Title
On Being ‘Fitter, Happier, and More Productive’: The Impact of Implicit Goals in Affective Personal Informatics

Author
Hollis, Victoria

Publication Date
2018

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ON BEING 'FITTER, HAPPIER, AND MORE PRODUCTIVE': THE IMPACT OF IMPLICIT GOALS IN AFFECTIVE PERSONAL INFORMATICS

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Victoria Hollis

December 2018

The Dissertation of Victoria Hollis is approved:

Professor Steve Whittaker, chair

Professor Doug Bonett

Professor Jean E. Fox Tree

Professor Leila Takayama

Lori Kletzer
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures	**v**
List of Tables	**vii**
Abstract	**viii**
Acknowledgments	**x**
Introduction	**1**
Theoretical Background and Related Works	**4**
Research Questions	**17**
Research Approach	**18**
Study 1: Stress Tracking	**20**
Study 1: Hypotheses	21
The Stress Tracking System	22
Measures	24
Timed Problem-Solving Test	30
Conditions	31
Procedure	36
Dataset and Analysis	37
Study 1: Stress Tracking Pilot Results	38
Study 1: Stress Tracking Main Results	**43**
Study 1 Discussion	**52**
Study 2: Impact of Positivity Goals on Emotional Well-being	**53**
Study 2: Hypotheses	56
The Mood Tracking System	57
Participant Recruitment	60
Measures	61
Conditions	67
Procedure	71
Dataset and Analysis	72
Study 2: Onboarding Materials Pilot	74

Study 2: Mood Tracking Results **75**

Study 2 Discussion **96**

General Discussion **97**

Appendices **110**

References **130**

List of Figures

Figure 1. Examples of Affective Personal Informatics (PI) technologies. 2

Figure 2. Survey results from (n=188) respondents on the self-relevant data types respondents want automatically collected by a personal informatics system. 7

Figure 3. The EmVibe system. 23

Figure 4. Retrospective data viewing option for the stress tracking system. 24

Figure 5. Example math word problems from the exam. 31

Figure 6. Improvement condition for an automatic stress tracking system. 32

Figure 7. Self-Efficacy condition for an automatic stress tracking system. 33

Figure 8. Self-Knowledge condition for an automatic stress tracking system. 35

Figure 9. Design of Study 1. 37

Figure 10. Average differences in pre-post scores for the STAI-State scale. 45

Figure 11. Adoption Interest and Data Interest for Study 2. 48

Figure 12. Average ratings of Perceived Success. 49

Figure 13. Human rated valence of participant free writes for Study 1. 51

Figure 14. Mood Tracking System Components. 58

Figure 15. Example app landing page for the Improvement condition. 59

Figure 16. Example app landing page for the Self-Knowledge condition. 60

Figure 17. Example of onboarding materials for the Improvement condition. 68

Figure 18. Example of onboarding materials for the Self-Knowledge condition. 70

Figure 19. Design of Study 2. 72

Figure 20. Average differences in well-being scales. 77

Figure 21. Average pre-post scores for the Values Happiness scale. 78

Figure 22. Reported engagement and avoidance with data visualizations. 80

Figure 23. Ratings of Adoption Interest and Continued Use. 80

Figure 24. Average word count and character length of text produced per entry. 82

Figure 25. Perceived mood prediction accuracy. 82

Figure 26. Ratings of Perceived Success for Study 2. 84

Figure 27. Human rated valence of participant free writes for Study 2. 86

Figure 28. Best uses for an emotion-tracking system. 93

List of Tables

Table 1. Results from a pilot with 100 survey respondents. 41

Table 2. Results of emotion outcomes for Study 1 from an in-lab pilot. 42

Table 3. Results of hypothetical intentions to use the device, view stress data, and perceived sensor accuracy. 42

Table 4. Emotion outcomes for Study 1. 46

Table 5. Means and Standard Deviations of pre-post Beliefs about Stress for Negative, Positive, and Controllability stress beliefs. 46

Table 6. Adoption Interest and Data Interest for Study 1. 48

Table 7. Examples of text free writes from Study 1. 52

Table 8. Online pilot for onboarding materials for Study 2. 75

Table 9. Examples of text free writes from Study 2. 87

Abstract

On Being 'Fitter, Happier, and More Productive': The Impact of Implicit Goals in Affective Personal Informatics

by

Victoria Hollis

Personal informatics (PI) technologies allow unprecedented opportunities to track and analyze complex data about ourselves. However, a concern is that these technologies can make normative assumptions about user goals and ideal outcomes. Such assumptions could be especially problematic for Affective PI, as there is a risk that technologies which reflect implicit goals for users be more positive or reduce stress could ironically decrease well-being (Mauss et al., 2011). Furthermore, users could actively avoid PI data if they feel unable to meet the demands of the system (Duval & Wicklund, 1972), running counter to the view that users will engage data for beneficial insights (Kersten-van Dijk et al., 2017). We tested whether Affective PI systems that reflect goals for particular emotion outcomes (Improvement) have counterproductive effects for well-being and user engagement. These outcomes were contrasted against systems that instead reflect goals for Self-Knowledge, a top user interest (Hollis et al., 2018). Study 1 examined the effects of implicit goals in the context of an automatic stress detection and feedback system used during an exam. Participants viewed instructions that either describes the system goal as stress reduction (Improvement), stress reduction with a relaxation strategy (Self-Efficacy), accurate self-knowledge (Self-Knowledge), or saw no system goal (Control). Study 2 was a 21-day field study during which participants used a manual emotion-tracking web app that either emphasized a goal of increased positivity (Improvement), a goal of accurate self-knowledge (Self-Knowledge), or only completed pre-post surveys (Control). For each study, participants completed measures of well-being and engagement with the experimental systems. Across both studies, there were no significant condition differences in well-being. However, participants in the Self-Knowledge conditions of both studies considered themselves significantly more successful at achieving the system goals. As a result, Self-Knowledge participants were also more engaged with the stress- and emotion-tracking systems. Unlike prior work showing the ironic effects of emotional positivity goals, we show such negative impacts do not occur in this real-world context. We discuss these results with design implications for self-tracking systems and deepen the theoretical understandings of how users engage with PI.

Acknowledgments

The acknowledgments section for an effort like this can be one of the hardest to write. I am indebted to a countless number of individuals who have helped me grow as a researcher, or otherwise keep me sane enough to pursue something as demanding, yet deeply rewarding, as a doctoral degree. This is as much my achievement as yours.

This thesis and all other research I've conducted over the past five years would not have been possible without the continuous love and support of my family: Elizabeth, Bob, and Tiffany Hollis. Thank you for reminding me of what is important in life. You mean the world to me.

To my partner, Matt: you have been by my side since the beginning of this wild journey. Thank you for the many nights you spent listening to practice talks, arguing about tech, or stealing me away from work to experience life. You've been an unwavering source of support, and the impact you've had on me is immeasurable. I could not imagine having a more loving, humorous, or brilliant partner by my side.

A massive thank you to my qualifying exam and dissertation committee members: Leila Takayama, Jean E. Fox Tree, Doug Bonett, Katherine Isbister, and Steve Whittaker. Doug, thank you for responding to my numerous questions over the years for statistical advice and encouraging the best practices for quantitative analyses. Jeannie, thank you for continuing to be an advocate for students and for selflessly sharing your time and expertise. Katherine, thank you for encouraging big picture questions for how academic research can impact people's lives. Leila, thank

you for your ongoing mentorship. Your humility, compassion, and excitement for research are contagious. You have positively impacted numerous student lives and I am lucky to know you.

To my advisor, Steve Whittaker: You have played a dramatic role in shaping my future for the better. Thank you for affording exciting research opportunities when I was an undergraduate and for your continued encouragement to pursue a doctoral degree with you. This dissertation and many other research efforts over the past five years benefitted from our countless discussions and brainstorming sessions. Thank you for teaching me the art of writing academic papers and presenting scientific findings to a diversity of audiences. Thank you for fostering a human-computer interaction community at UC Santa Cruz. And thank you for your continued support in pursuing research opportunities in graduate school and beyond. I am deeply grateful for the opportunity to have learned from you.

To my former and current labmates: Artie Konrad, Jeff Warshaw, Charlotte Massey, Ryan Compton, Aaron Springer, Lee Taber, and Joel Schooler. Thank you for the comradery and for providing a fun space to share research ideas. You made the journey through grad school just a little less scary and more relatable.

Research in human-computer interaction can be tricky, with requirements to build technologies that are reliable and comparable to commercial tools. We have been extremely fortunate to work with talented collaborators. Thank you to the team who made crucial contributions to the emotion-tracking web app: Rob Martin, Matt Antoun, Chris Antoun, and Aaron Springer. Several other studies were possible

through the efforts of talented research assistants who went above and beyond to develop research platforms. Thank you to Alon Pekurovsky and Nika Wu for your fantastic problem-solving and foundational review of technical options. Thank you to Rishita Roy, Cindy Tiet, and Huy Ngo for your development of data visualizations for various applications.

Several research assistants have also dedicated considerable time and effort to assist with data collection. A special thanks to Mandi Martin, Bonni Baichon, Angelica Lau, Eric Kellenberger, Erika Lin, Andrea Limon, and Millie Angulo. We were fortunate to work with undergraduate research assistants who could not only run participants in-lab, but also learn new technologies, be attentive to participant well-being, and conduct thoughtful interviews. Being able to trust you will treat participants with respect, but in a scientifically sound way, is a foundational component of research. I am indebted to you and will always be available to help you grow as researchers.

Finally, this thesis and all other prior research from my Ph.D. involves participants who are willing to share their time and personal aspects of their lives with us. None of this could have been possible without those participants. Thank you for your patience to try new technologies.

Introduction

Technologies to monitor one's self are pervasive. As of 2015, 2.6 billion individuals own smartphones, and 1 in 6 Americans own a dedicated fitness tracker or smartwatch (Nielsen, 2014). These new hardware capabilities are further supported by many of the 165,000+ mobile wellbeing apps available across Google Play and the iOS App Store (Aitken & Lyle, 2015). This rise in self-tracking capabilities is coupled with the emergence of communities (notably Quantified Self) who claim that rigorous data-centric personal monitoring enables people to better understand and perfect themselves. Furthermore, we see new lines of exciting research unfolding around the practices, tools, and consequences of these self-tracking technologies.

I refer to these self-tracking systems as Personal Informatics (PI), tools which support a wide range of methods to collect personal data and present this information back to individuals (Li et al., 2010). These systems can be manual, supporting self-reports of data such as emotional states or fatigue. PI systems can also automatically collect data, such as steps for exercise, or use of time for a measure of personal productivity. This variety of personal data can be presented back in several possible ways such as ambient displays (Snyder et al., 2015; Whittaker et al., 2016), real-time haptic feedback (Hollis et al., 2018), or digital records for later reflection (Hollis et al., 2015; Hollis et al., 2017; Isaacs et al., 2013). For example PI systems which demonstrate many of these features, see Figure 1.

