Mood Modeling: Accuracy Depends on Active Logging and Reflection

Aaron Springer
University of California Santa Cruz
1156 High St
Santa Cruz CA, 95060

Victoria Hollis
University of California Santa Cruz
1156 High St
Santa Cruz CA, 95060

Steve Whittaker
University of California Santa Cruz
1156 High St
Santa Cruz CA, 95060

Corresponding Author:
Aaron Springer
alspring@ucsc.edu
(309)613-2922

Keywords: mood, forecasting, system, reflection, machine learning, well-being, emotion regulation.

1.Abstract

Current behavior change systems often demand extremely advanced sense-making skills, requiring users to interpret personal datasets in order to understand and change behavior. We describe EmotiCal, a system to help people better manage their emotions, that finesses such complex sensemaking by directly recommending specific mood-boosting behaviors to users. This paper first describes how we develop the accurate mood models that underlie these mood-boosting recommendations. We go on to analyze what types of information contribute most to the predictive power of such models, and how we might design systems to reliably collect such predictive information. Our results show that we can derive very accurate mood models with relatively small samples of just 70 users. These models explain 61% of variance by combining: (a) user reflection about the effects of different activities on mood; (b) user explanations of how different activities affect mood; as well as (c) individual differences. We discuss the implications of these findings for the design of behavior change systems, as well as for theory and practice. Contrary to many recent approaches, our findings argue for the importance of active user reflection rather than passive sensing.

2. Introduction: Why we need accurate mood models to promote mental wellbeing

Many mobile healthcare applications aim to improve both physical and mental wellbeing, taking a behavior change approach [1–3] by supporting detailed monitoring of target behaviors, such as diet, mood or exercise. The goal of these healthcare applications is to enable users to derive insights about the consequences of specific behaviors on their health. These user insights can then inform decisions about appropriate behavior modifications to improve personal health.

The first wave of personal healthcare systems focused on simple behavior tracking. The ubiquity of mobile phones provides a straightforward way for people to monitor health-relevant daily behaviors to better understand how these might be improved. For example, a user experiencing sleep problems might employ an application that allows them to log multiple sleep-relevant factors including: exercise, diet, social activity, bedtime, work habits and so forth. By analyzing the effects of each of these factors, this sleep-deprived user it is argued should then be able to determine which of these trigger factors have critical impacts on sleep. In an ideal world, this careful analysis of system data should allow users to modify their behaviors to improve sleep.

While there have been important successes with such monitoring systems [4, 5], they have crucial limitations. One critical issue is that these systems place heavy analytic demands on users. For example users may struggle to disentangle what behavioral triggers affect their sleep, because there are so many possible factors, including: exercise, social interactions, diet, bedtime, and more [6–8]. Users may also have further difficulty interpreting the results of multi-factor tracking because factors can interact in complex ways. For example, it may be hard for users to generate the insight that a combination of minimal exercise and a late bedtime are maximally disruptive for sleep, if neither factor alone is deleterious. But how can we better support the majority of users who lack the advanced data analytic abilities to interpret time varying data with complex interactions between behavioral triggers [9]? This paper argues that such systems must provide user-centric data analysis tools allowing people to reason about these complex relations between triggers and target behaviors. In other words, these systems must support analytic *sensemaking*.

One response to the need for sensemaking has been to propose new types of analytic tools that provide interpretive support for complex personal health data. One class of tools offers simple visualizations illustrating correlations between trigger activities and health behaviors [10, 11]. Other systems provide natural language summaries describing how triggers are affecting target behaviors [12]. However there are limitations to these approaches. For example, these systems tend to explore simple relations between trigger activities and goals (e.g. late bedtime affects sleep), whereas in many cases there are complex interactions between multiple variables (low exercise combined with late bedtime reduces sleep) [6–8].

And even when users do correctly interpret the effects of combined behavioral triggers this does not guarantee behavioral improvements. Sensemaking is necessary but not sufficient; it must also be supplemented by a *remedial plan*. To return to our example, simply understanding that bedtime and exercise together influence sleep is not enough to positively change behavior. Rather, users also have to plan effective new ways to change those two behaviors to achieve positive effects. For example, a user might decide they need to implement a combination of an early bedtime of 10pm, allied with an exercise regime of at least 8000 steps to promote a good night's sleep. Other work has shown that such remedial planning works best if plans are concrete and executable [13], but such requirements are not always straightforward to satisfy.

Given these dual requirements of sensemaking and remedial planning, we therefore explore a different approach to behavior change in personal healthcare systems. Our novel approach uses predictive algorithmic modeling to provide recommendations to users about how to modify activities to promote effective behavior change. Rather than devolving the burden of analytic sensemaking and devising remedial plans to users themselves, instead we scaffold both these processes. To aid sensemaking, we provide predictive end-user models. These models allow users to straightforwardly determine which trigger factors affect their wellbeing. More importantly, we also offer users actionable recommendations about what remedial actions they might undertake to positively change behavior.

Our system addresses emotion regulation through the use of activity planning. It is well known that people experience major challenges in regulating their emotions; this has important consequences for emotional wellbeing. People are typically poor at predicting future emotions [14]; they overestimate the impacts that recent negative events will have on long-term affect. They also find it difficult to choose future activities that will improve long-term wellbeing [15]. Finally, when in a distressed state, many people tend to recall negative information rather than enhancing mood by remembering positive events [16, 17]. People who have difficulty overcoming these affective biases can experience severe negative consequences for their mental and physical wellbeing [18–20]. Access to algorithmic insights about one's own mood could alleviate these problems, especially if accompanied by remedial methods to positively improve mood [16]. These algorithmic insights may help users regulate mood by providing recommendations that are not tainted by the affective biases everyone experiences. Supporting mood regulation with software is promising, but relatively little research attempts to provide this mood regulation support for the general population.

To help people better understand and regulate their emotions, we designed and implemented a mobile phone-based system called EmotiCal (**Emoti**onal **Cal**endar, see Fig 1). Unlike many off-the-shelf applications, EmotiCal goes beyond simple mood and activity tracking. It supports predictive emotional analytics to help with sensemaking, allowing users to better understand how specific everyday activities influence their mood. EmotiCal also helps users generate remedial plans which take the form of personalized recommendations about new behaviors to improve mood.

Elsewhere we describe a 3-week intervention study demonstrating how engaging with EmotiCal's predictive analytics improves wellbeing and increases users' sense of control over their emotions [26]. In addition, users who engaged with EmotiCal's emotion forecasting and remedial activity recommendations were more successful at choosing activities that improved their mood [26]. The current paper instead focuses on system design, specifically the underlying mood modeling that underpins both sensemaking and remedial planning. To develop these models, we examine users' active evaluations of *trigger activities* that users believe affect mood, as well as the explanations they provide of how these factors influence mood. Trigger activities are common social, work and health related behaviors. Users log how often they engage in such activities within the application, and also provide information about each activity's effects on mood, e.g. a good night's sleep might enhance mood. We explore how these logged activities and explanations can be used to predict mood. Accurate mood models are critical to facilitating user sensemaking. With accurate mood models, users are able to better understand the factors underlying mood because of the reliable relationships those models capture between trigger activities and mood. More importantly, accurate models are necessary to provide compelling recommendations about remedial activities users might perform to improve mood. We explore exactly what types of data are needed to generate these models and what system designs might promote the collection of such data. While our focus here is on emotion regulation, the insights we generate have direct implications for a broad class of quantified self and behavior change systems that aim to help users gain insight into, and hence modify, important personal behaviors.

This paper describes how we developed accurate mood models to support sensemaking and remedial planning. Data were derived from deployments of the EmotiCal system with 70 total users who generated 2,875 logfiles to provide this data. One critical question we wanted to address was the role of active user evaluation. In our procedure, users actively engaged in reflecting about their mood and trigger behaviors in contrast to recent automatic approaches [3, 27]. This active reflection involved identifying which activities affect mood, weighing the effects of those activities, and generating explanations for those effects. This active user approach generates rich, systematic data but imposes additional logging burdens on users. Given these burdens, in addition to examining the role of active reflection, we also assess the additional effects such active data collection has on model accuracy.

