve models would be to increase the set of factors that are included in the explanatory model. Like many hedonic well-being models (Kahneman, 2000; Kahneman, Diener & Schwarz, 1999), our current approach is limited in the simple activity triggers it relies on to predict mood, including health, work, and social activities. Of course, these do not exhaust the many possible contributors to mood. We have therefore begun modeling work to extend these factors, exploring the role of long-term goals and identity factors (Deci & Ryan, 2000) in explaining mood. Expanding our models to include such factors could also make a significant contribution to social science theorizing about emotions.

Furthermore, there may be individual differences; not all users may be affected in the same way by our event triggers. So, mood models might also be improved by clustering data to identify patterns across subgroups of users. We have already begun experiments to identify different emotional styles, finding that work activities have varied impacts on different user's mood. For one subset of users, work has positive effects on mood, for others it has negative effects, and for a final subset it has little emotional effect. Clustering data in this way should improve mood forecasting. However, clustering could also assist with activity recommendations, again allowing us to identify novel but relevant activities. By clustering it might be possible to recommend to users activities that others with a similar emotional and activity profile have found to be effective. This profiling approach follows techniques used successfully in recommender systems (Herlocker, Konstan, Terveen, & Riedl, 2004; Rashid et al., 2002; Sarwar et al., 2000). Finally, longer term deployments might also allow more systematic use of feedback concerning the relations between planned activities and actual mood shifts. Determining that particular planned activities have strong predictable effects on mood might lead our system to more aggressively recommend these.

Aside from these issues concerning modeling and recommendation, we might also explore different forecasting UI designs. Given the difficulties of designing effective interfaces to support end-user analysis of personal data (Bentley et al., 2013; Epstein et al., 2014; McDuff et al., 2012), we were careful in the current study to base our forecasting and planning visualization around direct user feedback to initial designs and a prior deployment. Our participants were clear that their overall requirement was for a simple, easily comprehensible visualization, and exit interviews and surveys confirmed this. However, it may be that other, more complex time series visualizations, allowing users to explore longer term patterns or the emotional effects of specific activities, might also be effective, and future work might explore this. Such designs are present in other research systems (Bentley et al., 2013; Epstein et al., 2014; McDuff et al., 2012) and in some commercial products such as In Flow (In Flow, 2015) and Moodscope (Moodscope, 2015). However, these more complex alternatives need to be considered in the context of our application. Our goal is to support rapid impromptu activity planning rather than systematic scrutinizing of complex past personal datasets.

Future EmotiCal Research

An obvious systems improvement is to support record keeping of whether plans were actually executed, for example, using a simple probe ("Did you complete activity X?") or by sensor-based activity logging. In addition, our activity recommendations were brief statements (e.g., "Go for a run", "Interact with a friend"). Activity compliance may be further improved by providing a breakdown of activity recommendations into smaller steps (e.g., "Find your running shoes, put on workout clothes, fill your water bottle …"; Gollwitzer, 1999). It is also important in future work that we better dissect the impact of different aspects of the intervention by comparing designs with only mood forecasts versus only activity planning features.

In addition, we can gather valuable psychological insights into cases of when and why participant beliefs about their future moods diverge from algorithm predictions and why these discrepancies occur. Similarly, participants may be unaware of the effect of specific activities on mood. For example, past research has shown that semantic beliefs about what influences one's emotional state can contrast sharply with data collected from experience sampling measures (Robinson & Clore, 2002). Future work with EmotiCal can support greater user insight by addressing these discrepancies between *beliefs* about mood improving or impairing behaviors versus what is suggested in daily records.

In addition, our current system design was focused on increasing positive activities, yet we also gathered data regarding negative influences on behavior (e.g., activities or specific people who depress mood). It will be useful to understand how users react to this data, how to present this information tactfully to improve well-being and special considerations for these types of analytics. We now turn to the topic of presenting negative data.

5.2. Echo for Technology-Mediated Reflection

Our second intervention examined systems that support active reflection on prior experiences, exploring the effects of recommending positive versus negative experiences. Present commercial reflective systems do not currently focus on event valence. Our intervention was therefore motivated by concerns that systems that inadvertently encourage users to record and reflect on prior highly negative experiences may negatively affect well-being. We showed overall benefits for recording and reflecting on the positive but that recording and reflecting exclusively on intensely negative past experiences detracts from well-being. This confirms other unmediated work exploring expressive writing about very negative experiences (Gidron et al., 1996). We also replicated prior work showing that reflection adds little well-being benefit over recording alone (Isaacs et al., 2013; Konrad, Isaacs, et al., 2016). These findings have significant implications for designing systems to promote emotional reflection and well-being. In particular, they reinforce the importance of developing new systems similar to LiveHappy (Parks et al., 2012) that encourage users to record and reflect on positive experiences.