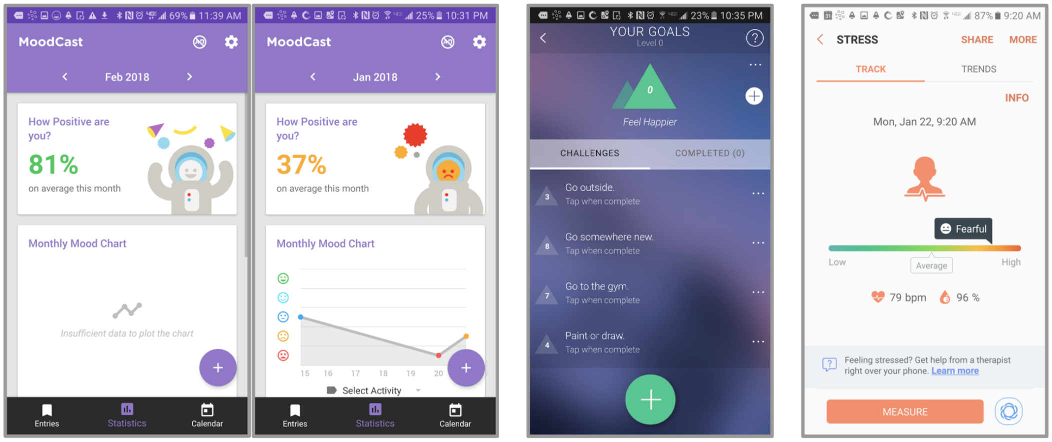


Figure 1. Examples of Affective Personal Informatics (PI) technologies. The leftmost image is the Moodcast application which encourages positivity. The center application is Pacifica which encourages personal mood and behavior goal setting. The rightmost image is a Heart Rate Variability (HRV) stress tracking function in the Samsung Health application.

It should be immediately apparent that these technologies provoke significant psychological questions that I address in this thesis. First, the prevailing discussions of PI make questionable assumptions for how individuals will interact with these systems, namely that users rationally engage data and that this promotes accurate insight and subsequent behavior change (i.e., the Self-Improvement Hypothesis, Kersten-van Dijk et al., 2017). However, not only do motives to adopt PI systems vary beyond pure self-improvement (Rooksby et al., 2014; Hollis et al., 2018), there may even be downfalls to set expectations that one will directly improve following exposure to data. As prior psychological work has shown, priming users with an expectation to improve mood and feel more positive can have paradoxical effects for one’s well-being (e.g., Ford and Mauss, 2014; Gruber et al., 2011; Mauss et al., 2011; Schooler et al., 2003). We have similarly seen in our own research participant concerns of being ‘pressured’ by systems to not feel a certain way, such as with stress tracking (Hollis et al., 2018).

Second, complementing this focus on improvement, there has been very little discussion surrounding the implicit goals embedded within these PI systems that one is expected to strive toward. Recent work has raised concerns about the prevailing goals and ideals assumed by these systems (Spiel et al., 2018). For example, that menstruation tracking apps assume a woman wants to conceive (Epstein et al., 2017), sexual activity tracking apps convey standards around an ideal sex life (Lupton, 2014), or mood tracking apps promote positivity which contrasts with mood stabilizing goals for those with bipolar disorder (Matthews et al., 2017).

Beyond consequences for well-being, there are important questions for how the goals reflected in PI systems can impact one’s willingness to engage with data. If one has failed to meet a personal or external standard, prior work suggests that people will strive to reduce the self-to-standard discrepancy, or alternatively avoid information about the self (e.g., Greenberg & Musham, 1981). Furthermore, psychological research has shown that feedback can be counterproductive when it is more evaluative of the self (Kluger & DeNisi, 1996). Real-world instances of this information avoidance of data have been reported by fitness tracker users (Sjöklint et al., 2015) and other cases of self-tracking systems for health (Elsden et al., 2014). However, it is possible to counter these motivational biases to avoid self-relevant information (Carlson, 2013) and such engagement is critical for a key tenet of the Personal Informatics vision: that people engage data for accurate self-insight. As a secondary set of questions, this thesis tested whether the presence of goals for emotion alters data engagement and interest to adopt an affective PI system.

These questions around the impact of implicit goals on well-being and data engagement were addressed within two experimental studies. The first was an in-lab experiment to test the effects of different implicit system goals on self-reported stress and engagement for an automatic stress tracker. The second explored the impact of different implicit system goals in a longitudinal field deployment of a mobile app to manually record mood. In each study, we directly contrasted systems which reflect Improvement goals to achieve particular emotion outcomes against an alternative system design that instead encourages accurate Self-Knowledge, a known primary motivator to use affective PI systems (Hollis et al., 2018).

Theoretical Background and Related Works

Personal Informatics (PI): Recent advances in computing have resulted in a surge of specialized devices to monitor physiological data and activities. Approximately 70% of US citizens track some form of health information and 20% use technology to do so (Fox & Duggan, 2013). Even if one does not actively engage in manual tracking, data from existing systems can be repurposed to quantify our lives (i.e., digital phenotyping, Onnela & Rauch, 2016). For example, it's possible to estimate driver stress from GPS traces of a mobile phone (Vhaduri et al., 2014), step count from smartphone accelerometers (Samsung S Health), or emotional states over time from twitter feeds (De Choudhury et al., 2013; Dodds et al., 2011; Golbeck, 2016).

These computing advances have led to the emergence of the Quantified Self, a movement which advocates for "*self knowledge through numbers*" by using technology to track and interpret personal data. The clear appeal of self-tracking is that this practice converts previously hidden (e.g., heart rate) or forgettable information (e.g., mood) into concrete records that can provide insight and self-awareness (Ruckenstein, 2014). This process is captured by the Self-Improvement Hypothesi*s* of PI which suggests that users obtain accurate insight from self-tracking and use these insights to change behavior (Kersten-van Dijk et al., 2017). Furthermore, we see this hypothesis underlies much of the foundational research around personal informatics use (e.g., Li et al., 2010; Li et al., 2011). This highlights an important assumption of many self-tracking technologies, which is that these tools are believed to generate accurate self-knowledge and successfully motivate behavior change.

Motivations to Adopt PI Systems: Prior work on personal informatics has addressed underlying motivations to use PI systems. In a seminal paper, Li et al. (2011) discuss two motivations of personal data. In directive tracking, goals are pre-identified and PI is used to track those goals to motivate behavior change (e.g., tracking steps for fitness goals). In diagnostic tracking, PI logging is exploratory, e.g., to identify goals or ways to achieve goals.

While improvement and goal-attainment is the dominant paradigm in discussing tools for self-tracking (e.g., Choe et al., 2014; Li et al., 2010; Li et al., 2011), there are other motivations to self-track that are not primarily for specific goal attainment. For instance, women who use menstrual tracking applications may not aim directly for behavior change but rather to understand their bodies and make sense of their daily lives (Epstein et al., 2017) or that tracking can be used to inform personal life narratives (Rooksby et al., 2014; Elsden et al., 2014). Recent work has highlighted other use cases of PI technologies (Elsden et al., 2014; Sjöklint et al., 2015; Rooksby et al., 2014) drawing attention to experiential aspects of tracking, and pointing out that tracking is often related to larger personal projects, and being tied to self-esteem and self-image. Despite the rhetoric of quantification, serious self-trackers are often emotionally invested in their tracking, expressing pride in their achievements and disappointment at their failures (Elsden et al., 2014; Sjöklint et al., 2015). The experiential benefits of tracking are also apparent in work that explores PI for medical conditions. For example, work on self-tracking in bipolar disorder and multiple sclerosis (Ayobi et al., 2017) have shown how such practices can promote feelings of control over a worrying health condition.

My prior work has shown that motivations to track the self can go beyond strictly narrow goal-striving (Hollis et al., 2018), showing that motivations to hypothetically adopt a PI system differ considerably by the type of data participants wanted to track (e.g., weight tracking versus stress). In Hollis et al. (2018), 188 survey respondents selected up to three types of data that they would want technology to automatically track from a list of 14 possible options (e.g., calories burned, stress, productivity). For each type of data selected, they could select one of ten possible *primary* motivations for why they would want to adopt that PI system. These primary motivations expanded upon Rooksby et al.'s (2014) categorization of 5 motives to use PI systems (e.g., to document one's life, to improve one's self, to try new technology, etc.).

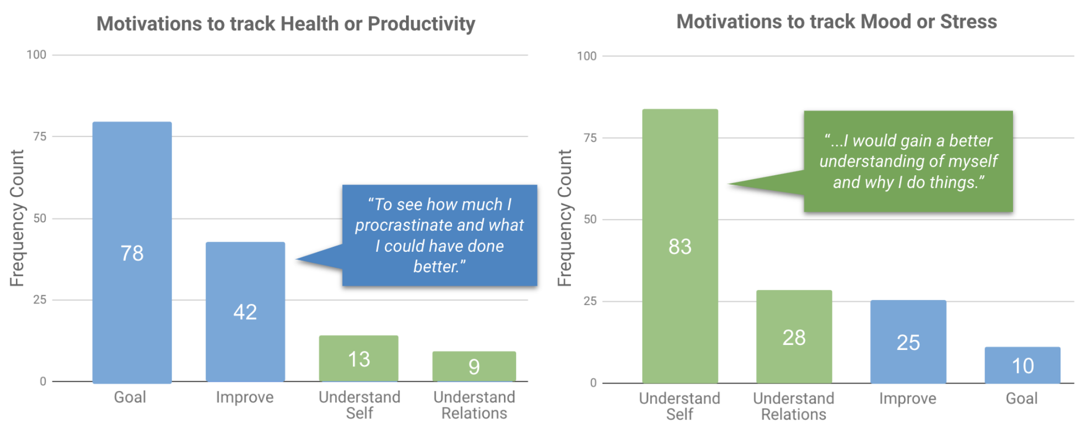


Figure 2. Survey results from (n=188) respondents on the self-relevant data types respondents want automatically collected by a personal informatics system. Each respondent could select up to 3 choices and identify a primary motivation for each. Motivations to track stress or mood were primarily about self- understanding, while motivations for health or productivity were for achieving specific goals.

In support of prior work (Choe et al., 2014; Gimpel et al., 2013; Epstein et al., 2015), adoption intentions for health- and productivity-related PI systems were found to be primarily for goal-attainment and self-improvement. For example, top motivations were either to 'achieve a goal' or 'improve in this area, without a specific goal' for data such as 'use of time' (79.4%), 'calories burned' (63.2%), or ‘dehydration’ (65.9%). However, top primary motivations to adopt stress (42.4%) or emotion trackers (63.3%) were instead mainly to understand oneself better (e.g., “*Sometimes i cannot tell even my own emotions...*”) or understand how this factor related to others. One possibility, as some survey respondents describe, is that self-understanding is a prerequisite to improvement (e.g., “*The more I understand myself, the better I can become with controlling my stress and mood*.”). Overall, these findings and other lines of research depart from the popular belief that self-tracking is strictly for behavior change goals (Li et al., 2011; Kersten-van Dijk et al., 2017) and presents a more complex landscape for people choose to quantify their lives (e.g., Rooksby et al., 2014).