We were also interested in individual differences between users; as it may be that certain classes of triggers have very different effects on different users. For example, for some people, activities related to work may be critical in determining their mood, whereas for others, their social behaviors might have much stronger effects. Our study therefore explored these differences. Finally, we were interested in the effects of temporal context on mood. One's current mood may be highly influenced by anticipated future events, such as an upcoming vacation, or by past events, such as a recent family reunion. We wanted to assess the contributions of recent past and upcoming future events on current mood.

We address the following questions:
- *Explanatory Models for Mood:* How do we derive accurate mood models? How do different sets of trigger activities, specifically social, health, and work, affect mood? Additional questions include: Does active user reflection about activities improve mood models? Do user explanations also improve models? Are there individual differences between users? How do past and future events affect current mood?

- *System design:* Based on the above models, how might we design effective systems to better track, capture, and predict mood? What types of data are needed for accurate modeling? What system design features allow such data to be collected?

3. Prior Approaches to Modeling Mood

A mood is a generalized emotional state that is typically described as having either a positive or negative valence. Moods tend to be longer lasting than discrete emotions such as anger, joy or fear. Moods are less specific, less intense, and less likely to be triggered by a particular stimulus than discrete emotions [28]. Energy is a different concept that relates the strength of the experienced emotion [24, 29], and is commonly synonymous with arousal. Mood and energy interact in complex ways, although there is no consensus about the exact nature of such interactions [24, 30, 31]. Theoretical and empirical work indicates that reporting both mood and energy together improves the accuracy of mood evaluations [24].

3.1 Relationships Between Activities, Mood and Wellbeing

Our main goal is to develop models that accurately characterize relations between user activities and mood. Few studies have modeled how daily personal activities affect mood; most of these have focused on non-digital contexts. Daily personal activities have been examined by Stone et al. [32] who developed the Day Reconstruction Method (DRM). The DRM involved participants writing both about the activities they engaged in during the previous day, along with their accompanying feelings. Activities with highly positive effects on mood included exercising, socializing, intimate relations, and relaxing. Activities that engendered negative moods included doing housework, commuting, and working. However, other factors that are unrelated to specific activities also affected general mood levels, including pressure to work quickly and sleep quality. Another study found that people are prone to inaccurately evaluate how normal activities affect mood [33]. People in that study were more likely to rely on folk theories about mood (e.g., Fridays are happier days, Wednesdays are unhappy) than their own direct personal experiences.

A very different approach to understanding and improving mood is the Pleasant Events Schedule (PES). The PES explores the effects of different activities on mood using a combination of retrospective reporting and intervention methods. The PES is a behavioral self-report inventory that focuses on positive events in participants' lives [34]. In the PES users are asked to retrospectively report how frequently they engage in various activities and also to rate the 'pleasantness' of each activity. For example, seeing old friends is reliably judged as a highly pleasant activity, whereas physical discomfort is judged to be highly unpleasant [35]. While participants' subjective evaluations provide useful information about how different activities affect mood, these evaluations can take place as long as a month after the activity took place. This delayed evaluation may be problematic, as we know that emotional appraisals of activities and events change over time; current evaluations are not always reliable indicators of how people feel immediately after an activity [2, 36, 37].

Despite potential problems with reporting accuracy, interventions based on the PES show clear positive results. In these interventions, participants are encouraged to deliberately and repeatedly engage in activities they previously rated as pleasant, and the resulting effects on mood and wellbeing are then evaluated. In one study, the PES was given to depressive patients along with a depression scale [38]. Patients were then reminded about their most positive uplifting activities and encouraged to improve their moods by engaging in the uplifting activities. Patients' depression scores improved by 19% following this intervention. These results also generalize to non-clinical populations. University students instructed to increase their number of pleasant events derived from the PES obtained more pleasure than a control group [22]. Interestingly, only the students who were instructed to cognitively reflect on their positive feelings while engaged in the activity decreased their depression inventory scores. People who simply engaged in pleasant events without reflection increased pleasure but did not improve on a depression inventory. In general, positive activity scheduling interventions have large positive effects. A meta-analysis of 16 studies with over 780 combined participants found an effect size of .87 [18]. Pleasant activity scheduling was found to be among the most effective treatments for depression, on the same level as traditional cognitive behavior therapy techniques. In contrast with our work, the goal of these prior studies was to evaluate the intervention effects of specifically positive activities rather than modeling how normal daily activities influence mood.

Similar intervention methods for boosting positive affect have been explored in positive psychology. These studies are relevant because they show which types of activities influence mood. Seligman et al. [39] analyzed the effects of five different positive intervention strategies to increase positive affect. These strategies included expressing gratitude, performing acts of kindness, noting three positive things that happened each day and so forth. All five strategies promoted significant improvements in positive affect with some effects persisting up to 6 months after the intervention. In a rare digital intervention, Parks et al. [40] also conducted Positive Activity Interventions (PAI) through the iPhone application LiveHappy. Parks et al. tracked self-selected LiveHappy users using wellbeing scores to assess the effects of the recommended positive affect strategies. In the intervention, LiveHappy compared 8 mood enhancing strategies (expressing gratitude, writing about gratitude, focusing on meaningful goals, savoring the moment, replaying happy days, performing acts of kindness, nurturing interpersonal relationships, "considering one's best possible self"). All strategies led to improvements in overall mood. In both these interventions, the goal is to evaluate the effects of high level strategies for improving affect; in contrast we examine relationships between routine daily activities and mood.

3.2 Systems For Tracking Mood That Aim To Change Behavior

Until recently, most systems for wellbeing focused on physical health, for example, tracking goals relating to weight loss by documenting food consumption or daily step count. The implicit assumption behind these systems is that simply capturing and presenting detailed records of physical behavior will be sufficient to allow users to change target behaviors to meet goals. This approach has not always been successful. First, data about daily activities is often complex, requiring users to interpret multiple streams of time-varying data. Second, much of the general population has low numeracy skills, creating challenges for end-users when making sense of such complex data, and making it critical to design effective end-user analytics [9]. A further limitation of basic tracking systems is that they are overly rational; ignoring both users' affective reactions to their logged data as well as their emotional motivations for tracking and changing behavior [41]. It is clear that behavior change applications must facilitate better sensemaking for personal data.

In addition to systems that promote physical wellbeing, there are now more systems that primarily focus on emotional wellbeing. Echo was one of the first systems that addressed emotional reflection for psychological wellbeing [2, 16]. Echo is a digital diary in which users post about recent experiences as well as their current mood, writing text to explain their mood evaluation. Echo deployments have shown that both posting and reflection upon prior posts improve emotional wellbeing. Furthermore, the effects of Echo are long-lasting with improvements still evident after 4 months [16]. Other work has explored mood-dependent reflection, showing that reflection on positive memories can elevate a current negative mood, and that reflecting on negative events when in a positive mood leads those prior negative events to be more positively viewed [16]. Another reflection application, Pensieve, used a different design approach; emotional reflection was encouraged by having users reflect on prior social media content or respond to targeted prompts [42].

Early systems mainly supported manual text entry about mood. Recently, however, we have seen the emergence of new automatic sensing applications that aim to model mood based on a variety of automatically detected behaviors. MoodScope analyzes phone usage patterns such as number of SMS messages, application usage, call frequency, and call length to infer mood [3]. BeWell is another application that incorporates passive sensing to detect physical, social, and sleep activities. BeWell blurs the line between physical and mental wellbeing in an attempt to create a holistic system; it aims to explore how different activities affect mood as well as physical wellbeing. However, experimental trials to validate the models underlying BeWell are lacking [27]. Furthermore, there may be major limitations to approaches that involve passive data collection. Recall that one of the major questions we examine in this work is the role of active personal data recording compared to more passive approaches such as BeWell and MoodScope.