Recording Negative Emotional Events

The design lessons are far more complex for negative emotions, however. There has been some discussion of systems that accidentally subject users to highly negative past experiences (Haimson et al., 2015; Sas & Whittaker, 2013). Our results inform this debate by showing that people who continually record and reflect on highly negative experiences may suffer compromised well-being. How might negativity be addressed? One technical option is that systems could monitor posts, identifying patterns of extreme negativity, similar to suicide watch services operating on Facebook and Reddit. Another approach is to modify the overall design to increase user control over *when* and *if* they revisit negative events. With Echo, users were automatically presented with content to reflect on. Increasing user agency is important to avoid unintended consequences of negative reflection. Allowing users to strategically delete, edit, or hide content for deferred review could prevent the decreases to well-being we observed. There are also interesting analogies with other types of systems that help users control their impulses that have proposed similar design solutions (Hollis et al., 2015; Sas & Whittaker, 2013). Another possibility is that the reflections the system recommends are mood dependent, with negative experiences only being re-presented when users' mood is positive. Other work has explored such mood-dependent systems that suggest reflections adaptively based on current mood (Konrad, Tucker, et al., 2016). Reflecting on a negative post when in a positive mood leads the content of that post to be evaluated more positively, possibly promoting a redemption sequence (Konrad, Isaacs, et al., 2016; Pennebaker, 2004, 2016; Wildschut et al., 2006). However, the Konrad, Tucker, et al. (2016) results suggest that negative reflection may need to be used sparingly. Processing negative experiences remains a challenge, and more interventions and new designs are needed.

Future Work on the Valence of Mood Records

Future work exploring reflection valence should also involve control conditions to compare these results against a no-logging baseline. In addition, an important next step is to look at what happens when recording behavior is equally positive, negative, or neutral, and only the information that users reflect upon is biased. Also, although like Isaacs et al. (2013) we found no benefits of expressive writing for negative events, this result may be inherent to how technology-based monitoring typically operates in comparison to Pennebaker's expressive writing therapies. There are three main differences: (a) users write less using Echo than they would in a offline expressive writing therapy; (b) with Echo, prior logs are reviewable, whereas in expressive writing no records are kept; and (c) with expressive writing, participants choose what events to reflect on and typically focus on those events over multiple sessions, whereas Echo reflection events are chosen for them and participants reflect on a wider variety of events. These are important considerations to take into account if designers intend to create effective emotion-disclosure systems.

Comparing the two studies, one important lesson we can draw is that emotional valence is critical and that positive and negative emotions are very different. Recording and reflecting on positive experiences and carrying out enjoyable activities promoted well-being. In contrast, recording and reflecting on negative experiences, specifically intensely negative experiences, detracted from well-being. There is also an intriguing contrast between the two studies. In the forecasting study we saw through logfile analysis that underlying processes of insight, cognition, and appraisal promoted positive emotions and well-being. However, in the reflection study, for those who recorded intensely negative experiences, these same processes were correlated with reductions in well-being. Why was this? One possibility is that our Echo intervention was too short and that appraisal benefits would have emerged even for highly negative events if we had continued that study for a longer period. Other work on expressive writing suggests such writing has short-term negative impacts before long-term benefits emerge (Sloan & Marx, 2004). In terms of system design, we need to better understand this time course, possibly deferring re-presentation of recent negative events to allow participants some distance from negative events before reappraising these.

Our results also contribute to our emerging scientific understanding of online behavior. People are increasingly spending huge parts of their lives using digital technologies, and it is important that we understand how this affects emotions and well-being. For example, the emotional content Facebook posts affects other people's online conduct, as well as the poster's social networks, relationships, and well-being (Burke & Develin, 2016; Gonzales & Hancock, 2011; Kim & Lee, 2011; Kramer et al., 2014). Although many of these prior studies have been correlational in nature, our work adds to this literature using intervention methods, in showing the effects of different forecasting, recording, and reflection behaviors on well-being, as well as the underlying mechanisms that give rise to these effects.

Finally, we want to return to some of the original motivations for developing well-being systems. We began by reviewing social science research showing that people have difficulty in tracking, controlling, and understanding their emotions, and in some cases they have problems in processing past events. These difficulties have important negative consequences for well-being. We have presented two interventions and systems that potentially address these problems. By externalizing, tracking, and reflecting on emotions in this way, participants can gain greater understanding and control over their emotions, results that hold great promise for future human–computer interaction design research and well-being interventions.