Reflection of Ideals and Impact on Well-being: Beyond the issue of reflecting different primary motivations to track, systems also can present different underlying values and expectations for what goals one should strive towards (Spiel et al., 2018). Fitness trackers often assume users want to lose weight, being a potential danger to those recovering from eating disorders (Simpson & Mazzeo, 2017). Mood tracking applications can also similarly reflect cultural ideals, such as Western views of positivity, which can clash against different cultural norms (Oishi, 2002), or aspects of a destabilizing mood disorder (e.g., Matthews et al., 2017).

A wealth of prior psychological research has shown particularly counterproductive effects of setting emotion standards which makes the problem of goals embedded within affective PI systems especially important. For example, primes to encourage positivity correspond to lower reports of positive affect during a viewing of happy films (Mauss et al., 2011). Other strong lines of research have shown that trait and state effects of setting standards for increased positivity have counterproductive effects for emotional well-being (Ford & Mauss, 2014; Gruber et al., 2011).

Furthermore, we see similar effects for stress in that those who have more negative beliefs about stress, including beliefs they should reduce stress, have greater reports of stress during final exam periods in an academic school year (Fischer et al., 2016). The pursuit of ideal emotional states is not entirely futile though as simple attentional shifts to focus on specific actions rather than emotion outcomes can result in the experience of more positive emotional states (Catalino et al., 2014; Goldsmith et al., 2013).

These effects are complicated by mixed evidence on the effectiveness of digital emotion monitoring. A daily email reminder to orient one to how happy they are ("*How happy were you today?*") resulted in no improvement in reported happiness (Goldsmith et al., 2013) and our own work has shown that simple daily tracking of mood does not correspond to increases in positive affect (Hollis et al., 2017). In some cases, tools for emotion-oriented tracking are counterproductive to well-being. Faurholt-Jepsen et al. (2015) deployed the MONARCA mood tracking application for participants with bipolar disorder and found that users of the monitoring-only version of the application had greater post-test depressive symptoms compared to a control group. Similarly, Conner & Reid (2012) found that participants with depression who recorded their mood daily via SMS showed a decline in daily mood reports throughout the study period, relative to controls. In the case of a system to automatically detect and visualize stress, participants who received stress feedback had greater reports for subjective appraisals of stress relative to a no-feedback group (MacLean et al., 2013).

Theories of Self Awareness and Regulation: There are multiple theories of self-regulation which can potentially explain this paradoxical effect of emotion goals and monitoring (e.g., Objective Self-Awareness Theory, Duval & Wicklund, 1972; Control Theory: Carver & Scheier, 1982; Encoding Theory: Hull & Levy, 1979; Multi-Level Theory: Gibbons, 1990; for a review, see Kluger & DeNisi, 1996).

One theory of self-regulation, Objective Self-Awareness Theory (OSA) describes the process of focusing attention towards one's self, to make comparisons against a particular standard, and subsequent possible efforts to close a self-to-standard discrepancy arising from this comparison (Duval & Wicklund, 1972; Silvia & Duval, 2001). OSA theory argues that increased focus on the self leads individuals to [a] compare their current self to some standard (internal or external), [b] respond with some emotional reaction and, if a negative discrepancy is found, [c] alleviate this discrepancy by changing their own behavior, altering the standard or decreasing self-awareness. OSA is largely synonymous with 'self-focused attention' or 'self-consciousness' and can stem from a variety of possible situational or dispositional factors.

OSA theory predicts such heightened self-awareness promotes observable differences in behavior, emotional reactions, or awareness-dampening responses

(Duval & Wicklund, 1972; Scheier & Carver, 1977). For example, an individual who has a goal to do well in an academic course may receive self-relevant feedback (quiz grade of a 'C'), which is lower than their goal grade for the course (a 'B'), and as a result feel a negative emotional consequence (e.g., disappointment). That individual can engage in a variety of ways to manage the self-to-standard discrepancy such as by avoiding similar classes in the future, avoiding feedback on future tests, modifying their goal grade they intend to receive ('a C is fine'), or increase efforts to improve their grade.

However, in the domain of emotional goals, negative affect as a result of self-to-standard discrepancies is contrary to the standard itself (have a positive mood) and has been shown to result in unique counterproductive effect. In other words, those who set a standard for themselves to feel positive, are more likely to suffer from disappointment, leading to further unhappiness. This theoretical prediction is supported by psychological research which has demonstrated that trait and experimentally manipulated changes to value happiness corresponds to lower reports of positive affect (Gruber et al., 2011; Mauss et al., 2011; Schooler et al., 2003). From the perspective of design, this question is additionally important as psychological research shows that people naturally hold different personal standards for how they want to ideally feel, varying by demographic characteristics and culture (Tsai et al., 2006). As a result, assuming standard emotional goals in PI systems across a diverse range of users may be inherently problematic.

The application of OSA theory to PI is also valuable as it can provide useful predictions for engagement. A major assumption of PI advocates is that PI users act rationally with their data and engage self-relevant information to facilitate goal achievement (Sjöklint et al., 2015; Li et al., 2011; Kersten-van Dijk et al., 2017). For example, on becoming aware from PI data that one's caloric intake is too high or exercise too minimal, one should adapt behavior to improve physical health. However, this adaptation may not occur in practice as users could instead choose to overlook data or modify their standards. Despite the apparent power of self-awareness manipulations to modify behavior, previous OSA research also shows that individuals can exercise a strategic response to avoid self-relevant information or revise a particular standard they are being compared to (Duval et al., 1992; McDonald, 1980).

Greenberg & Musham (1981) tested the avoidance or engagement of self-awareness when people engaged in behaviors that were consistent or inconsistent with their beliefs about women's rights. In a control condition, participants read aloud nonsensical statements unrelated to attitudes toward women. In experimental conditions, participants instead read aloud statements that either matched or conflicted with their individual beliefs about women's rights (e.g., liberal: "*Both husband and wife should be allowed the same grounds for divorce*" or traditional: "*The intellectual leadership of a community should be largely in the hands of men.*").

Engagement with self-relevant information was shown to be significantly impacted by whether participant behavior was consistent with their personal standards. Participants viewed their reflection in a nearby mirror most often when reading statements consistent with their views of women's rights and avoided their reflection when statements and personal views were inconsistent. In addition, participants who read statements inconsistent with their beliefs gave significantly lower ratings for a desire to hear the tape recording after the study session.

These findings on self-awareness lead to crucial questions regarding the impact of implicit goals are user well-being. OSA theory predicts a range of strategies by which individuals can resolve self-to-standard discrepancies. As a result, this thesis addresses not only whether implicit goals in PI systems cause self-to-standard discrepancies, but also how such discrepancies are resolved by users. In the face of perceived failure, participants could increase effort to meet the goals of the system, discredit measurement accuracy, or avoid self-relevant information altogether. These potential outcomes lead into a following closely related set of psychological theories regarding motivational barriers to gain accurate self-knowledge.

Informational and Motivational Barriers to Self-Knowledge: Research on self-knowledge broadens the value of studying implicit goals in PI systems. Prior work, argues that accurate views of the self can be impaired by lack of access to self-relevant information (i.e., informational barriers) or motivational biases that can promote avoidance or rejection of self-relevant information (i.e., motivational barriers) (Vazire, 2010; Carlson, 2013). These motivational biases can result in behaviors to protect an existing self-concept (i.e., self-verification bia*s*) or to protect a positive self view (i.e., ego-protective bias) (Sedikides & Strube, 1997; Swann & Read, 1981; Vazire, 2010).

Informational barriers relate to one's ability to access relevant data to inform a self-view. Access to self-relevant data can be impaired by many factors, e.g., inaccurate historical views of the self due to forgetting (Robinson & Clore, 2002), limited attentional processes (Van Boven & Robinson, 2012), or reliance on incorrect sources of information to judge the self (Oishi, 2002; Robinson & Clore, 2002). Limits for how well one can access self-relevant information can also be context-specific, e.g., when one lacks attentional resources to monitor the self (Van Boven & Robinson, 2012). For example, Van Boven and Robinson (2012) showed that under high cognitive load, people are more likely to rely on gender-stereotypes to recall how they felt only moments prior. Forgetting can also impair self-knowledge regarding general beliefs about factors impacting mood. Female participants tend to rely on general beliefs about the negative impact of mood on menstruation when recalling how they felt in the prior month, yet daily participant logs showed no relationship between the two factors (McFarland et al., 1989). Wilson et al. (1982) found that even active monitoring of multiple data streams results in no benefits to self-knowledge, but prompting participants to reflect directly on the relationship between sleep and mood helps users more accurately draw correlations. These results highlight the important opportunity PI systems provide to improve access to information about the self and reflect on data for insight. However, prior work also shows that individuals can be additionally biased in how they seek out and interpret information to protect views of the self.

Motivational barriers include biases to protect an existing self-concept (i.e., self-verification bias; Swann & Read, 1981) or to protect a positive self view (i.e., ego-protective bias; Vazire, 2010). Preferences for positive or self-verifying information can influence information-seeking behaviors (Swann et al., 1992; Swann et al., 1989), attribution of events (Bradley, 1978), and accuracy in assessing one's self (Vazire, 2010). Vazire (2010) demonstrated such ego-protective biases in an experimental context in which participants evaluated themselves on personal traits such as intelligence, neuroticism, and extraversion. To serve as ground truth, participants took a series of psychological tests and were separately evaluated by peers following a group activity. Participants were significantly less accurate in appraising themselves on highly evaluative traits (intelligence) compared to less evaluative traits (extraversion) and overall were biased to appraise themselves favorably.

Motivational biases that prevent engagement with self-relevant data have clear connections to the success of PI systems and potential value these tools can offer to users. Much of the research on improving self-knowledge has focused on information access, which is consistent with the goal of many PI systems. However, there are extremely limited examples of PI interventions that have identified ways to reduce avoidance with self-relevant information. In our prior research, we have found that users of PI systems strongly prefer positive information rather than negative (Hollis et al., 2015; Hollis et al., 2017) and that users are skeptical of the accuracy of PI analytics when they represent negative information about one's self (Hollis et al., under review). Experimentally assessing the system features that amplify these barriers to engage with data is an area of key importance to PI system research. We see from qualitative studies of PI systems that such information avoidance occurs (Sjöklint et al., 2015; Elsden et al., 2014), but lack rigorous experimental research to identify factors that can make such PI data avoidance more or less likely.

In summary, PI can potentially improve self-knowledge and, in particular, reduce informational barriers by compensating for forgetting, sensing new information about the self not previously possible, and tracking complex patterns. For example, accurate PI tracking might address forgetting and provide computational support to help users understand how they might do in the future, based on current progress (Hollis et al., 2017). In principle, PI might therefore support improved self-insight and decision making (Kersten-van Dijk et al., 2017). However, there are multiple theories suggesting whether users will reference system data to better understand the self. Users can be resistant to access information about the self if they feel unable to meet a standard (Duval & Wicklund, 1972; Silvia & Duval, 2001; Greenberg & Musham, 1981), they could alternative discredit system accuracy (Arkin, & Maruyama, 1979), or completely disengage with using the system to avoid self-relevant information. These issues around engagement, paired with concerns that implicit goals can be harmful to well-being, motivate critical issues addressed in this thesis on the consequence of different underlying goals in self-tracking systems for emotion.