These research systems purport to support end user sensemaking by providing correlations between data streams, e.g. relations between mood and sleep, but a different approach has been taken in MONARCA [21, 43, 44]. MONARCA is focused on emotional analytics; it allows bipolar patients to actively 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 they skip medication, fail to exercise, or sleep poorly. Unlike many of the prior systems, MONARCA was deployed to a targetted clinical population to explicitly test the effects of such analytic support. Seventy-eight 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 [44]. While an ongoing trial is exploring improved analytic support, this MONARCA evaluation highlights the need for additional work on emotional analytics and a greater exploration of possible benefits for nonclinical users.

4. Methods

4.1 EmotiCal System Overview

EmotiCal users actively record mood, energy level, and up to 14 trigger activities that users believe have influenced their mood. For example, they can track social interactions (e.g., time spent with a friend or coworker), aspects of physical health (e.g., sleep or exercise), and work activities (e.g. meetings) to log these activities' effects on mood. EmotiCal also encourages active reflection by evaluating exactly which activities have affected their mood. EmotiCal also prompts users to generate short explanations of how and why they think those activities have affected mood. This active reflection has been shown to be important for behavior change [2, 22]. EmotiCal uses this logged information about mood and activities to create an individualized mood model for each user, predicting how different trigger activities influence the user's mood. Users are also encouraged to report energy levels separately from mood valence as this has been shown to provide more accurate information about one's emotional state [23, 24].

Fig 1 illustrates the main functions of EmotiCal. The left-hand panel shows the landing page visualization for EmotiCal users. This visualization allows users to engage in *affective forecasting* about their future mood, as

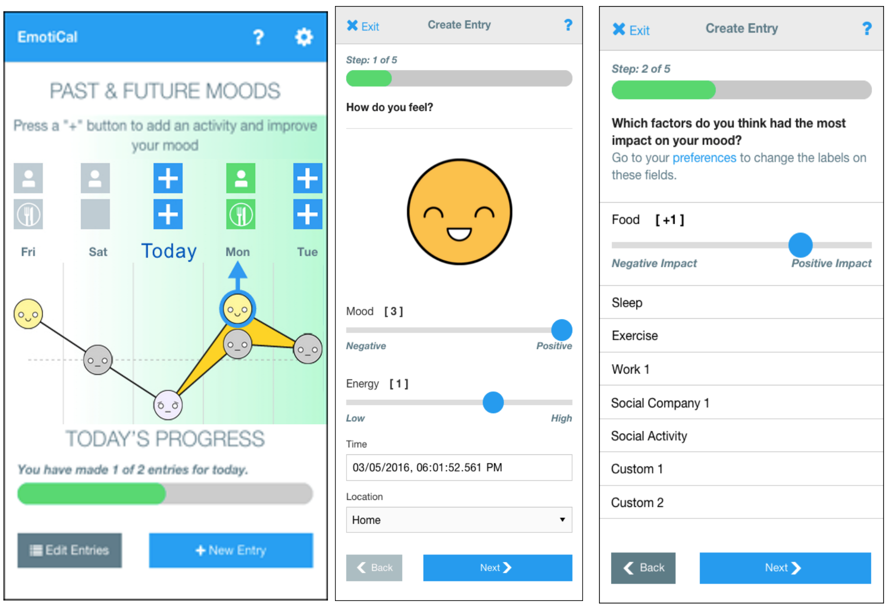

Fig. 1 EmotiCal System Components. The first screen shows the mood-forecasting component. The second and third screens show parts of the logging.

well as sensemaking analytics to plan future remedial activities to improve mood. The center panel shows the *mood-monitoring* interface with options to rate mood and energy level, as well as contextual information, e.g. time and location. The right-hand panel shows the UI for *evaluating trigger activities* that led to current mood (e.g., that food had a positive impact on current mood). There are a total of 14 possible activities the user might select as affecting mood, although not all are shown in this UI view. Models of the relations between these activities and mood are used to generate recommendations about potential remedial plans shown in the left panel. We discuss mood scales and trigger activities in more detail below.

EmotiCal uses this historical monitoring information to predict each user's expected general mood for two upcoming days. Past and future expected moods are presented to users in a visualization (see left-hand panel of Fig 1). Sensemaking and remedial planning are supported through interaction with this visualization. Users are encouraged to actively manipulate their future mood by adding recommended mood-enhancing activities to their schedule. Two slots (+s) are displayed above the visualization showing moods of today, tomorrow and the day after tomorrow. Users can click on a slot and see a list of recommended activities. Recommendations are derived from a user's logs (*history-based*) or their psychological needs profile (*needs-based*). History-based recommended activities are derived specifically from the user's own past data; this allows EmotiCal to propose actions that the user's own logging data indicate have positive past relationships with mood. Needs-based activities are generated by profiling each user's psychological needs as assessed using the Basic Psychological Needs Scale (BPNS) [25], and generating activities that meet those needs. This paper focuses on history-based profiling as these recommendations are directly derived from modeling.

After selecting a recommended activity by pressing a '+' above the visualization, users are prompted to schedule that activity. Past research shows that concrete implementation intentions improve the likelihood of following through with a plan [13]. Textual feedback then summarizes the activity plan (e.g., "*At 9am tomorrow, I will go for a run.*"). The user then writes a brief description of the expected benefits from engaging in that activity and any additional planning information necessary; prior work again shows this to improve intervention effectiveness [20]. After finishing activity planning, the visualization then updates to show the predicted changes in mood resulting from adding the new action. In Table 1 are two examples of planned

User ID	**Recommendation**	**Activity**	**User Explanation of Anticipated Mood Benefits of Planned Activity**
80126	History-Based	Food	It helps me gain more energy and feel happiness. I will go to my favorite restaurant around 6pm tonight.
42968	History-Based	Work	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.

Table 1: Remedial Plans Created by Participants. These remedial activities were recommended by EmotiCal based on the user responding positively to this activity in the past.

remedial activity entries showing sources of activity recommendations derived from log file modeling. The Activity column shows the type of mood boosting activity planned, and the Explanation column shows user-generated expectations of how activities will help along with anticipated benefits. Thus User 80126 plans to boost mood by scheduling a Food related activity involving going to her favorite restaurant at 6pm that night and 42968 plans to write each day to master that important skill. EmotiCal recommends user-specific enjoyable activities; adding an activity increases the expected mood for the planned day. For example, User 42968's Work activity plan of daily writing would increase his estimated mood for the next day from 'slightly happy' (+1 on the mood scale) to 'happy' (+2 on the mood scale) on the later-explained full mood scale of -3 (very negative) to 3 (very positive). The magnitude of this estimated mood change is calculated by the mood models we describe in this paper.

We now turn to our system deployment in which users used EmotiCal to monitor their emotions and activities over a 3-week period.

4.2 Mood Monitoring, Energy Level and Trigger Activities

EmotiCal monitoring involved users logging information about mood and trigger activities that potentially contribute to current mood, and actively weighting the extent to which that trigger affected mood. Users also generated short text benefit descriptions explaining how and why they thought those triggers activities affected their mood. EmotiCal prompted users to create at least 2 mood entries per day with notifications that encouraged them to submit at least one morning entry and one evening entry. Prior research using similar methods indicated that allowing users to make entries on their own schedule led to more carefully considered entries and better compliance than system-generated prompts [2, 16]. However, this approach may preclude users from recording in specific situations, if they are stressed, busy or engaged in social interaction [2, 16]. Users were prompted via automatic text messages if they did not spontaneously submit a minimum of 2 entries per day.