NOTES

HCI Editorial Record. First received 22 April 2016. Revisions received 30 October 2016. Accepted by 27 December 2016. Final manuscript received 27 December 2016. — *Editor*

REFERENCES

Alonso, J., & Lépine, J. P. (2007). Overview of key data from the European Study of the Epidemiology of Mental Disorders (ESEMeD). *The Journal of Clinical Psychiatry*, *68*(Suppl. 2), 3–9.

Bannon, L. (2006). Forgetting as a feature, not a bug: The duality of memory and implications for ubiquitous computing. *CoDesign*, *2*(1), 3–15. doi:10.1080/15710880600608230

Bauer, J. S., Consolvo, S., Greenstein, B., Schooler, J., Wu, E., Watson, N. F., & Kientz, J. (2012, May). ShutEye: Encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. *Proceedings of the CHI 2012 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Baumeister, R. F., Vohs, K. D., Nathan Dewall, C., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review*, *11*, 167–203. doi:10.1177/1088868307301033

Bell, G., & Gemmell, J. (2009). *Total recall: How the E-memory revolution will change everything*. New York, NY: Penguin.

Bentley, F., Tollmar, K., Stephenson, P., Levy, L., Jones, B., Robertson, S., … Wilson, J. (2013). Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction*, *20*(5), 1–27. doi:10.1145/2503823

Bryant, F. B., Smart, C. M., & King, S. P. (2005). Using the past to enhance the present: Boosting happiness through positive reminiscence. *Journal of Happiness Studies*, *6*, 227–260. doi:10.1007/s10902-005-3889-4

Burke, M., & Develin, M. (2016). Once more with feeling: Supportive responses to social sharing on Facebook. *Proceedings of the CSCW 2016 Conference on Computer-Supported Cooperative Work & Social Computing*. New York, NY: ACM.

Campbell, R., & Pennebaker, J. (2003). The secret life of pronouns: Flexibility in writing style and physical health. *Psychological Science*, *14*, 60–65. doi:10.1111/1467-9280.01419

Choe, E. K., Lee, N. B., Lee, B., Pratt, W., & Kientz, J. A. (2014, April). Understanding quantified-selfers' practices in collecting and exploring personal data. *Proceedings of the CHI 2014 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Cosley, D., Akey, K., Alson, B., Baxter, J., Broomfield, M., Lee, S., & Sarabu, C. (2009, September). Using technologies to support reminiscence. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology* (pp. 480–484). British Computer Society.

Cosley, D., Schwanda Sosik, V., Schultz, J., Peesapati, S., & Lee, S. (2012). Experiences with designing tools for everyday reminiscing. *Human–Computer Interaction*, *27*, 175–198.

Cuijpers, P., Van Straten, A., & Warmerdam, L. (2007). Behavioral activation treatments of depression: A meta-analysis. *Clinical Psychology Review*, *27*, 318–326. doi:10.1016/j.cpr.2006.11.001

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*, 227–268. doi:10.1207/S15327965PLI1104_01

Deci, E. L., & Ryan, R. M. (2008). Hedonia, eudaimonia, and well-being: An introduction. *Journal of Happiness Studies*, *9*, 1–11. doi:10.1007/s10902-006-9018-1

Depp, C. A., Ceglowski, J., Wang, V. C., Yaghouti, F., Mausbach, B. T., Thompson, W. K., & Granholm, E. L. (2015). Augmenting psychoeducation with a mobile intervention for bipolar disorder: A randomized controlled trial. *Journal of Affective Disorders*, *174*, 23–30. doi:10.1016/j.jad.2014.10.053

Diener, E. (1984). Subjective well-being. *Psychol. Bull.* *95*, 542–575.

Dobson, K. S., & Joffe, R. (1986). The role of activity level and cognition in depressed mood in a university sample. *Journal of Clinical Psychology*, *42*, 264–271. doi:10.1002/(ISSN)1097-4679

Doryab, A., Frost, M., Faurholt-Jepsen, M., Kessing, L. V., & Bardram, J. E. (2015). Impact factor analysis: Combining prediction with parameter ranking to reveal the impact of behavior on health outcome. *Personal and Ubiquitous Computing*, *19*, 355–365. doi:10.1007/s00779-014-0826-8

Durkin, L. A. (2006). Relationship between quality of mood monitoring and treatment outcomes in clients with bipolar disorder. *Dissertation Abstracts International: Section B: the Sciences and Engineering*, *67*, 3447.