Research Questions

To summarize, there are three main questions addressed in the context of Study 1 and Study 2. First (RQ1): What are the emotional well-being consequences of systems that promote 'improvement' towards an emotional goal versus more accurate 'self-knowledge'? Improvement conditions involve presenting an emotion goal for participants to strive towards (e.g., increased positivity or reduced stress). In Study 1, we additionally included a second improvement condition that also supports remedial activities to regulate emotion (increasing potential self-efficacy). Well-being consequences were measured in the form of self-reports on validated measures for symptoms of stress (Study 1), the perceived frequency for experiencing different emotional states over a time period (Study 2), and general self-reports of subjective well-being (Study 2).

The second key question (RQ2) asks about a potential mechanism for RQ1. To what extent do users of an affective PI internalize the goals they are presented with? This is addressed by pre-post measures of beliefs about emotion. For example, do users of a stress tracking system that encourages 'improvement' towards reduced stress show increases in beliefs that stress should be reduced and controlled? Do users of a mood tracking application that encourages positivity consequently report valuing positivity more than those who did not interact with such a system? These measures were included to potentially explain a mechanism for any well-being effects we find for RQ1. Furthermore, participants were asked in each study to submit qualitative accounts about their experiences and what they like or dislike about the system. We could then use these accounts to see whether participants describe the implicit goals of these systems as playing a role in their experiences.

The final question regards system and data engagement (RQ3): How is engagement with the PI system impacted by the presence of different implicit goals for emotion? This question allowed us to broaden these findings to not only the immediate impact of the system on well-being, but also data engagement, interest to adopt PI systems, and PI system usage. This is a key question as there are qualitative instances of motivational biases in PI systems (Sjöklint et al., 2015; Hollis et al., 2015). This is the first line of research to address possible system designs to attenuate such potential barriers to self-knowledge. Also, as predicted by Objective Self-Awareness Theory, people may be more likely to engage with self-relevant information if they consider themselves close to or exceeding a standard (Davis & Brock, 1975; Duval, Wicklund & Fine, 1972; Greenberg & Musham, 1981). To account for this as an explanatory factor for condition differences in engagement (RQ3) participants were asked to rate the extent to which they considered themselves successful or unsuccessful at achieving the aim outlined in the application (i.e., *Perceived Success*). We could then assess whether feelings of perceived success varied by the different goals participants were exposed to.

Research Approach

To address these research questions, we conducted two separate experiments. In each study, participants were exposed to different implicit goals within an affective PI system. Across both studies, one condition was improvement-oriented and presented a standard for success (increased positivity or reduced stress). For Study 1, there was an additional improvement-oriented condition that included a remedial activity to achieve the standard present in the system (i.e., the 'self-efficacy' condition). This condition was included to test whether suggestions for remedial activities reduce the counterproductive effects of implicit system goals. Finally, another experimental condition was 'self-knowledge' oriented and communicated no specific emotion outcomes to achieve. Rather, it suggested that the goal of the system is to have accurate information about the self and that all possible emotion outcomes are acceptable.

The first study explored the impact of implicit goals in an automatic affect-monitoring system. As used in prior research (Hollis et al., 2018), we presented participants with different experimental onboarding framings of a stress-feedback system. Participants then participated in an activity designed to be challenging, while receiving wrist-based vibration feedback that they were told indicates emotional stress. Outcomes were measured for symptoms of anxiety, ratings of perceived stress, beliefs about stress, data engagement, and interest to use the PI system. The advantage of Study 1 is that we could rapidly test hypotheses within a controlled lab setting, and this additional control allowed us to collect observational measures of data engagement with first time exposure to stress data.

The other advantage of Study 1 was to broaden the potential impact of this research. Automatic stress monitoring is a top interest (CCS Insights, 2016; Hollis et al., 2018). Drawing conclusions about the impact of implicit goals in a stress monitoring application could therefore have potentially significant consequences for future PI systems in the stress-tracking space.

Study 2 explored the impact of implicit goals in an affective PI system that supports manual emotion tracking and visualized analytics of one's data. The benefit of Study 2 is that it is more ecologically valid than Study 1, allows us to triangulate study results, and we could ask different questions for this deployment. For instance, behavioral measures of daily PI system usage.

Study 1: Stress Tracking

Automatic stress monitoring is an increasingly common application which makes it a key area to study these research questions. Automatic stress monitoring is a top user interest (CCS Insights, 2016; Hollis et al., 2018) and despite recent deployments of working stress detection products (e.g., Bellabeat, Spire, or Pip) theories of self-regulation predict negative outcomes when a person has a standard to reduce stress, yet does not feel capable of doing so (Duval et al., 1992; McDonald, 1980).

Study 1 tested whether systems that convey goals for stress reduction ironically result in greater perceptions of stress. If the goal of reducing stress is de-emphasized (Self-Knowledge condition) or the stress is perceived as more controllable by presenting remedial strategies (Self-Efficacy condition) then OSA theory predicts that these counterproductive effects should be reduced.

In addition, goals embedded in these systems can impact how people react to and engage with system data. Key additional questions included whether system goals for stress reduction influence perceived stress sensor accuracy, intention to use the system (i.e., adoption interest), and interest in exploring their data (i.e., data interest). One prediction from prior psychological research is that with a self-to-standard discrepancy, users may be motivated to reject accuracy (Arkin and Maruyama, 1979) or avoid their data (e.g., Greenberg and Musham, 1981).

Study 1: Hypotheses

Hypothesis 1: Anxiety scores, subjective stress ratings, and negative-valence language will be highest in the *Improvement Frame*.

Hypothesis 2: Participants will avoid data more often and have lower interest in adopting the system in the *Improvement* and *Self-Efficacy Frames* and this will be mediated by condition differences in Perceived Success or post-task stress ratings. As predicted by OSA, with a self-to-standard discrepancy, they can be motivated to avoid information (e.g., Greenberg and Musham, 1981). If participants do not consider themselves unsuccessful, then data avoidance should not occur.

Hypothesis 3: Following the same logic of Hypothesis 2, we will also see the lowest adoption intentions for Improvement and Self-Efficacy conditions. These effects will be again explained by condition differences in Perceived Success or post-task stress ratings.

Hypothesis 4: Pre-post changes in beliefs about stress in terms of negative stress beliefs and controllability will increase in the *Improvement* and *Self-Efficacy Frames*. This will address the research question of whether participants, at least temporarily, internalize the goals presented in the system.

The Stress Tracking System

The stress tracking system (EmVibe) involves three components: a handheld sensor, smartwatch, and corresponding Android application. This system was developed through prior research and piloting (Hollis et al., 2018).

Measuring Electrodermal Activity: ElectroDermal Activity (EDA) assesses sympathetic nervous system activity by measuring the moisture level of skin (i.e., skin conductance) (Bouscein, 2012). In Study 1, EDA is measured using the commercial Pip sensor (Galvanic Ltd., 2016), which has been successfully deployed multiple times in related research (Dillon et al., 2016; Hollis et al., 2018; Matthews et al., 2015; Snyder et al., 2015). The Pip is a handheld device consisting of two gold-plated sensors, pinched between thumb and forefinger. The Pip samples EDA at 8Hz, applying a proprietary algorithm to classify EDA events as 'increased', 'decreased', or 'constant'. The output from the Pip is transmitted wirelessly over Bluetooth to an Android smartphone application that communicates increased EDA events to the smartwatch. The Android application stores a logfile for each participant session including time-stamped EDA events and raw data.

Presenting Haptic Feedback: Given participants might be motivated to avoid stress feedback *during* the main task (MacLean et al., 2013; Snyder et al., 2015), participants received haptic vibration to reduce this variance. When the Pip detects increased EDA, feedback is provided in the form of wrist-directed vibrations using a Motorola Moto 360 Sport smartwatch.

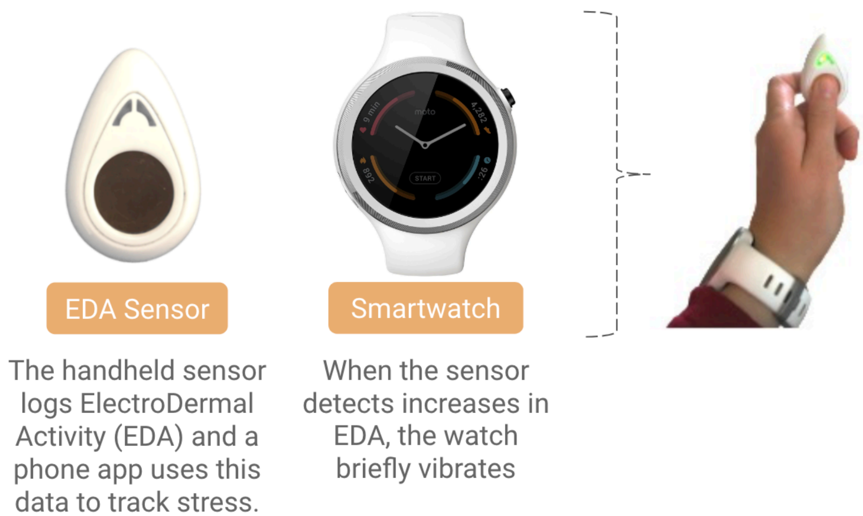


Figure 3. The pip is held between the thumb and forefinger and a smartwatch is worn on the left wrist. When the pip detects an increase in ElectroDermal Activity (EDA), the smartwatch briefly vibrates.

Retrospective Graph: Participants were given the option to explore their EDA data during the post-test survey. The EDA data was presented as a scatterplot in an Android mobile application with the x-axis as timestamps and y-axis as raw the EDA values. Points on the scatterplot were color-coded based on the classification by the Pip: ‘stressed’, ‘relaxed’, or ‘constant’.