Making a mood entry was lightweight and could typically be done in about 40 seconds. To create a mood entry, users first make a simple mood valence and strength decision, choosing a mood ranging from -3 (very negative) to +3 (very positive) (see center panel of Fig 1). This scale is similar to the valence row of the Self-Assessment

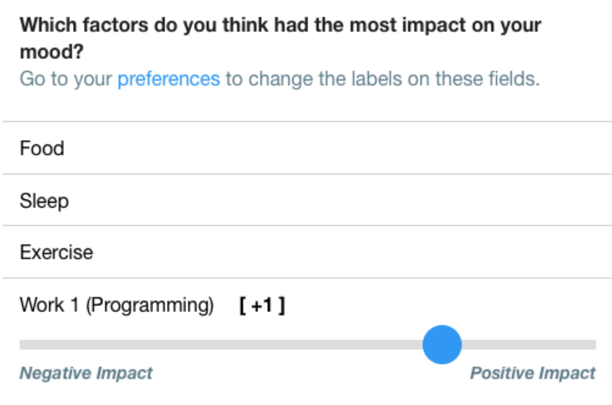

Fig. 2. Users record which trigger activities influenced their current mood on this screen of the application

Manikin (SAM) though extended to 7 ratings rather than 5 [45] to allow for finer mood granularity. Similar to SAM, we showed a face that changed as different moods were selected. Following prior work, we also included energy level evaluations to increase the precision of mood ratings (e.g., a cheerful emoticon may have a connotation of excitement when the user wants to convey feeling calm) [23]. Energy ratings ranged from -3 (low energy) to +3 (high energy). Users could also optionally change the time and date of the entry or set a location (Home, Work or Other).

After selecting mood, users were prompted to engage in active mood analysis. Users were asked to identify which of 14 possible trigger activities influenced their mood and to rate that influence on a scale of -2 (negatively impacted mood) to +2 (positively impacted mood). The (-2, 2) scale was based loosely on the Positive Events Schedule [34], we extended this from a 3 item rating to a 5 item rating in order to allow users provide more sensitivity for our modeling algorithms. Users could choose as many activities as they felt were relevant, although most users chose relatively few per entry. The mood-analysis component of the UI is shown in Fig 2.

The 14 trigger activities that we incorporated into the system for mood analysis were identified from three different sources: a log file analysis from a previous study of mood-tracking [16], surveys (n=39) and interviews (n=12). In each of these contexts, informants were asked to identify activities that affected their mood. By far the most frequent trigger activities discussed were food, sleep, exercise and general social activity. However in addition to these common triggers, all informants mentioned several more esoteric personal activities that also had emotional impacts. Informants discussed the effects of highly customized, specific activities such as particular leisure activities (e.g., knitting) or socializing with a particular person (e.g., a romantic partner). These esoteric activities fell into three main categories: leisure, work and social domains, with informants generally mentioning more social factors than leisure or work triggers. Survey respondents who described work largely emphasized negative impacts on mood. Given the goal of EmotiCal was to use log files for positive activity recommendations, we prioritized customized leisure options in order to motivate the later activity recommendations. A final class of triggers fell outside these three domains, for example alcohol intake, menstrual cycle, or finances. To account for the wide range of other triggers for mood, we included 2 custom options. All participants were given an instructions document that explained how to set mood triggers, and discussed with a researcher in an onboarding phone call the trigger creation process. Participants were instructed to not change triggers once they were initially set.

This requirements data informed the design of the EmotiCal UI for tracking and analyzing triggers, where there were 3 main classes of trigger.

- Default activities: The UI first probed the four commonly reported default trigger activities (food, sleep, exercise, general social activity).
- Non-default: Then, to address the issue of esoteric triggers, we also allowed people to track non-default activities. Following our requirements analysis, and to provide users with some guidance, we explained during setup that non-default activities could be of three general types, work, leisure and social, and we provided some examples. A user could therefore decide that ‘Playing Music’ was often important for

their mood, and so we allowed them to set it as a non-default trigger activity. Or they could decide that 'Processing Email' was a work activity that affected mood and decide to track that.

- Custom: Finally, because we didn't want to restrict users, we also told them that they could also define other Custom activities that didn't fall into these prior categories.

To ease tracking, non-default and custom activities have an editable title. Thus triggers discussed above might show up in the system as 'Leisure 2 (Playing Music)', or 'Work 1 (Processing Email)'. Overall users could set up to 10 customized triggers of which 3 were social, 3 were leisure and 2 were work, and two were totally uncategorized, motivated by the triggers identified in our initial requirements samples.

After choosing activities that affected their mood, the UI encouraged users to submit a brief free-write explanation about how those triggers activities impacted their mood (e.g., "*I really love the TV shows I watch. Class today was too demanding and draining.*" – User 13489). Again, data entry was lightweight and took around 40 seconds on average, as users tended to select a small number of triggers (Mean=2.3, SD=1.21). For the trigger activities UI, see Fig 2.

4.3 Users

Ninety-two users were recruited through Craigslist, Facebook, Quantified Self forums, university classroom announcements and flyering. Users were eliminated from our analysis if they were noncompliant, which we define as entering fewer than 10 entries over the course of the three-week study. This criterion eliminated 22 users from our analysis, resulting in 70 compliant users. These compliance levels are consistent with those reported in similar [2, 16]. The final sample consisted of 48 females and 21 males and 1 unspecified person, (Mean age=30.7, SD=10.26). Users received a $20 Amazon gift card as compensation for participating.

4.4 Procedure

Users were told that the research goal was to beta-test a new technology to help regulate mood and improve wellbeing. Users first completed an online pretest, consisting of a set of surveys to assess emotional wellbeing and behavior frequencies with enjoyment ratings for those behaviors [26]. We then emailed users a web-link to EmotiCal with login information. To maintain user compliance, researchers individually contacted users by text and phone within the first week; this ensured that users were consistently submitting entries and addressed any technical errors or confusion over study instructions. We also scanned server logs to confirm that users were indeed making daily entries, correctly following instructions and were not submitting records that would raise concern (e.g., self-harm). We contacted users to answer the post-test survey three weeks after the start date; they were debriefed, thanked, and given the opportunity to delete or modify any data they wished to keep private before data analysis. Overall users generated a total of 2,875 logfiles.

5. Results

Accurate mood models are necessary to support compelling activity recommendations. If models are perceived to be highly accurate, users are more likely to engage with the system and use its recommendations [26]. Multiple different types of data potentially affect our models, including trigger activities, user explanations, and individual differences. We now explore the effects of each of these different types of information to model mood, with the aim of developing the most accurate models possible.

5.1 Modeling the Impact of Activities on Daily Mood

Mood ratings were recorded on a 7 point scale (-3 to +3). We chose regression rather than classification based machine learning to better quantify the exact contributions of each feature towards mood. Regression models were trained using Ordinary Least Squares regression from the statsmodels Python library [46]. Feature selection for textual models was done using scikit-learn [47]. All R^2 reported are adjusted for the number of features.

We first analyze trigger activities that users logged. Next we examine the textual explanations associated with each mood post, in which users analyzed how and why they felt that those activities would benefit mood. Finally, we explore individual differences in how different activities affect mood.

5.2 Activities are Critical for Explaining Mood: Health and Social are Important

We begin by regressing on the 14 trigger activities in order to predict overall mood across all users. The model was highly predictive $R^2 = .434$ and highly statistically significant ($p < .000001$). We refer to this model in Table 2 as the Activities model because it solely uses the trigger activities that users record. In order to compare beta coefficients of activities in this model, trigger activities, and mood were both normalized to [0,1]. The regression was calculated using all the entries created across users. These coefficients are graphed in Fig 3. We examined the model to determine which types of activities most affected mood, by exploring the relative weightings of the activity categories on the overall model. The beta weights for the different categories are shown in Fig 3, indicating that users felt that most activities had positive effects on mood. All 14 factors were significant at $\alpha = .001$.