Ekers, D., Richards, D., & Gilbody, S. (2008). A meta-analysis of randomized trials of behavioural treatment of depression. *Psychological Medicine*, *38*, 611–623. doi:10.1017/S0033291707001614

Epstein, D., Cordeiro, F., Bales, E., Fogarty, J., & Munson, S. (2014). Taming data complexity in lifelogs. *Proceedings of the DIS 2014 Conference on Designing Interactive Systems*. New York, NY: ACM.

Facebook Inc. (2015a). *On this day*. Retrieved from https://www.facebook.com/help/439014052921484/

Facebook Inc. (2015c). *Year in review*. Retrieved from https://www.facebook.com/help/1551882718390433/

Faurholt-Jepsen, M., Frost, M., Ritz, C., Christensen, E. M., Jacoby, A. S., Mikkelsen, R. L., … Kessing, L. V. (2015). Daily electronic self-monitoring in bipolar disorder using smartphones—The MONARCA I trial: A randomized, placebo-controlled, single-blind, parallel group trial. *Psychological Medicine*, *45*, 2691–2704. doi:10.1017/S0033291715000410

Faurholt-Jepsen, M., Munkholm, K., Frost, M., Bardram, J. E., & Kessing, L. V. (2016). Electronic self-monitoring of mood using IT platforms in adult patients with bipolar disorder: A systematic review of the validity and evidence. *BMC Psychiatry*, *16*(1), 1. doi:10.1186/s12888-016-0713-0

Frijda, N. H. (1988). The laws of emotion. *American Psychologist*, *43*, 349–358. doi:10.1037/0003-066X.43.5.349

Frisina, P. G., Borod, J. C., & Lepore, S. J. (2004). A meta-analysis of the effects of written emotional disclosure on the health outcomes of clinical populations. *The Journal of Nervous and Mental Disease*, *192*, 629–634. doi:10.1097/01.nmd.0000138317.30764.63

Gidron, Y., Peri, T., Connolly, J. F., & Shalev, A. Y. (1996). Written disclosure in posttraumatic stress disorder: Is it beneficial for the patient? *The Journal of Nervous and Mental Disease*, *184*, 505–506. doi:10.1097/00005053-199608000-00009

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (2002). Durability bias in affective forecasting.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*, 493–503. doi:10.1037/0003-066X.54.7.493

Gonzales, A., & Hancock, J. (2011). Mirror, mirror on my Facebook wall: Effects of exposure to Facebook on self-esteem. *Cyberpsychology, Behavior, and Social Networking*, *14*, 79–83. doi:10.1089/cyber.2009.0411

Google Inc. (2015). *Rediscover this day*. Retrieved from https://support.google.com/photos/answer/6128811?hl=en

Haimson, O. L., Brubaker, J. R., Dombrowski, L., & Hayes, G. R. (2015, February). Disclosure, stress, and support during gender transition on Facebook. *Proceedings of the CSCW 2015 Conference on Computer Supported Cooperative Work & Social Computing*. New York, NY: ACM.

Harris, A. H. S. (2006). Does expressive writing reduce health care utilization? A metaanalysis of randomized trials. *Journal of Consulting and Clinical Psychology*, *74*, 243–252. doi:10.1037/0022-006X.74.2.243

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, *22*, 5–53. doi:10.1145/963770

Hollis, V., Konrad, A., & Whittaker, S. (2015). Change of heart: Emotion tracking to promote behavior change. *Proceedings of the CHI 2015 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Hyndman, R. J., & Khandakar, Y. (2007). *Automatic time series for forecasting* (No. 6/07). Melbourne, Australia: Department of Econometrics and Statistics, Monash University.

In Flow. (2015). *In flow*. Retrieved from http://www.inflow.mobi/

Isaacs, E., Konrad, A., Walendowski, A., Lennig, T., Hollis, V., & Whittaker, S. (2013, April). Echoes from the past: How technology mediated reflection improves well-being. *Proceedings of the CHI 2013 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Isen, A. M. (2004, April). Positive affect facilitates thinking and problem solving. In *Feelings and emotions: The Amsterdam symposium* (pp. 263–281). Cambridge, UK: Cambridge University Press.

Kahneman, D. (1999). Objective happiness. In E. Diener, N. Schwarz, and 0. Kahneman (Eds.), *Well-being: the foundations of hedonic psychology* (pp. 3–27). New York, NY: Russell Sage Foundation.

Kahneman, D. (2000). Evaluation by moments: past and future. In D. Kahneman and A. Tversky (Eds.), Choices, values and frames (pp. 293-308). New York, NY: Cambridge University Press and the Russell Sage Foundation.

Kahneman, D., Diener, E., & Schwarz, N. (1999). Well-being: Foundations of hedonic psychology. New York, NY: Russell Sage Foundation.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (2002). Durability bias in affective forecasting.