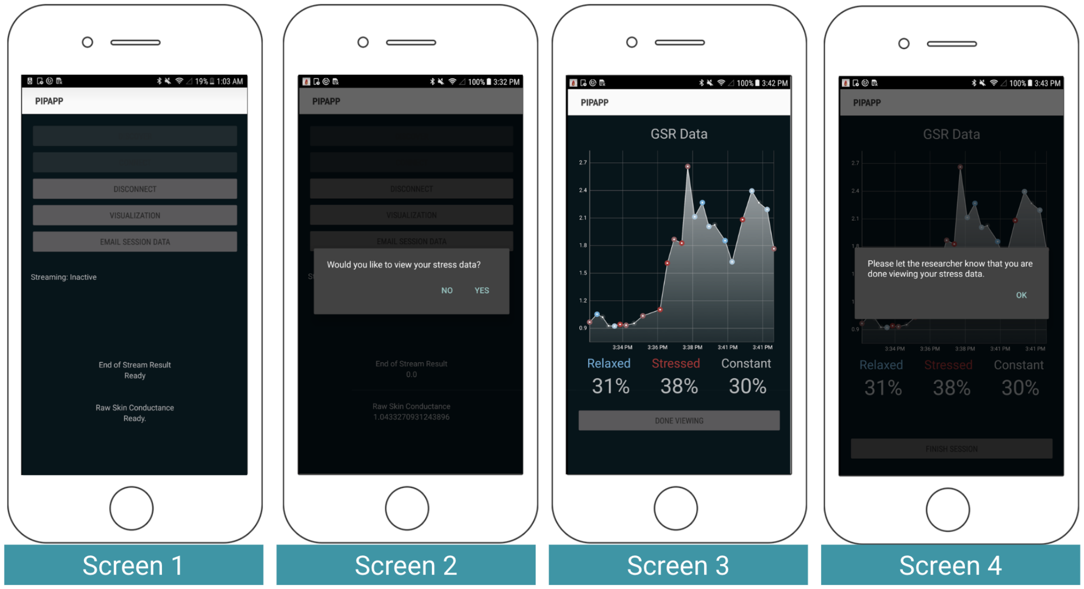


Figure 4. Retrospective data viewing option for the stress tracking system. Participants can choose to access their stress data and explore graphs of their raw data. Corresponding logs were generated from the optional data exploration step to measure how long they spent viewing their data on screen 3.

Measures

Pre-test Materials

In order to situate findings with respect to participants who might be interested in stress tracking, they were first asked to rate the extent to which they are interested in using a system to automatically measure their stress and how much they know about technologies to track mood or stress.

Participants then answered pretest surveys to assess baseline levels of anxiety (both situational and dispositional), motivation to perform well, and beliefs about stress.

Awareness of Stress Monitoring Technologies: As used in prior work (Hollis et al., 2018) and to probe for existing knowledge of stress PI systems, participants could multi-select from a list which commercial PI systems they had heard of. The list included 19 different systems with options spanning popular self-tracking technologies (e.g., Fitbit, Jawbone, MyFitnessPal, Runkeeper), and included options for commercial stress-monitoring systems (e.g., Pip). Respondents could select up to 3 aspects of themselves that they would want a tool to track automatically. These options spanned 14 possible categories including health (e.g., calories burned, time spent sitting, dehydration), work (e.g., productivity, focus time), and affective categories (e.g., stress, mood). Respondents also indicated their perceived familiarity of technologies to track mood or stress on a 5-point scale ranging from 'Not familiar at all' (1) to 'Extremely familiar' (5).

State-Trait Anxiety Inventory (STAI) Survey: The STAI for Adults short-form (Spielberger, 1983) is a 20-item scale to assess mental and physical symptoms of anxiety. Responses are given on a 4-point scale with higher scores indicating greater anxiety. The State subscale (10-items) assesses the degree to which one is experiencing *current* symptoms of anxiety (e.g., "*I am jittery*"). The Trait subscale (10-items) assesses the degree to which one experiences anxiety symptoms in general life (e.g., "*I feel nervous and restless*"). The STAI has good internal reliability with Cronbach's alpha coefficients of 0.86-0.95. It has also been used in prior work to measure the effects of false heart-rate feedback on anxiety (Costa et al., 2016).

Beliefs About Stress Scale (BASS): The BASS is a questionnaire to assess beliefs about stress in terms of negative beliefs (e.g., "*Being stressed is something I need to avoid*"), positive beliefs (e.g., "*Beings stressed makes me more productive*"), and perceived control of stress (e.g., "*Being stressed is something I am helpless to control*"). Ratings are given on a 4-point Likert scale ranging from 'completely disagree' (1) to 'definitely agree' (4). In a validation study by Laferton et al. (2016), the BASS showed good internal reliability with coefficients for the subscales ranging from .73 to .87. The original reliability and validity studies with the BASS were conducted with student populations and looked at how exams evoked subjective stress.

After reading the test instructions and corresponding system descriptions per condition, participants completed the following additional measure.

Motivation: After reading about the exam and reading system descriptions per condition, participants estimated their motivation ("*How motivated are you to perform well on this test?*") with responses on a 5-point scale from 'Not at all' (1) to 'A great deal' (5). Participants may differ in the effort they intend to apply toward the test and we wanted to account for this as a potential confound for our results.

Post-test Materials

After completing the exam, participants again completed the STAI-state scale to determine if different implicit system goals influenced symptoms of anxiety. They also completed the BASS scale to test whether they internalized the system goals they were presented with. Given OSA predicts people will disengage with information about the self if they feel unsuccessful, participants also rated perceived success at achieving the goals outlined in the system. Finally, participants answered questions assessing perceived stress sensor accuracy, recall of the test-taking experience, and interest to use the stress tracking system outside of a research setting.

Perceived Stress Ratings: Participants gave ratings for Perceived Stress in response to the question "*During the test, rate the extent to which you felt the following:*" with separate response options for "*stressed*", "*calm*", "*relaxed*", and "*tense*". Each of the four emotions was rated on a 5-point scale (1: 'Not at all' to 5: 'Extremely'). Scores for 'calm' and 'relaxed' were reverse scored and a total score across the four options was computed for a single Perceived Stress score. This scale has good to acceptable internal consistency based on Study 1 pilot results (Cronbach's $\alpha = 0.850$) and Study 1 main results (Cronbach's $\alpha = 0.785$). A similar scale has been used in prior stress research, also showing that the scale has good internal consistency (Kersten-van Dijk, 2018).

Number of Perceived Vibrations: Outcomes could be affected by variability in perceived levels of feedback. To address this potential issue, participants estimated "*Approximately how many times did you notice the band vibrate?*" with an open text option for participants to enter a number estimating the number of vibrations they experienced throughout the exam.

Test-Taking Experience Free-Writes: To understand participants' experiences and spontaneous use of emotion language, they submitted a free-write for the following question: “*Please describe your experience just now with trying this technology. What did you like or dislike?*”. This allowed us to assess qualitative differences in how participants emotionally appraised the test-taking experience based upon which system goal they were exposed to.

Accuracy of Feedback: Participants also answered a single item question to rate sensor feedback accuracy: “*How would you rate the overall accuracy of the stress detection technology you just tried?*”. Responses were given on a 5-point scale, ‘not accurate at all’ (1) to ‘extremely accurate’ (5). This was followed by a free-write for an explanation of system accuracy.

Reported Adoption Interest: Participants were asked to rate Adoption Interest: “*If you had the opportunity to wear *only* a smartwatch for stress detection (without holding a sensor), rate your interest to wear this smartwatch in your daily life:*”. Ratings were given on a 5-point scale ranging from ‘Not interested at all’ to ‘Extremely interested’. They could then provide a free-written answer to explain their Adoption Interest rating.

Reported Data Interest: Participants also indicated interest to view their stress data from the study: “*Rate your interest in seeing your stress data from this study:*” with responses given on a 5-point scale ranging from ‘Not at all interested’ (1) to ‘Extremely interested’ (5). After which, they provided a free written explanation for their answer.

Perceived Success: As predicted by Objective Self-Awareness Theory, people may be more likely to engage with self-relevant information if they consider themselves close to or exceeding a standard (Davis & Brock, 1975; Duval, Wicklund & Fine, 1972; Greenberg & Musham, 1981). To account for this as an explanatory factor for condition differences in engagement, in the post-test survey, participants in the experimental framing conditions were asked to rate on a 5-point scale the extent to which they considered themselves successful or unsuccessful at achieving the aim outlined in the application (i.e., *Perceived Success*). Ratings were given on a 5-point scale ranging from 'Not at all successful' (1) to 'Extremely successful' (5).

Length of Data Engagement: Participants were prompted within the survey whether they would like to review their stress data as a visualization. They could then view their visualized data in the Android application which produced a logfile for how long the participant viewed their data visualization.

Demographics: Participants entered their age, gender identity (e.g., Female, Male, Trans, Genderqueer, etc.), and rated their English fluency.

If participants choose to view their stress data visualization, they were also asked within the survey a set of closed- and open-ended questions about their interpretation of the data. The final questions of the post-test survey included a suspicion check question, asking participants what they thought the study was about. On the closing screen, participants were thanked, presented with a written debrief of the study, and given Victoria Hollis' contact information if they had any follow-up questions.

Timed Problem-Solving Test

The problem-solving test involved participants completing 20 math word problems from a website with practice exams for the SAT (4tests.com). A math test was chosen as previous pilots showed that these tests were considered more challenging than abstract reasoning tasks which we had used in prior work (Hollis et al., 2018). An additional pilot using the Raven's Progressive Matrices was previously conducted for the Study 1 procedure which showed that this task was not sufficiently challenging, resulting in floor effects. We are also not the first to use math questions to induce challenge, as prior work has shown significant differences in physiological measures of arousal when student samples are presented with either positively- or negatively-framed arousal reappraisal primes before taking the Graduate Record Examination (GRE) math section (Jamieson et al., 2010; Jamieson et al., 2013). In addition, while a popular method to induce stress involves a public speaking exercise (Costa et al., 2016; Hua et al., 2014), previous work has also shown that exams are the most often selected specific context in which students want automatic stress tracking (Hollis et al., 2018). A key additional benefit of using an exam here is increased potential real-world validity of findings by testing these effects in a situational context in which students may want stress tracking.

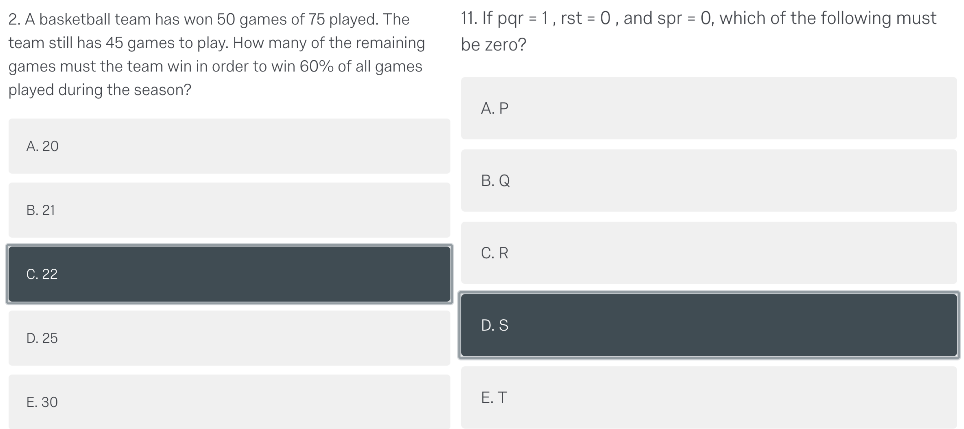


Figure 5. Example math word problems from the exam. In these cases, the correct answers are C (leftmost figure) and D (rightmost figure).

Conditions

Participants were randomly allocated to one of the following conditions. One hypothesis in this study was that the Improvement condition (C1) would result in higher stress ratings than goals which focused on Self-Knowledge (C3). However, existing systems do not simply provide improvement-focused designs and can also present suggestions for how to dampen stress. An improvement and suggested strategy condition (C2) allowed for the opportunity to compare the extent of potential ironic effects when participants are also given a recommendation for how to regulate their stress responses.