The next critical question we wanted to address was the role of active user evaluation. In our procedure, users actively evaluate and weight the role of trigger activities; they determined both which activities affect mood, as well as weighting the effects of those chosen activities. How critical are such active weightings for generating

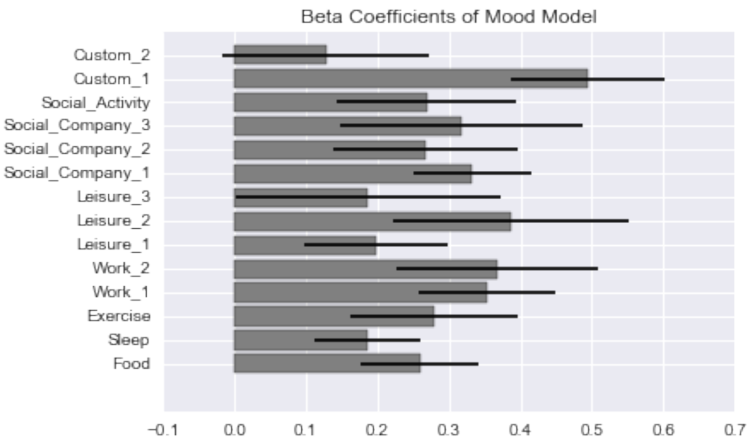

Fig. 3: Coefficients of Trigger Activities Predicting Mood.
Bars indicate the standard error of each coefficient

accurate models, as opposed to less burdensome automatic approaches that track activities automatically?

The accuracy of the model appears to result from users' active appraisal of the extent to which activities affected mood, rather than the simple fact that they engaged in these activities. For example, a user simply recording that they slept is not highly predictive of mood. In contrast, adding the user's active appraisal informs us both how well they slept and how that user feels that sleep affected their mood. User 86753 does this exactly in one of their entries "*I got enough sleep last night, … so I am in a good mood.*" To evaluate the model difference between active evaluation of activities and simple presence/absence of activities, we removed the self-evaluation in user entries; we did this by mapping the [-2,2] scale for activities to a simple binary [0,1]. Ones replaced any non-zero entry and zeroes remained, representing an effect/no effect contrast. In other words, we ignored all weightings and treated all cases where users rated an activity as equivalent. We then calculated the relationship between mood and these binary effect/no effect features. This model without active evaluative weightings loses nearly all its predictive power, resulting in an R^2=.037. Active user weighting is therefore critical for accurate mood models.

In order to better compare differences between different classes of activities, we constructed a hierarchical model that aggregates the data from the 14 individual trigger activities into superordinate categories. We did this by combining the effects of the different health related activities, so that sleep, food, and exercise trigger activities were aggregated into a common 'Health' category. Likewise, we combined other activities into Work, Social, Leisure, and Custom. We modeled the relations between these superordinate categories on mood, to determine the mood effects of each. The model using superordinate categories was again highly predictive with $R^2 = 0.414$, $p < 0.000001$.

We then calculated significant differences between the superordinate categories; we did this by comparing the Pearson correlations between pairs of Activity triggers and mood while controlling for the correlation between each pair. Overall Health and Social factors have a larger impact than all other factors. There was no difference between the effects of each of the other activities. The other 3 activities: Work, Custom, and Leisure, were not significantly different from each other. This result supports Parks et al. [40] whose users reported that developing and maintaining social relationships was the most important and meaningful activity for user happiness.

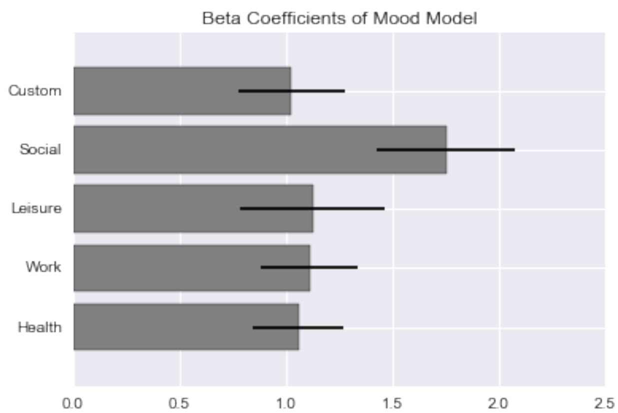

Fig. 4: Category Model Predicting Mood. Bars indicate the standard error of each coefficient

Type(s) of Data Included in Mood Model	R^2	Significance (p value)
Activities	.434	<0.000001
Explanations	.442	<0.000001
Activities + Explanations	.565	<0.000001
Activities + Explanations + Individual Differences	.613	<0.000001

Table 2. Contribution of Different Types of Information to Predicting Mood

Our primary goal was to develop as accurate model of mood as possible to enable sensemaking and remedial planning, where planning requires proposing specific activities. We therefore return to the original 14-factor Activity model to explore other factors that contribute to mood. The next factor we examined was user explanations.

5.3 User Explanations Improve Mood Models

In addition to actively weighting activity categories, users also added a short text description explaining how and why specific activities affected mood. Below are a few examples from users:

- "I woke up feeling really sick and then dealing with an hour commute on public transit made me feel worse" [User 10606]

- "I had four meetings today but being around my friends made it a little better." [User 13489]

- "Secured a much coveted freelance gig" [User 71153]

We analyzed the text in the user explanations to determine whether it offered information that improved our prior activity based mood models. We took two approaches to analyzing this text in these explanations. The first approach used Linguistic Inquiry Word Count (LIWC). LIWC is a tool that analyzes individual words and categorizes them into different linguistic categories, e.g. the words 'hate', 'fear', and 'rage' are all classified by LIWC as negative emotions [48]. Using LIWC, we first analyzed the percentage of words that mentioned positive (e.g., happy, enjoyed, bliss) and negative emotions (e.g., sad, depressed, angry). Overall, 6.79% of total words were positive and 2.03% were negative. Sixty-six and a half percent of explanations contained positive and 25.3% negative words, with 15.4% containing both. We wanted to see the extent to which providing explanations that referred to positive or negative feelings predicted users' mood judgements. So for each user explanation we determined the percentage of words within each explanation that were negative and the percentage that were positive and regressed this against the mood rating. Use of positive or negative emotion words was indeed related to mood in a multivariate regression model, $R^2 = .18$. Adding these linguistic categories to the Activity model, led to a modest improvement in that model from $R^2 = .434$ to $R^2 = .474$.

However, one limitation of LIWC is that it relies on fixed mappings between specific words and predefined emotional categories. However, people often talk implicitly about their emotions [49]. To address this implicit expression of emotion, we next modeled language in posts using unigrams in a second analysis. Unigram

Feature	Coefficient
negative	-3.980933
issues	-3.033704
negatively	-2.87173
all	-2.37859
don [don't]	-2.349698
sick	-2.168835
headache	-2.075584
not	-1.993266
able	1.600763
didn [didn't]	-0.883192
with	0.74027
tired	-0.492941
enough	0.416083

Table 3: Unigram Correlations with Mood from User Explanations After Removing LIWC Words. Many of these unigrams point towards the importance of implicit emotion recognition in text.

modeling examines whether the use of specific words by users correlates with changes in mood. Using unigrams also improved the basic activities model, but this time more significantly. Unigrams of user explanations alone generates a highly predictive model (R^2 = .442) and adding unigrams to the Activities base model improves it by .131 to R^2 = .556. We denote the added unigrams of user explanations as 'Explanations' in Table 2.

We hypothesized that this additional improvement with unigrams occurred because implicit emotions in users' text entries might be missed by LIWC. To examine this, we identified the top 30 most predictive unigrams chosen by F-score. F-score is a common way to measure feature importance [50]. After excluding 1 proper name, we subtracted the intersection of the top unigrams with the list of words that LIWC categorizes as positive or negative emotion. This allowed us to identify terms not captured by LIWC classification. This subtraction left us with 13 unigrams in Table 3. Following Goyal et al. [51] we note that many of these unigrams are implicit expressions of events related to negative emotions. For example, negatively weighted terms refer to experiences such as having a 'headache' or being 'sick'. These negative experiences are likely to affect our emotions negatively even though they aren't an explicit expression of emotion. In the same way, positively weighted terms such as 'with' may refer to important social experiences or 'able' may express competence, both of which are important for psychological wellbeing [25]. Such expressions would not be captured by LIWC, nor would they be captured by the Activities analysis, which focuses on specific (generally positive) activities. However, it's clear these Explanations provide valuable data for our mood models.