Kalnikaite, V., Sellen, A., Whittaker, S., & Kirk, D. (2010). Now let me see where I was: Understanding how lifelogs mediate memory. *Proceedings of the CHI 2010 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Khovanskaya, V., Baumer, E. P., Cosley, D., Voida, S., & Gay, G. (2013). "Everybody knows what you're doing": A critical design approach to personal informatics. *Proceedings of the CHI 2013 Conference on Human Factors in Computer Systems*. Paris, France: ACM.

Kim, J., & Lee, J. (2011). The Facebook paths to happiness: Effects of the number of Facebook friends and self-presentation on subjective well-being. *Cyberpsychology, Behavior, and Social Networking*, *14*, 359–364. doi:10.1089/cyber.2010.0374

Klein, K., & Boals, A. (2001). Expressive writing can increase working memory capacity. *Journal of Experimental Psychology: General*, *130*, 520–533. doi:10.1037/0096-3445.130.3.520

Konrad, A., Bellotti, V., Crenshaw, N., Tucker, S., Nelson, L., Du, H., … Whittaker, S. (2015). Finding the adaptive sweet spot: Balancing compliance and achievement in automated stress reduction. *Proceedings of the CHI 2015 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Konrad, A., Isaacs, E., & Whittaker, S. (2016). Technology Mediated Memory: Is technology destroying our memory and interfering with well-being? *Transactions on Computer Human Interaction*, *23*(4) , 1073–1516. doi:10.1145/2934667

Konrad, A., Tucker, S., Crane, J., & Whittaker, S. (2016). Technology and reflection: Mood and memory mechanisms for well-being. *Psychology of Well-Being*, *6*, 1–24. doi:10.1186/s13612-016-0045-3

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*, 8788–8790. doi:10.1073/pnas.1320040111

Kruijshaar, M. E., Barendregt, J., Vos, T., De Graaf, R., Spijker, J., & Andrews, G. (2005). Lifetime prevalence estimates of major depression: An indirect estimation method and a quantification of recall bias. *European Journal of Epidemiology*, *20*(1), 103–111. doi:10.1007/s10654-004-1009-0

Lewinsohn, P. M., & Amenson, C. S. (1978). Some relations between pleasant and unpleasant mood-related events and depression. *Journal of Abnormal Psychology*, *87*, 644–654.

Lewinsohn, P. M., & Libet, J. (1972). Pleasant events, activity schedules, and depressions. *Journal of Abnormal Psychology*, *79*, 291–295. doi:10.1037/h0033207

Li, I., Dey, A., & Forlizzi, J. (2011). Understanding my data, myself. Supporting self reflection with Ubicomp technologies. *Proceedings of the UbiComp 2011 Conference on Ubiquitous Computing*. New York, NY: ACM.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*, 705–717. doi:10.1037/0003-066X.57.9.705

Lyubomirsky, S., Kasri, F., Chang, O., & Chung, I. (2006). Ruminative response styles and delay of seeking diagnosis for breast cancer symptoms. *Journal of Social and Clinical Psychology*, *25*, 276–304. doi:10.1521/jscp.2006.25.3.276

Lyubomirsky, S., & Layous, K. (2013). How do simple positive activities increase well-being? *Current Directions in Psychological Science*, *22*, 57–62. doi:10.1177/0963721412469809

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, *46*, 137–155. doi:10.1023/A:1006824100041

MacPhillamy, D. J., & Lewinsohn, P. M. (1982). The pleasant events schedule. *Journal of Consulting and Clinical Psychology*, *50*, 363–380. doi:10.1037/0022-006X.50.3.363

Mayer-Schönberger, V.. (2009). *Delete: The virtue of forgetting in the digital age*. Princeton, NJ: Princeton Press.

McDuff, D., Karlson, A., Kapoor, A., Roseway, A., & Czerwinski, M. (2012). AffectAura: An intelligent system for emotional memory. *Proceedings of the CHI 2012 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

Meads, C. (2003, October). *How effective are emotional disclosure interventions? A systematic review with meta-analyses*. Paper presented at the 3rd International Conference on The (Non) Expression of Emotions in Health and Disease, Tilburg, the Netherlands.

Meyer, E. (2014). *Inadvertent algorithmic cruelty*. Retrieved from http://meyerweb.com/eric/thoughts/2014/12/24/inadvertent-algorithmic-cruelty/

Michie, S., van Stralen, M. M., & West, R. (2011).The behaviour change wheel: A new method for characterising and designing behaviour change interventions. *Implementation Science 6*, 42.