All participants first viewed an image of the hardware (see Figure 3) with the following description: "*The mobile app you will review today involves a physical sensor, smartwatch, and visualization. The handheld sensor logs ElectroDermal Activity (EDA) and the phone app uses this data to track stress. When you experience* *an increase in stress levels, you will get real-time vibration feedback from a smartwatch.*”.

Participants were then randomly allocated through the Qualtrics platform to view one of the following sets of condition materials.

Condition 1: Improvement

For the Improvement condition, participants received system explanations that convey a goal of stress reduction (see Figure 6). Participants also received a text description to reaffirm this experimental framing: “*The goal of this system is to help you decrease stress. By making you aware of stress, you can better manage it and reduce negative stress effects.*”

Figure 6. Improvement condition for an automatic stress tracking system.

Condition 2: Self-Efficacy

Participants in the Self-Efficacy condition also received system descriptions that described the system goal as stress reduction (see Figure 7) but differed from

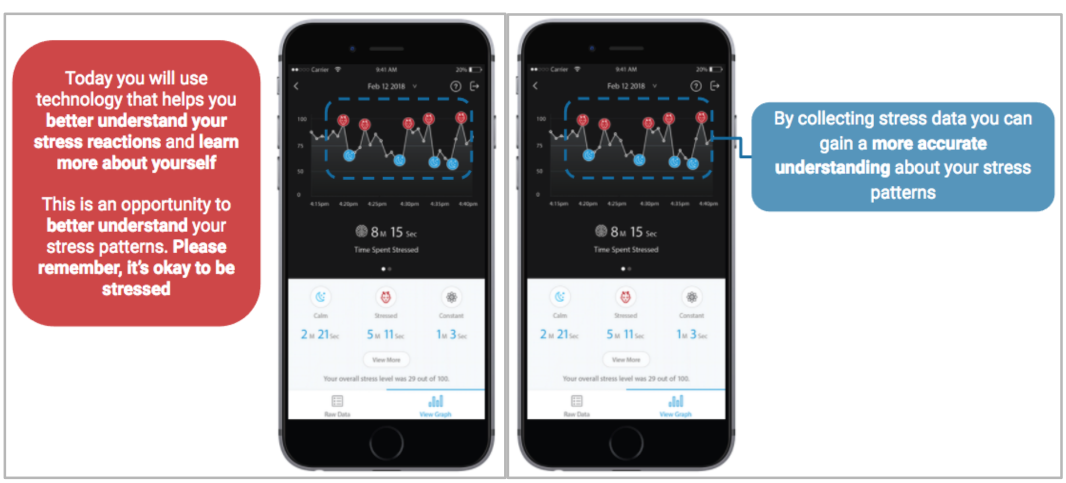
Condition 1 as it also recommended a strategy to regulate stress (Fisher & Newman, 2013). They were shown a text description on the following page to reaffirm the system goal: "*The goal of this system is to help you decrease stress in order to reduce negative stress effects. Whenever you feel a vibration, take a deep breath from your stomach rather than from your chest. Also, slow your breathing down to a rate slower than usual but not so slow that it is unpleasant or uncomfortable. You might do this by counting from one to three as you breathe in evenly and then again as you evenly exhale.*".

Figure 7. Self-Efficacy condition for an automatic stress tracking system.

Condition 3: Self-Knowledge

The Self-Knowledge condition did not emphasize a goal of stress reduction. Instead, the system description emphasized the goal as increased self-knowledge and purposefully downplayed goals for stress reduction (see Figure 8). Participants saw the following page with a text-only description to reaffirm the system goal. "*It's okay to feel stressed and you don't need to fix that! The goal of using this system is to help you learn more about yourself. This is not intended to help you decrease stress, but instead to learn more about what makes you stressed, or not. When you have access to data, you can see and understand your limits a little easier. We think this is the most central thing where technology can benefit us – to teach us and to increase our understanding of ourselves.*"

Pilots revealed that comprehension of the Self-Knowledge system description was least consistent as some participants still attributed a stress reduction goal to the system description. We included additional written content to further distance the Self-Knowledge condition from the improvement-focused conditions (1 and 2).



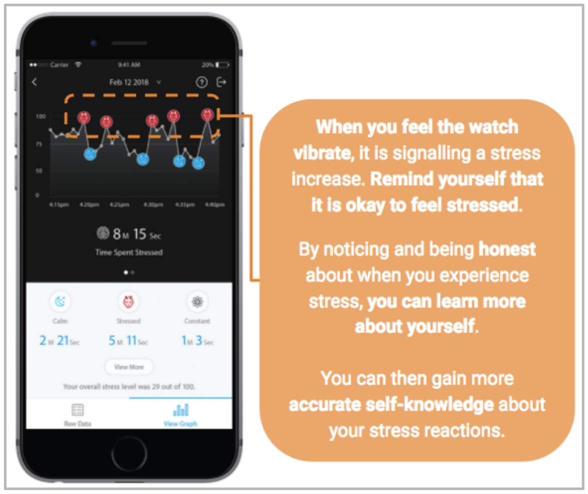


Figure 8. Self-Knowledge condition for an automatic stress tracking system.

Condition 4: No frame

Finally, a control condition was simply given a description of the system as measuring stress but was given no additional content with regards to the goal of the system. This allowed us to compare results against a control who receives apparent stress information with no specific goals described for the system.

Frame Comprehension and Manipulation Checks

To ensure that participants have read and understood the different system descriptions, they were given comprehension check questions after viewing the onboarding content per their condition. Participants in the experimental conditions (Improvement, Self-Efficacy, and Self-Knowledge) were asked "*(Select all that apply) What is/are the goal(s) of this system?*" with possible multi-select options for (1) 'To help me learn more about myself', (2) 'To increase my stress', (3) 'To decrease my stress', (4) 'To increase my awareness of stress', (5) 'To distract me from feeling stressed', and (6) 'None of the above'. Options 1 through 5 were presented in a randomized order, with option 6 anchored at the bottom of the list. We should see that selections for answer #3 (''To decrease my stress') as most frequent in the improvement-oriented conditions (C1 and C2) and answer #1 (''To help me learn more about myself') should be most frequent in the Self-Knowledge condition.

Participants in the Self-Efficacy condition had an additional comprehension question: "*(Select all that apply) What should you do whenever you feel stressed?*". Answer options were multiple choice and ranged from (1) 'Ignore the stress', (2) 'Take a slow, deep breath from my stomach', (3) 'Answer questions more quickly', and (4) 'None of the above'. Participants who have read and understood the Self-Efficacy system description are expected to select option 2 '*Take a slow, deep breath from my stomach*'.

Procedure

Participants first submitted a consent form and then completed the pre-test anxiety scale (STAI) and beliefs about stress scale (BASS). Participants were randomly allocated to one of the four conditions. After reading the corresponding system descriptions, each participant read instructions for the timed test. This included multiple-choice questions to check they had read the system description and study instructions. Following instructions, participants gave ratings of anticipated motivation.

We then set the participant up with the stress tracking system (Emvibe) and proceeded with the timed test. After which, the participant removed Emvibe, submitted the post-test questionnaire, was thanked, and debriefed through a web form. Participants were prompted within the post-test questionnaire for an optional step where they could view their stress data. Research assistants conducting the experiment were blind to conditions and did not know about hypotheses at the time of the study.

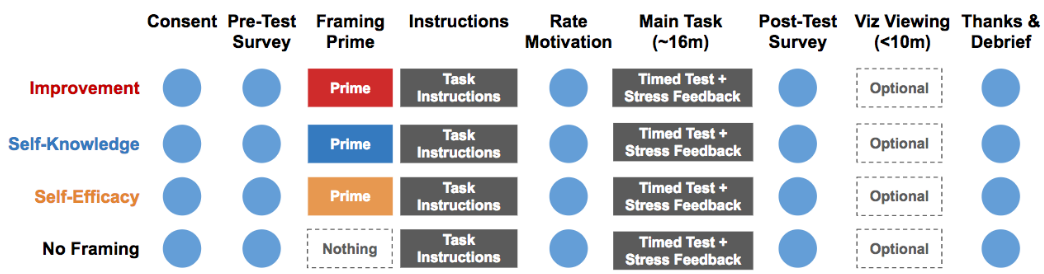


Figure 9. Design of Study 1 to compare the impact of different implicit system goals for stress on perceived stress and engagement.

Dataset and Analysis

The dependent variables for this study were (1) pre-post changes in STAI-State scores (i.e., Anxiety), (2) changes in BASS scores, (3) Perceived Stress rating, (4) text analysis of free-written descriptions of the study experience, and (5) measures of data engagement and interest in continuing to use the system. To be kept in the final dataset, participants needed to pass the multiple questions to confirm that they understood the system descriptions.

Processing Qualitative Responses: Participants could submit free-written descriptions for how they felt during the test-taking experience and what they liked or disliked about the experience. As in prior research, we used the Linguistic Inquiry and Word Count (LIWC) software (Pennebaker, Francis, & Roger, 2001) to automatically process these text responses. LIWC is a widely used lexical analysis tool, which includes up to 93 different linguistic categories and has good reliability compared with human judges (Pennebaker et al. 2015). The LIWC analysis only included categories that are relevant to our research questions: positive emotion terms (e.g., 'happy', 'enjoyed'), negative emotion terms (e.g., 'hate', 'unhappy'), and anxiety-specific language (e.g., 'nervous', 'anxious').

There are limitations to using such automatic approaches, which can fail to detect implicit expressions of emotion (Wiebe, Wilson, & Cardie, 2005) so two researchers also individually coded the written participant descriptions of the test experience, categorizing descriptions by emotional valence as 1 ("Negative"), 2 ("Neutral or Mixed Valence"), or 3 ("Positive"). Inter-rater agreement was good (Cohen's kappa over 0.61), so an average of the two ratings is used to assess valence differences across framings.

Study 1: Stress Tracking Pilot Results

Two pilot studies were conducted to refine the procedure and materials for Study 1. The first pilot was an online survey to assess whether different beliefs about stress are *perceivable* within each of the condition frames. We found from the online pilot for study materials that the Improvement and Self-Efficacy system descriptions were comprehensible, but that the Self-Knowledge system description needed additional content to correctly convey the intended goal.

The second pilot was an in-lab study conducted with UCSC student participants (n=24) for only the three experimental frames: Improvement, Self-Efficacy, and Self-Knowledge.

Study 1a Pilot: Web Testing of Experimental Frames

Participants: 100 respondents were recruited through mechanical turk to complete an anonymous online survey. Survey respondents were roughly an equal distribution of self-identified gender (41 females, 57 males, 2 declined to state) and ranged from 18 to 61 years of age (Mean = 34.21). In the survey, respondents were randomly presented with one of these stress tracker system descriptions for either Improvement (n=35), Self-Efficacy (n=31), or Self-knowledge (n=34). No identifiable data was collected and they were asked no personal questions.