5.4 Individual Differences

We explored model personalization to see whether incorporating individual differences would further improve model quality. During the deployment, we built individual regression models on a per-user basis rather than using the general models trained across all participants mentioned above. Combining and averaging the metrics of these 70 individual models gives us a rough picture of how they performed. Averaging the R^2 of the 70 individual models resulted in a R^2 of .45 with a standard deviation of .19. However, creating a model for each individual user has both advantages and disadvantages. One major disadvantage of creating individual models is the cold start problem [52]. User models don't become accurate until they accrue enough data, which may take weeks. An alternative to this is to use the previous generalized Activity model and add a feature for each user.

To account for individual differences we use our general Activities model but add a binary feature for each user into the overall model, therefore adding 70 binary features. One unique binary feature of these 70 is set to 1 for each entry, depending on which user made the entry. The linear regression then learns a baseline for each of these 70 binary features, i.e. each user. The binary feature denoting each user essentially acts as an individualized intercept. This model takes into account individual differences in base mood and increases the general model's R^2 to .515. This is notably higher than the initial Activities mood model where R^2 = .434. Table 2 shows that when we added Individual Differences to the Activities plus Explanations model, the R^2 increased from .565 to .613.

5.5 User Subgroup Modeling

Another way to tackle the cold start problem is by segmenting users into groups and building group mood models. This could allow a system to quickly categorize a starting user and provide them with an accurate model from the group. For example, one group of users' mood may be affected primarily by Social and Work Activities, whereas for others, mood may depend on Health and Leisure. We analyzed the group structure in our population using k-means clustering. Recall that each user provided a subjective emotional appraisal of some of the 14 activities for each entry; this appraisal ranged from -2 (very negative) to +2 (very positive). We applied Principle Components Analysis (PCA) to this data set to reduce the number of extraneous features. Following standard methods, the number of principal components (n=2) was chosen through examination of the scree plot [53]. The principal components of each user's activity evaluations were clustered using k-means. The number of clusters (k=3) was chosen by its silhouette score [54]. Silhouette scores were generated from k=2 to k=10 and were highest at k=3, averaging .72 from a maximum score of 1. A silhouette score of .72 indicates that a strong cluster structure has been found [55]. Figure 6 illustrates the clusters that resulted.

We see marked differences between the clusters as indicated by their mean subjective scores for the activity categories in Figure 5. One of the clusters (shown in blue) is generally less positive about the activities they engage in than the other two clusters. The other two clusters (red and green) are generally more positive but differ significantly on a few categories (Work 2, Social Company 2, Leisure 2).

We then used these clusters to build group models for mood. We anticipated that these models would improve

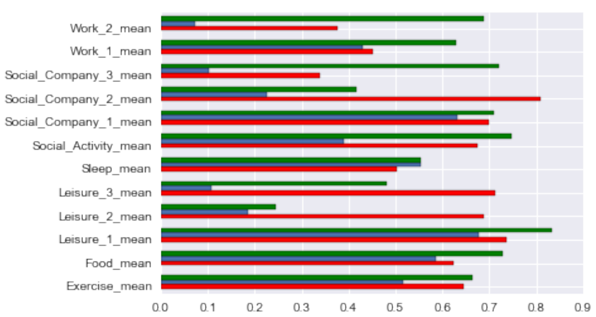

Fig. 5 Mean Scores for Activities from Different Clusters Showing Differences in Aggregate Ratings by Cluster

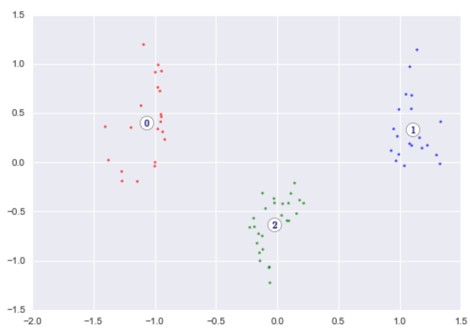

Fig. 6 Cluster Distribution Illustrated by First 2 Principle Components

the accuracy of the generalized mood model across all users by taking advantage of the differences we saw in how those groups of users evaluated different activities. We used the same features and process as the Activities model, and we trained mood prediction models at the cluster level. This resulted in three cluster models with varying R^2 of (.426, .421, .470). Models trained on random subsets of equivalent size to the clusters had R^2 = (.450, .438, .422). As an additional test, we added the cluster labels back into the original regression as a feature; this only improved the original R^2 by .005. Despite the good separation of clusters according to the silhouette score, we fail to find cluster specific regression models that improve on the generalized models. It may be because clusters are redundant with information already present in our models.

These prior mood models have all taken a situational perspective, that mood is affected solely by activities that have occurred immediately prior to the mood judgment. However prior studies suggest that mood might also be affected by non-proximal events [16], which we now explore.

5.6 Inertia and Anticipation Have Small Effects on Mood

We went on to examine the inertia of everyday activities on mood. Our prior analyses looked exclusively at effects of current activities on mood. However, it is possible that mood is affected by activities that were carried out before or after the current time. For example, early morning exercise may have long-term inertia creating a positive mood throughout the day. We define these longer term effects as *mood inertia effects*. In the same way, anticipating upcoming events may prospectively influence current mood. Thinking about spending an evening with a good friend might elevate mood throughout the day. We define these prospective effects as *anticipatory mood effects*.

To examine mood inertia effects, we calculated how activities from previous entries influenced a future entry's moods. Given a series of entries $[t_1, t_2, \ldots t_n]$ where t is a feature vector and chronological entry number is denoted by the subscript. We predict mood of t_i using a lag of one entry by regressing on the concatenated feature vectors t_i and t_{i-1}. Mood inertia for 2 previous entries of activities was calculated in a similar fashion by concatenating feature vectors t_i, t_{i-1}, t_{i-2} to predict the mood score of entry t_i. To calculate anticipatory mood effects, the reverse was done (e.g., the mood of t_i is predicted by concatenating the features of t_i, t_{i+1}, t_{i+2}).

To examine the predictivity of mood inertia and anticipatory mood effects we augment the Activities model with entries from before or after. For mood inertia, if someone made mood entries in the morning and afternoon of a particular day, when predicting their afternoon entry, we added activities from their early morning entry to the model to measure inertia. Likewise we included the afternoon's activities in the morning mood model to examine anticipatory mood effects.

There was a marginal effect of augmenting Activities with these additional sources of information. We noted a small .016 increase in adjusted R^2 of the Activities model to R^2=.451. Extending mood inertia effects by adding both the entry prior and 2 entries prior's features further increased the R^2 by .003 to .454. These results give credence to the view that mood is not lastingly affected by the past activities we asked users to log. For anticipatory mood effects we also noted rather small increases. Adding the next entry's features resulted in a .01 increase in R^2 to R^2=.444 and adding the two consecutive entries further increased R^2 by .004 to .448.

We expected that mood inertia and anticipatory mood effects would add valuable information to our mood models. There are a few possible reasons that this analysis did not produce more predictive results. One reason may be that users choose when they want to make entries; this can lead to inconsistent timings between users' entries which makes effects hard to detect. Some users occasionally missed consecutive days of entries, which would lead to our inertia predictions using stale data from multiple days before. With enough variation in this inconsistency, it could lead to more noise in the data rather than signal. A final possibility is simply that the activities that we record are mundane. It is plausible that only more major life events would have more notable temporal effects.

6. Discussion

Our modeling results are very encouraging in the context of systems like EmotiCal. Our models were indeed accurate as shown by the amount of variance in mood they could objectively explain, as well as by users' subjective evaluations of their accuracy. Simple Activities explain 43.4% of the variance and models additionally incorporating Individual Differences and User Explanations account for 61.3% of the variance. These are important findings because users are only likely to adopt recommended mood boosting activities when a system makes accurate predictions about projected effects of those activities on mood [26].

These mood models derive much of the predictive power from users' careful weighing of the extent to which each activity has affected mood, rather than the simple fact that the user engaged in those activities. We showed this by comparing models in which activities are given weights by users, with an alternate model in which each activity is represented in a binary fashion as relevant or irrelevant to mood score. The binary model had extremely weak ability to predict mood, showing the importance of active user weightings. The highly predictive nature of the active user evaluations is further shown when we examined the users' textual explanations. It is clear that users are actively appraising their actions in text rather than simply describing the activities they engaged in. This active appraisal also explains the increase in predictive power that results from including textual features; suggesting users are more nuanced in text compared with a simple appraisal using an integer scale. Active evaluation provides essential information for accurate mood modeling.