Moodscope. (2015). *Moodscope*. Retrieved from http://www.moodscope.com

Mulligan, B. (2014). *MorningPics*. Retrieved from http://www.morningpics.com/

Munson, S. A., Lauterbach, D., Newman, M. W., & Resnick, P. (2010, June). Happier together: Integrating a wellness application into a social network site. *Proceedings of the PERSUASIVE 2010 Conference on Persuasive Technology*. Berlin, Germany: Springer

Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depressive episodes. *Journal of Abnormal Psychology*, *100*(*4*), 569–582.
Parks, A., Della Porta, M., Pierce, R. S., Zilca, R., & Lyubomirsky, S. (2012). Pursuing happiness in everyday life: The characteristics and behaviors of online happiness seekers. *Emotion*, *12*, 1222–1234. doi:10.1037/a0028587
Peesapati, S. T., Schwanda, V., Schultz, J., Lepage, M., Jeong, S., & Cosley, D. (2010). Pensieve: Supporting everyday reminiscence. *Proceedings of the CHI 2010 Conference on Human Factors in Computer Systems*. New York, NY: ACM.
Pennebaker, J. W. (2004). Theories, therapies, and taxpayers: On the complexities of the expressive writing paradigm. *Clinical Psychology: Science and Practice*, *11*(*2*), 138–142.
Pennebaker, J. W., & Beall, S. (1986). Confronting a traumatic event: Toward an understanding of inhibition and disease. *Journal of Abnormal Psychology*, *95*, 274–281. doi:10.1037/0021-843X.95.3.274
Pennebaker, J., Booth, R., & Francis, M. (2007). *Linguistic inquiry and word count: LIWC* [Computer software]. Austin, TX: liwc. net.
Pennebaker, J. W., & Chung, C. K. (2007). Expressive writing, emotional upheavals, and health. *Handbook of Health Psychology*, 263–284.
Pennebaker, J. W., & Chung, C. K. (2011). Expressive writing: Connections to physical and mental health. In H. S. Friedman (Ed.), *Oxford handbook of health psychology* (pp. 417–437). Oxford, England: Oxford University Press.
Pennebaker, J., & Francis, M. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & Emotion*, *10*, 601–626. doi:10.1080/026999396380079
Peters, E., Hibbard, J., Slovic, P., & Dieckmann, N. (2007). Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affairs*, *26*, 741–748. doi:10.1377/hlthaff.26.3.741
Petrie, K. J., Booth, R. J., & Pennebaker, J. W. (1998). The immunological effects of thought suppression. *Journal of Personality and Social Psychology*, *75*, 1264–1272. doi:10.1037/0022-3514.75.5.1264
Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist*, *47*, 1102–1114. doi:10.1037/0003-066X.47.9.1102
Rabbi, M., Pfammatter, A., Zhang, M., Spring, B., & Choudhury, T. (2015). Automated personalized feedback for physical activity and dietary behavior change with mobile phones: A randomized controlled trial on adults. *JMIR Mhealth and Uhealth*, *3*, e42. doi:10.2196/mhealth.4160
Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A., & Riedl, J. (2002, January). Getting to know you: Learning new user preferences in recommender systems. *Proceedings of the IUI 2002 Conference on Intelligent User Interfaces*. New York, NY: ACM.
Robinson, M. D., & Clore, G. L. (2002). Episodic and semantic knowledge in emotional self-report: Evidence for two judgment processes. *Journal of Personality and Social Psychology*, *83*, 198–215. doi:10.1037/0022-3514.83.1.198
Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. C. (2014, April). Personal tracking as lived informatics. *Proceedings of the CHI 2014 Conference on Human Factors in Computer Systems*. New York, NY: ACM.
Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161–1178. doi:10.1037/h0077714

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, *52*, 141–166. doi:10.1146/annurev.psych.52.1.141

Ryff, C., & Keyes, C. (1995, October). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*, 719–727. doi:10.1037/0022-3514.69.4.719

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000, October). Analysis of recommendation algorithms for e-commerce. In *Proceedings of the 2nd ACM conference on Electronic commerce* (pp. 158–167). New York, NY: ACM.

Sas, C., & Whittaker, S. (2013). Design for forgetting: Disposing of digital possessions after a breakup. *Proceedings of the CHI 2013 Conference on Human Factors in Computer Systems*. Paris, France: ACM.

Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion: Theory, Methods, Research*, *92*, 120.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition & Emotion*, *14*, 433–440. doi:10.1080/026999300402745

Seligman, M. E., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: Empirical validation of interventions. *American Psychologist*, *60*(5), 410–421. doi:10.1037/0003-066X.60.5.410

Sellen, A., Fogg, A., Aitken, M., Hodges, S., Rother, C., & Wood, K. (2007). Do life-logging technologies support memory for the past? *Proceedings of the CHI 2007 Conference on Human Factors in Computer Systems*. New York, NY: ACM Press.