Measures and Procedure: The respondents were presented with experimental system descriptions and asked questions related to usefulness ("*Rate the extent to which a system like this seems useful or not useful to you:*"), likeability ("*To what extent do you like or dislike this system description?*"), and given an option to free-write what they dislike or like about the system description. Responses were given on a 5-point scale with higher scores indicating greater likeability or perceived usefulness. The goal of these questions was to address whether one condition was particularly disfavored which could confound our results.

Participants were also instructed to provide a free-write for what they consider the goals of the system to be. This free written explanation was used to check if participants largely understood the different goals described by the system and to what extent they might have incorrectly interpreted the system descriptions.

Respondents were then instructed to rate the Beliefs About Stress Scale (BASS) from the perspective of the phone application: "*For the following items, please rate the degree to which each you think each statement matches the views of THIS PHONE APPLICATION*". They were then given modified instructions for the BASS "*The following items complete this statement: 'This PHONE APPLICATION thinks that being stressed...*"

As a reminder, the BASS is a scale with three subcomponents to measure negative beliefs about stress (e.g., "*Being stressed is something I need to avoid*."), positive beliefs about stress (e.g., "*Being stressed enables me to reach my full potential*."), and perceived stress controllability (e.g., "*Being stressed is something I am able to control through my actions*.").

Results: As expected, a series of one-way ANOVAs showed that Improvement and Self-Efficacy system descriptions had highest ratings for stress controllability ($F(2, 97)=7.917$, $p=.001$) and negative stress beliefs ($F(2, 97)=5.676$, $p=.005$). There were no differences across the three system descriptions for positive stress beliefs ($F(2, 97)=.660$, $p=.519$).

There were also no significant differences in likeability of the different system descriptions ($F(2, 97)=1.703$, $p=.188$) or perceived usefulness ($F(2, 97)=.014$, p=.986). Mean liking and perceived usefulness scores of the descriptions were between neutral and moderately positive.

	Stress is Controllable*	**Negative Beliefs***	**Positive Beliefs**	**Perceived Usefulness**	**Likable**
Improvement	9.85 (1.22)	24.91 (6.55)	8.34 (3.6)	3.34 (1.21)	3.97 (1.01)
Self-Efficacy	10.39 (1.56)	23.13 (7.71)	8.71 (3.93)	3.39 (1.15)	4.29 (0.78)
Self-Knowledge	8.91 (1.75)	19.41 (6.46)	9.32 (3.18)	3.38 (1.18)	3.85 (1.11)

Table 1. Means and standard deviations of results from a pilot with 100 survey respondents. A series of one-way ANOVAs showed that participants perceived the Improvement and Self-Efficacy framed systems as believing that stress is more negative and also controllable. The '*' indicates a significant difference between conditions at p<.05.

Study 1b Pilot: In-Lab Test of Full Procedure

A second pilot was conducted in-lab with 24 total participants (12 female, 12 male), aged 18-23. Participants were randomly allocated to the Improvement (n=9), Self-efficacy (n=5), or Self-Knowledge (n=10) conditions. Participants completed the full procedure with the exception of the visualization interaction step as that portion was under development at the time of the pilot.

These results suggested that at least 10 minutes could be dedicated to the visualization exploration step. We also saw from this small data set that results pointed in expected directions, with participants in the Self-Knowledge frame having the lowest increases in STAI-state scores, higher reports of feeling calm/relaxed, and lower reports of feeling stressed/tense. However, this was a small pilot so results could be by chance.

	STAI Change	Stress	Tense	Calm	Relaxed
Improvement	6.22 (6.65)	3.56 (0.53)	3.22 (0.67)	2.00 (0.87)	1.44 (0.53)
Self-Efficacy	7.80 (4.97)	3.80 (0.84)	4.00 (1.23)	1.80 (0.45)	1.60 (1.34)
Self-Knowledge	2.60 (3.50)	3.10 (1.10)	2.90 (1.37)	2.70 (1.34)	2.50 (1.43)

Table 2. Means and standard deviations of emotion outcomes for Study 1 from an in-lab pilot with 24 participants.

	Data Interest	Adoption Interest	Sensor Accuracy
Improvement	4.11 (0.78)	2.56 (1.51)	3.67 (0.71)
Self-Efficacy	4.20 (0.84)	3.20 (1.48)	4.00 (0.71)
Self-Knowledge	3.8 (0.92)	3.10 (1.37)	3.20 (0.92)

Table 3. Means and standard deviations of hypothetical intentions to use the device, view stress data, and perceived sensor accuracy. These results are from an in-lab pilot for Study 1 with 24 participants.

From these results, sample size planning was conducted at 80% power and 5% significance level looking at changes in STAI-state (i.e., Anxiety) scores from pre-post surveys. The two conditions used for sample size planning were pilot results from the Self-Knowledge and Improvement conditions as these were expected to have the greatest differences in stress reports. The equation used for sample size planning is from Sealed Envelope, a collection of online support tools for clinical trials. The sample size planning calculator used was for superiority testing between two conditions on a continuous variable (Sealed Envelope Ltd., 2012). The advantage of using a sample size planner from Sealed Envelope is that these calculators have been tested for accuracy against published results. In addition, the code used for sample size planning here is open source and available publically online. The sample size estimate is based on the formula $n = f(\alpha/2, \beta) \times 2 \times \sigma^2 / (\mu_1 - \mu_2)^2$. Here μ_1 and μ_2 are mean outcomes for the Self-Knowledge (Mean: 2.6) and Improvement (Mean: 6.22) conditions in the pilot, and σ is the standard deviation of the outcome (SD: 5.08). The result of this calculation suggested a minimum of 31 participants per condition for a total minimum of 124 participants across the four conditions.

Study 1: Stress Tracking Main Results

Participants

186 participants from a large U.S. university volunteered for the study. Of these, 48 were removed for issues such as failing attention check questions, or malfunctioning hardware during the study. The final sample consisted of 139 participants (Female: 92, Male: 43, Genderqueer: 2, Decline to State: 2). Participant ages in the final sample ranged from 18 to 33 (Mean: 20.44 SD: 2.29). Participants received course credit for completing the study.

Given that approximately 25% of participants were excluded from analysis, a set of *t*-tests, Mann-Whitney U tests, and chi-squared tests were used to compare those who were excluded versus retained. Those who were retained were more anxious than those excluded from the study. Retained participants had higher *General Anxiety* (i.e., STAI-Trait) scores (Mean: 23.74 SD: 5.24) than those who were excluded (Mean: 21.87, SD: 4.79), $t(184)=-2.150$, $p=.033$. Retained participants also had higher *Perceived Stress* scores (Mean: 15.5, SD: 2.92) than those who were excluded (Mean: 14.51, SD: 2.80), $U=2467.5$, $z=-2.520$, $p=.012$. There were otherwise no differences in pre-test characteristics, age, or gender (ps .265 to .846).

Participants who were retained showed some interest in stress-tracking tools with 82 participants selecting that they would want a tool to automatically track stress. However, participants who were retained did not report having much knowledge of commercial options. Only two participants checked in the pre-test that they had heard of the Pip sensor before and on a 5-point scale to indicate familiarity with tools to track stress or emotion the average score was low (Mean: 1.86, SD: 0.91). There were no differences across conditions in pre-test motivation to perform well on the exam ($X^2(3)=3.791$, $p=.285$) or number of perceived stress vibrations ($X^2(3)=1.156$, $p=.764$).

Emotion Outcomes: No differences in Anxiety or Perceived Stress

Dispositionally, participant individual characteristics did show the expected relationship between stress beliefs and stress outcomes. A series of Pearson's correlations showed that higher scores on the *Negative Stress Beliefs* subscale of the BASS was associated with greater pre-test *Anxiety* (i.e., STAI-state) scores ($r=.323$, $p<.001$), post-test *Anxiety* scores ($r=.302$, $p<.001$), and overall *Perceived Stress* ($r=.253$, $p<.001$). This supports prior work that negative beliefs about stress, including beliefs that stress should be reduced, is associated with more extreme stress experiences (Fischer, Nater & Laferton, 2016).

A series of repeated measures ANOVAs was next conducted to look at effects of Time (i.e., pre-post scores) and interaction between Time and Condition on *Anxiety* scores. There was a main effect of time with *Anxiety* scores increasing on average by 5.39 points (SD: 5.28), $F(1, 135)=145.170$, $p<.001$, $\eta_p^2=.518$. However, there were no significant interaction between Time and Condition for *Anxiety*, $F(3, 135)=1.403$, $p=.245$, $\eta_p^2=.030$. A one-way ANOVA similarly showed no difference across conditions in final their final ratings of overall *Perceived Stress*, $F(3, 135)=.064$, $p=.979$, $\eta_p^2=.001$.

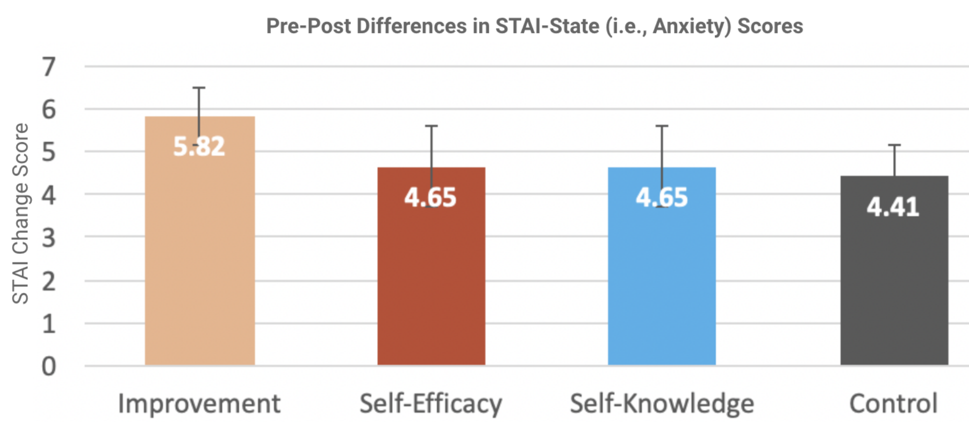


Figure 10. Average differences in pre-post scores for the STAI-State scale as a measure of immediate Anxiety. A repeated measures ANOVA showed no Time by Condition interactions for STAI-state scores. Error bars represent ± 1 standard error.

	Anxiety (Pre)	Anxiety (Post)	Perceived Stress
Control	20.32 (5.23)	24.73 (4.87)	15.70 (2.97)
Improvement	19.66 (4.65)	25.47 (5.24)	15.53 (2.75)
Self-Efficacy	19.81 (5.06)	24.45 (4.58)	15.41 (2.69)
Self-Knowledge	18.78 (5.49)	25.48 (5.51)	15.45 (3.35)

Table 4. Means and Standard Deviations of emotion outcomes for Study 1 (n=139). Statistical tests showed a main effect for time, with Anxiety scores increasing at post-test, but no significant difference in Anxiety change scores or Perceived Stress scores across conditions.