These results have important general implications for the design of new technologies to detect and model mood. Our results argue for the importance of designs that encourage users to actively reflect and appraise the effects of activities on mood. However this design recommendation runs counter to proposals for new systems that include sensors and methods to *automatically* detect users' actions, e.g. physical activity, sleep, or social interaction using sensors [43, 56, 57]. A clear motivation for such automatic approaches is to reduce the burden on users to actively log their activities. In contrast, our findings suggest that simple activity detection has low explanatory value compared with active user evaluations. While passive tracking of activities is lightweight for users, we show that this approach overlooks important information contributed by a user's self-evaluation. Incorporating individual users' active weightings of the effects of activities on mood along with their textual self-evaluations increased our model's accuracy by 1657% from the R^2=.037 from binary activity models to the R^2=.613 of the full actively user weighted Activities + Explanations + Individual Differences model. Having said this, carefully reflecting and weighing what affects mood is a clear imposition on users. The demands of such active reflection may potentially reduce user compliance and willingness to use such systems. While this clearly depends on the behaviors being analyzed, system designers must carefully consider how the accuracy of their models trades off against this user burden. Perhaps clearly informing users about the benefits of active user reflection may serve to motivate this behavior.

There are other design approaches that might promote active reflection while reducing the burden on users. We might, for example, seek to reduce the number of activities that users' actively track. We have shown that Health and Social activities are the most influential factors that need to be recorded. Tracking a few items relating to Health and Social activities might be done very quickly without overloading the user. Another approach might be to focus on individual differences. Modeling on a per user basis added considerably to the models' predictive power. Perhaps the tracking and reflecting interface might be adapted on a per-user basis to include a smaller number of user-relevant activities. Again there are trade-offs here, for example if users' lives undergo major changes it may be that new untracked activities begin to have effects on mood.

Individual mood modeling is an important consideration for both theory and for the development of future intervention systems, which will of course need to be highly personalized. We found three reliably distinct clusters of users, and although including such clusters did not improve our models overall, it may be because clusters are redundant with information already present in our individual user models. However, we believe that this clustering procedure holds promise for tackling the cold start problem in similar systems.

In addition, our findings have significant implications for mental health interventions. We have already incorporated activity models into our own EmotiCal system, a novel mobile application to present personalized activity recommendations and mood forecasts. A field deployment has shown that people who use EmotiCal for a month display increased mood ratings and report greater insight into their emotions suggesting that our application overcame users' limitations in users' inherent abilities to forecast their own affect [26]. Increased well-being following the use of EmotiCal may have arisen because of increased affective awareness following active emotional reflection or because of direct changes to behavior resulting from carrying out mood boosting activities, or a combination of both. Our intervention does not enable us to disentangle these different mechanisms underpinning well-being improvements as EmotiCal led participants to do both. Furthermore, our approach extends current positive psychology application approaches. Many of these applications currently suggest thought-based exercises for mood improvement, such as gratitude exercises [39, 40]. Instead, our approach reflects Lewinsohn's behavioral approach to mood regulation [38] in recommending specific, personalized activities that are intended to increase positive mood and structuring recommendations with concrete planning to improve compliance [13]. Accurate mood modeling supports improved recommendations for mood-boosting activities, leading to measurable changes in wellbeing.

There are also important implications for social science. Our results show that we are able to accurately predict user mood from self-reported activity data. Better modeling of relations between activities and mood are critical for improved scientific understanding of health and wellbeing; in particular, in supporting effective emotion regulation. We also add to our scientific understanding about how activity influences mood using a new method. Our deployment allowed users to log exactly what they felt in the moment, giving our data fidelity with regards to how mood and activities interact.

Our findings also contribute to studies that characterize relations between activities and mood. We support findings using retrospective data [32, 40] showing that Social Activities contribute one of the largest positive effects on mood. However, unlike the Stone et al. study [32], we find that Health activities also impart large effects on mood. Nevertheless, the large positive effect of Health activities on mood is consistent with the findings of Parks et al. where users, in aggregate, found “Doing physical exercise or sports” to be the second “most important or meaningful” activity [40]. In contrast with Stone et al., we did not find that Work depressed mood [32]. These discrepancies with prior work may arise from a difference in subject populations. Stone et al. recruited all female users, from the same geographical region that largely worked as teachers, nurses or telemarketers. In contrast, our sample included both males and females from areas across the United States.

Nevertheless, there are limitations to our approach: in our procedure users choose which events to log, which may introduce logging biases. For example, users may be more likely to log positive rather than negative moods [41]. We addressed this by allowing user choice in log scheduling which results in more carefully considered entries [2]. Users could also navigate through the application to a page listing their previous entries, so it is possible they engaged in self-reflection facilitated by having these past entries.

Overall, our study presents a successful new method of modeling emotions that we deployed in the context of a successful emotional regulation system to promote wellbeing. We were able to accurately model which activities influenced mood by collecting active user logs about activities, as well as user explanations of how different activities influenced mood. Models were also improved by personalization. Users largely complied with active data entry over a three-week deployment, suggesting the viability and promise of designing new personal health systems that analyze our activities and recommend new actions that can improve health.

7. References

1. Lin JJ, Mamykina L, Lindtner S, et al (2006) Fish’n’Steps: Encouraging physical activity with an interactive computer game. In: Int. Conf. Ubiquitous Comput. Springer, pp 261–278

2. Isaacs E, Konrad A, Walendowski A, et al (2013) Echoes from the past: how technology mediated reflection improves well-being. ACM Press, p 1071

3. LiKamWa R, Liu Y, Lane ND, Zhong L (2013) Moodscope: Building a mood sensor from smartphone usage patterns. In: Proceeding 11th Annu. Int. Conf. Mob. Syst. Appl. Serv. ACM, pp 389–402

4. Barwais FA, Cuddihy TF, Tomson LM (2013) Physical activity, sedentary behavior and total wellness changes among sedentary adults: a 4-week randomized controlled trial. Health Qual Life Outcomes 11:183. doi: 10.1186/1477-7525-11-183

5. Cadmus-Bertram LA, Marcus BH, Patterson RE, et al (2015) Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. Am J Prev Med 49:414–418. doi: 10.1016/j.amepre.2015.01.020

6. Peuhkuri K, Sihvola N, Korpela R (2012) Diet promotes sleep duration and quality. Nutr Res 32:309–319. doi: 10.1016/j.nutres.2012.03.009

7. Brockton West Roxbury VA (1997) Sleep, Sleep Deprivation, and Daytime Activities A Randomized Controlled Trial of the Effect of Exercise on Sleep. Sleep 20:95–101.

8. Cacioppo JT, Hawkley LC, Berntson GG, et al (2002) Do Lonely Days Invade the Nights? Potential Social Modulation of Sleep Efficiency. Psychol Sci 13:384–387.

9. Peters E, Hibbard J, Slovic P, Dieckmann N (2007) Numeracy Skill And The Communication, Comprehension, And Use Of Risk-Benefit Information. Health Aff (Millwood) 26:741–748. doi: 10.1377/hlthaff.26.3.741

10. Fitbit Fitbit App & Dashboard. https://www.fitbit.com/app. Accessed 7 Mar 2017

11. Jawbone Jawbone | JAMBOX Wireless Speakers | UP Wristband | Bluetooth Headsets. In: Jawbone. https://jawbone.com/support/articles/000005227/understanding-your-data. Accessed 7 Mar 2017

12. Bentley F, Tollmar K, Stephenson P, et al (2013) Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change. ACM Trans Comput-Hum Interact 20:1–27. doi: 10.1145/2503823

13. Gollwitzer PM (1999) Implementation intentions: Strong effects of simple plans. Am Psychol 54:493.

14. Gilbert DT, Pinel EC, Wilson TD, et al (1998) Immune neglect: a source of durability bias in affective forecasting. J Pers Soc Psychol 75:617.