Sellen, A., & Whittaker, S. (2010). Lifelogging: What are we doing and why are we doing it? *Communications of the ACM*, *53*(5), 70–77. doi:10.1145/1735223.1735243

Sheldon, K. M., Ryan, R., & Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin*, *22*, 1270–1279. doi:10.1177/01461672962212007

Sin, N. L., & Lyubomirsky, S. (2009). Enhancing well-being and alleviating depressive symptoms with positive psychology interventions. *Journal of Clinical Psychology*, *65*, 467–487. doi:10.1002/jclp.20593

Sloan, D. M., & Marx, B. P. (2004). A closer examination of the structured written disclosure procedure. *Journal of Consulting and Clinical Psychology*, *72*, 165–175. doi:10.1037/0022-006X.72.2.165

Smyth, J. M. (1998). Written emotional expression: Effect sizes, outcome types, and moderating variables. *Journal of Consulting and Clinical Psychology*, *66*, 174–184. doi:10.1037/0022-006X.66.1.174

Sosik, V. S., & Cosley, D. (2014). Leveraging social media content to support engagement in positive interventions. *The Journal of Positive Psychology*, *9*, 428–434. doi:10.1080/17439760.2014.910826

Spotify. (2015). *Year in music*. Retrieved from https://yearinmusic.spotify.com

Ståhl, A., Höök, K., Svensson, M., Taylor, A., & Combetti, M. (2009, June). Experiencing the affective diary. *Personal Ubiquitous Computing*, *13*, 365–378. doi:10.1007/s00779-008-0202-7

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials*, *24*, 182–199. doi:10.1016/S0197-2456(02)00320-3

Tausczik, Y., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24–54. doi:10.1177/0261927X09351676

Tice, D., Bratlavsky, E., & Baumeister, R. (2001). Emotional distress regulation takes precedence over impulse control: If you feel bad, do it! *Journal of Personality and Social Psychology*, *80*, 53–67. doi:10.1037/0022-3514.80.1.53

Timehop. (2014). *Timehop*. Retrieved from http://timehop.com/

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research*, *27*, 247–259. doi:10.1023/A:1023910315561

Tsanas, A., Saunders, K. E. A., Bilderbeck, A. C., Palmius, N., Osipov, M., Clifford, G. D., … De Vos, M. (2016). Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *Journal of Affective Disorders*, *205*, 225–233. doi:10.1016/j.jad.2016.06.065

Turner, R., Ward, M., & Turner, D. (1979). Behavioral treatment for depression: An evaluation of therapeutic components. *Journal of Clinical Psychology*, *35*, 166–175. doi:10.1002/(ISSN)1097-4679

van Dierendonck, D. (2004). The construct validity of Ryff's Scales of Psychological Well-being and its extension with spiritual well-being. *Personality and Individual Differences*, *36*, 629–643. doi:10.1016/S0191-8869(03)00122-3

Van House, N., & Churchill, E. F. (2008). Technologies of memory: Key issues and critical perspectives. *Memory Studies*, *1*, 295–310. doi:10.1177/1750698008093795

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant–and memory helps to keep it that way! *Review of General Psychology*, *7*, 203–210.

Watkins, P. C., Vache, K., Verney, S. P., & Mathews, A. (1992). Mood-congruent memory in depression. *Journal of Abnormal Psychology*, *101*, 581–586. doi:10.1037/0021-843X.101.3.581

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. doi:10.1037/0022-3514.54.6.1063

Wildschut, T., Sedikides, C., Arndt, J., & Routledge, C. (2006). Nostalgia: Content, triggers, functions. *Journal of Personality & Social Psychology*, *91*, 975–993. doi:10.1037/0022-3514.91.5.975