15. Tice DM, Bratslavsky E, Baumeister RF (2001) Emotional distress regulation takes precedence over impulse control: If you feel bad, do it! J Pers Soc Psychol 80:53–67. doi: 10.1037//0022-3514.80.1.53

16. Konrad A, Tucker S, Crane J, Whittaker S (2016) Technology and Reflection: Mood and Memory Mechanisms for Well-Being. Psychol Well-Being. doi: 10.1186/s13612-016-0045-3

17. Watkins PC, Mathews A, Williamson DA, Fuller RD (1992) Mood-congruent memory in depression: Emotional priming or elaboration? J Abnorm Psychol 101:581.

18. Cuijpers P, van Straten A, Warmerdam L (2007) Behavioral activation treatments of depression: A meta-analysis. Clin Psychol Rev 27:318–326. doi: 10.1016/j.cpr.2006.11.001

19. Chung C, Pennebaker JW (2007) The psychological functions of function words. Soc Commun 343–359.

20. Turner R, Ward M, Turner D (1979) Behavioral Treatments for Depression: An Evaluation of Their Therapeutic Components. J Clin Psychol 35:167–175.

21. Bardram JE, Frost M, Szántó K, Marcu G (2012) The MONARCA self-assessment system: a persuasive personal monitoring system for bipolar patients. In: Proc. 2nd ACM SIGHIT Int. Health Inform. Symp. ACM, pp 21–30

22. Dobson, Joffe THE ROLE OF ACTIVITY LEVEL AND COGNITION IN DEPRESSED MOOD IN A UNIVERSITY SAMPLE.

23. Russell JA, Pratt G (1980) A description of the affective quality attributed to environments. J Pers Soc Psychol 38:311–322. doi: 10.1037/0022-3514.38.2.311

24. Russell JA (1991) Culture and the categorization of emotions. Psychol Bull 110:426–450. doi: http://dx.doi.org/10.1037/0033-2909.110.3.426

25. Deci EL, Ryan RM (2000) The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. Psychol Inq 11:227–268. doi: 10.1207/S15327965PLI1104_01

26. Hollis V, Konrad A, Springer A, et al (2017) What Does All This Data Mean for My Future Mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being. Human–Computer Interact. doi: 10.1080/07370024.2016.1277724

27. Lane ND, Lin M, Mohammod M, et al (2014) BeWell: Sensing Sleep, Physical Activities and Social Interactions to Promote Wellbeing. Mob Netw Appl 19:345–359. doi: 10.1007/s11036-013-0484-5

28. Larsen RJ (2000) Toward a Science of Mood Regulation. Psychol Inq 11:129–141. doi: 10.1207/S15327965PLI1103_01

29. Bradley MM, Greenwald MK, Petry MC, Lang PJ (1992) Remembering pictures: Pleasure and arousal in memory. J Exp Psychol Learn Mem Cogn 18:379–390. doi: http://dx.doi.org/10.1037/0278-7393.18.2.379

30. Plutchik R (2001) The Nature of Emotions. Am Sci 89:344–350.

31. Tellegen A (1985) Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In: Tuma AH, Maser JD (eds) Anxiety Anxiety Disord. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US, pp 681–706

32. Stone AA, Schwartz JE, Schkade D, et al (2006) A population approach to the study of emotion: Diurnal rhythms of a working day examined with the day reconstruction method. Emotion 6:139–149. doi: 10.1037/1528-3542.6.1.139

33. Wilson T, Laser P, Stone J Judging the predictors of one's own mood: Accuracy and the use of shared theories. J Exp Soc Psychol 18:537–556.

34. MacPhillamy D, Lewinsohn P (1982) The Pleasant Events Schedule: Studies on Reliability, Validity, and Scale Intercorrelation. J Consult Clin Psychol 50:363–380.

35. Lewinsohn PM, Amenson CS (1978) Some relations between pleasant and unpleasant mood-related events and depression. J Abnorm Psychol 87:644–654. doi: 10.1037/0021-843X.87.6.644

36. Robinson MD, Clore GL (2002) Belief and feeling: Evidence for an accessibility model of emotional self-report. Psychol Bull 128:934–960. doi: 10.1037//0033-2909.128.6.934

37. Walker WR, Skowronski JJ, Thompson CP (2003) Life is pleasant--and memory helps to keep it that way! Rev Gen Psychol 7:203–210. doi: 10.1037/1089-2680.7.2.203

38. Zeiss AM, Lewinsohn PM, Muñoz RF (1979) Nonspecific improvement effects in depression using interpersonal skills training, pleasant activity schedules, or cognitive training. J Consult Clin Psychol 47:427.

39. Seligman MEP, Steen TA, Park N, Peterson C (2005) Positive Psychology Progress: Empirical Validation of Interventions. Am Psychol 60:410–421. doi: 10.1037/0003-066X.60.5.410

40. Parks AC, Della Porta MD, Pierce RS, et al (2012) Pursuing happiness in everyday life: The characteristics and behaviors of online happiness seekers. Emotion 12:1222–1234. doi: 10.1037/a0028587

41. Hollis V, Konrad A, Whittaker S (2015) Change of Heart: Emotion Tracking to Promote Behavior Change. ACM Press, pp 2643–2652

42. Peesapati ST, Schwanda V, Schultz J, et al (2010) Pensieve: Supporting Everyday Reminiscence. In: Proc. SIGCHI Conf. Hum. Factors Comput. Syst. ACM, New York, NY, USA, pp 2027–2036

43. Doryab A, Frost M, Faurholt-Jepsen M, et al (2015) Impact factor analysis: combining prediction with parameter ranking to reveal the impact of behavior on health outcome. Pers Ubiquitous Comput 19:355–365. doi: 10.1007/s00779-014-0826-8

44. Faurholt-Jepsen M, Vinberg M, Christensen EM, et al (2013) Daily electronic self-monitoring of subjective and objective symptoms in bipolar disorder--the MONARCA trial protocol (MONitoring, treAtment and pRediCtion of bipolAr disorder episodes): a randomised controlled single-blind trial. BMJ Open 3:e003353–e003353. doi: 10.1136/bmjopen-2013-003353 1.

45. Bradley MM, Lang PJ (1994) Measuring emotion: the self-assessment manikin and the semantic differential. Journal of behavior therapy and experimental psychiatry 25:49–59.

46. statsmodels.

47. Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825−2830.

48. Pennebaker J, Booth R, Francis M (2007) Linguistic Inquiry and word count: LIWC [Computer Software]. Austin, TX

49. Balahur A, Hermida JM, Montoyo A (2012) Building and Exploiting EmotiNet, a Knowledge Base for Emotion Detection Based on the Appraisal Theory Model. IEEE Trans Affect Comput 3:88–101. doi: 10.1109/T-AFFC.2011.33

50. Chen Y-W, Lin C-J (2006) Combining SVMs with various feature selection strategies. In: Feature Extr. Springer, pp 315–324

51. Goyal A, Riloff E, Daumé III H (2010) Automatically producing plot unit representations for narrative text. In: Proc. 2010 Conf. Empir. Methods Nat. Lang. Process. Association for Computational Linguistics, pp 77–86

52. Schein AI, Popescul A, Ungar LH, Pennock DM (2002) Methods and metrics for cold-start recommendations. ACM Press, p 253

53. Jackson DA (1993) Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. Ecology 74:2204–2214. doi: 10.2307/1939574

54. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 20:53–65.

55. Liu S, Xie Y, Mcgree J, Ge Z (2016) Computational and statistical methods for analysing big data with applications. In: CERN Doc. Serv. http://cds.cern.ch/record/2204430. Accessed 13 Apr 2017

56. Wang R, Chen F, Chen Z, et al (2014) StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. ACM Press, pp 3–14

57. Zisook M, Taylor S, Sano A, Picard R (2016) SNAPSHOT Expose: Stage Based and Social Theory Based Applications to Reduce Stress and Improve Wellbeing.