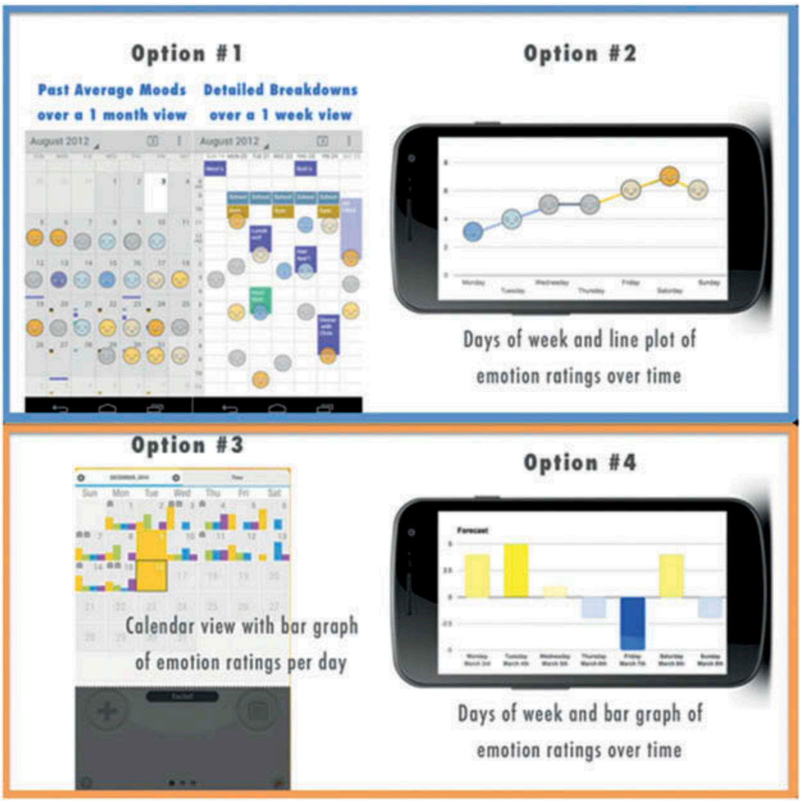
Wilson, T., & Gilbert, D. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, *14*, 131–134. doi:10.1111/j.0963-7214.2005.00355.x

World Health Organization. (2012). *Depression*. Retrieved from http://www.who.int/mediacentre/factsheets/fs369/en.

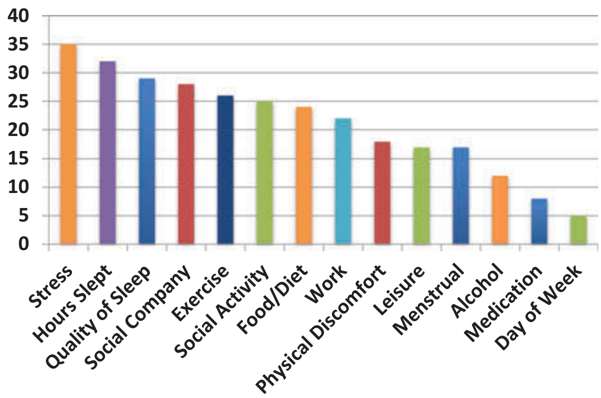
Zhao, X., & Lindley, S. E. (2014, April). Curation through use: Understanding the personal value of social media. *Proceedings of the CHI 2014 Conference on Human Factors in Computer Systems*. New York, NY: ACM.

APPENDIX A. EMOTICAL VISUALIZATION DESIGNS

Presented here are four designs of the EmotiCal visualization that were used to gain feedback from 39 participants via an online questionnaire, in addition to other quantitative and qualitative data we collected about expectations for which behaviors seem to affect their personal mood. From these designs, we extracted three possible templates that were prototyped and tested in-person with 15 participants.



APPENDIX B. USER RATINGS OF POSSIBLE TRIGGER ACTIVITIES



APPENDIX C. HAND-CODED POSITIVE ACTIVITY RECOMMENDATIONS

The following list is the pool of hand-coded activity recommendations. Hand-coded recommendations were based on individual ratings of enjoyment scores from the Positive Activities Schedule and pretest Basic Psychological Needs Profiles (ratings of autonomy, competence, and relatedness).

- Practice something you know you can do well
- Focus on doing some solo work today
- Find an interesting recipe to bake or cook
- Garden either outdoors or for an indoor plant
- Try to learn something new—You can find classes in your area or online courses
- Tidy up around home or try redecorating to make the space more comfortable and organized
- Play a board game, card game, or puzzles
- Watch a movie or TV show that you are looking forward to
- Listen to some favorite music or try finding new songs online
- Treat yourself to a relaxing shower or bath
- Explore a scenic place in nature
- Go for a leisurely drive by car or motorcycle

- Spend time outside by going for a walk. You can try exploring a nearby park or garden
- Visit a new place (like a nearby city or town)
- Engage in an artistic activity such as drawing or painting
- Find a relaxing place and meditate or do yoga
- Ride a bike or skateboard outside
- Hang out with a friend or family member or contact them by phone
- Play video games either alone or with friends
- Invite a friend or relative to the movies.
- Invite a friend or go solo to a museum, exhibit, or art gallery
- Invite someone or go solo on a picnic or hike outdoors