A series of repeated measures ANOVAs were conducted to look at impact of Time and Condition by Time interactions for beliefs about stress. There was a main effect of Time with scores decreasing for *Perceived Stress Controllability* ($F(1, 135)=19.642$, $p<.001$, $\eta_p^2=.127$), decreasing for *Positive Stress Beliefs* ($F(1, 135)=44.563$, $p<.001$, $\eta_p^2=.248$), and increasing *Negative Stress Beliefs* ($F(1, 135)=6.583$, $p=.011$, $\eta_p^2=.046$). There were no significant interactions between Time and Condition (*p*s .205 to .377).

	Positive (Pre)	Positive (Post)	Negative (Pre)	Negative (Post)	Control (Pre)	Control (Post)
Control	10.35 (2.26)	8.81 (2.59)	23.65 (3.8)	23.89 (3.84)	8.00 (1.55)	7.43 (1.94)
Improvement	9.94 (2.3)	9.21 (2.93)	24.5 (3.34)	25.00 (3.92)	7.82 (1.41)	6.61 (1.73)
Self-Efficacy	9.58 (2.28)	7.84 (3.15)	23.87 (3.6)	25.19 (3.64)	7.29 (2.39)	6.84 (2.24)
Self-Knowledge	11.09 (2.14)	10.09 (2.42)	21.97 (4.36)	22.30 (4.17)	8.15 (2.03)	7.69 (2.49)

Table 5. Means and Standard Deviations of pre-post Beliefs about Stress for Negative, Positive, and Controllability stress beliefs.

Engagement: Self-Knowledge Condition had Greatest Reported Engagement, No Differences in Behavioral Engagement

Kruskal-Wallis H tests show that participants in the Self-Knowledge Frame had the highest distribution of ratings for *Data Interest* ($X^2(3)= 8.657$, p=.034) and *Adoption Interest* ($X^2(3)= 8.017$, p=.046). Subsequent pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple pairwise comparisons. The adjusted p-values are presented here. The post hoc analysis revealed that differences in *Data Interest* were driven by differences between the Self-Efficacy and Self-Knowledge conditions (adjusted *p*=.028). There were no statistically significant pairwise comparisons for *Adoption Interest* (adjusted *p*s .071 to 1.0). The behavioral measures for engagement involved whether participants chose to look at their stress data and number of seconds that participants spent viewing their data. Overwhelmingly, a majority of participants selected to view their stress data (n=134) and there was no significant difference across conditions in amount of time spent viewing visualizations ($F(3, 112)=.555$, p=.646, $\eta_p^2=.015$).

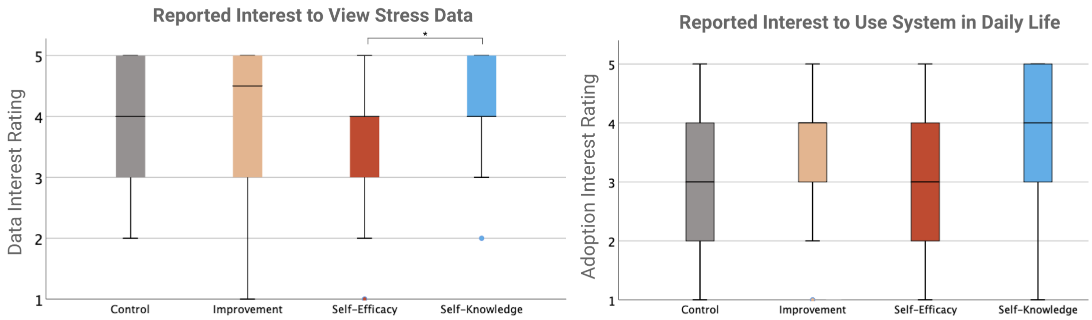


Figure 11. Distribution of participant ratings in interest to use stress tracking system in daily life (*Adoption Interest*), and interest in viewing their data from the study (*Data Interest*). Participants in the Self-Knowledge condition had the highest ratings of *Adoption Interest* and *Data Interest*. The '*' indicates a significant difference at p<.05.

	Data Interest*	Adoption Interest*	Sensor Accuracy
Control	4.05 (.88)	3.08 (1.32)	3.41 (.73)
Improvement	3.95 (1.29)	3.47 (1.23)	3.42 (.92)
Self-Efficacy	3.45 (1.21)	2.90 (1.39)	3.29 (.90)
Self-Knowledge	4.27 (.84)	3.76 (1.15)	3.67 (.82)

Table 6. Means and standard deviations of rated interest to view stress data (Data Interest) hypothetical intentions to use the device (Adoption Interest), and perceived Sensor Accuracy. The '*' indicates significant differences across conditions (p<.05).

Explanatory Mechanism: Self-Knowledge Participants Rated Themselves More Successful

A Kruskal-Wallis H test was conducted to look at differences in ratings of *Perceived Success* across experimental conditions (Improvement, Self-Efficacy, Self-

Knowledge) showing a significant difference in ratings ($X^2(2)$= 43.741, p<.001). Pairwise comparisons with Bonferroni corrections for multiple comparisons shows that this was driven by differences between the Self-Knowledge and Self-Efficacy condition (p<.001) and Self-Knowledge and Improvement condition (p<.001).

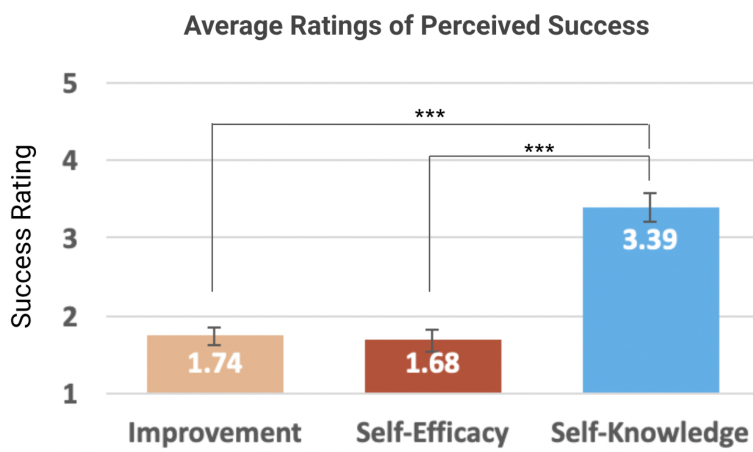


Figure 12. Average ratings of *Perceived Success.* Higher scores indicated greater self-perceived success at achieving the aim of the system. Participants in the Self-Knowledge condition rated themselves as more successful at the end of the study. Error bars represent ± 1 standard error. The '*' symbol indicates p <.001.**

A mediation analysis was performed using an ordinal logistic regression to assess the impact of condition on *Data Interest*, and *Adoption Interest* when controlling for *Perceived Success*. An ordinal logistic regression was first conducted with condition as the predictor variable and *Data Interest* as the outcome variable, showing that the Self-Efficacy condition was predictive of lower *Data Interest* ratings (Wald X^2 = 7.395, b=-1.272, 95% CI [-2.188, -.355], p = .007). When controlling for *Perceived Success*, the Self-Efficacy condition is no longer predictive of Data Interest suggesting that differences in *Perceived Success* fully mediate these condition differences for *Data Interest* (Wald X^2 = .528, b=-.425, 95% CI [1.571, .721], p = .467). A similar test was conducted for *Adoption Interest* again showing that while the Self-Efficacy condition was originally predictive in changes in *Adoption Interest* (Wald's X^2 = 6.608, b=-1.174, 95% CI [-2.069, -.279], p = .010), it is no longer predictive when controlling for *Perceived Success* (Wald's X^2 = .318, b=.326, 95% CI [-.808, 1.460], p = .573). Again, this suggests a full mediation of *Perceived Success*.

Experience Freewrites: Largely No Differences in Emotion Descriptions

In this dataset, we found no statistically significant differences between conditions for *Negative Emotion Language* in descriptions of how participants felt (F(3, 135)=.354, p=.787, η_p^2 =.001). However, there was a difference in rates of *Negative Emotion Language* across the descriptions of the study experience, (X^2(3)=8.139, p=.043). Subsequent pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple pairwise comparisons. The adjusted p-values are presented here. The post hoc analysis revealed that differences in *Negative Emotion Language* had no significant pairwise comparisons (adjusted *ps* =.067 to 1.0).

Free-written responses were also coded by two human judges who were not the author and who were blind to condition allocation. There was good agreement between the two raters (Cohen's kappa =.643, 95% CI [.539, .747], p<.001). An average of codes from the human raters was used to look at condition differences. We found no statistically significant difference across conditions for valence of *Experience Freewrites* ($X^2(3)=4.535$, $p=.209$).

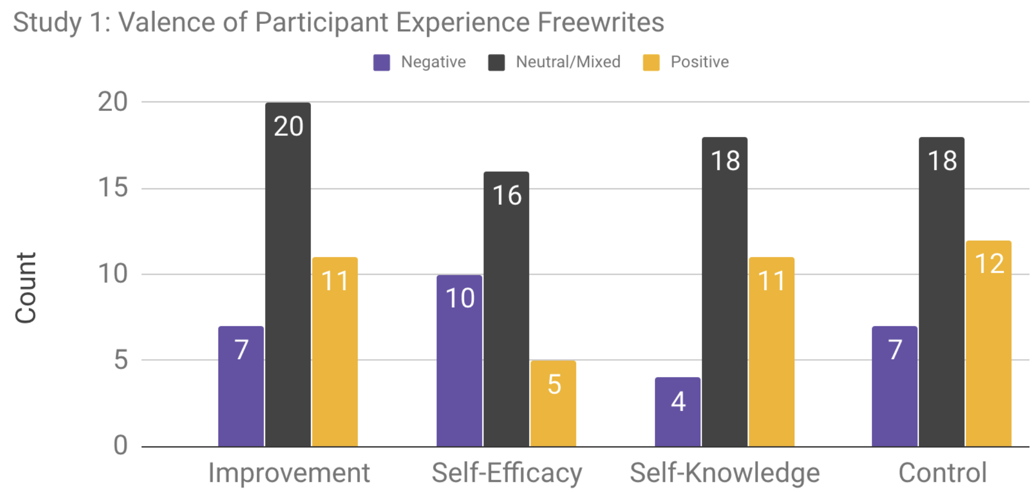


Figure 13. Two human coders who were not the author and were blind to conditions individually coded participant freewrites as 'Positive', 'Neutral/Mixed', or 'Negative'. An average score between two coders was used for statistical tests, showing no differences in valence of participant *Experience Freewrites* across the conditions. For ease of interpretability, results from one human coder are shown here to display the distribution of 'Positive', 'Neutral/Mixed', and 'Negative' freewrites of participants across the conditions.

ID	Quote	Rater 1	Rater 2
P107	*I actually really enjoyed the experience with the stress detecting technology. I think it is really interesting as to how a simple watch could detect that you are feeling stressed over things. I liked that it vibrates to let you know when you are stressed because it is like a reminder that you should probably chill out and breathe.*	**Positive**	**Positive**
P91	*I didn't particularly dislike or like anything about the technology, I was just holding something and wearing a watch. I suppose this is a good thing since it is low maintenance and easy use and set up.*	**Neutral/ Mixed**	**Neutral/ Mixed**
P39	*I hated the buzz. It really just felt patronizing in this context. "Like hey reminder you are sucking at this test." I think if i implemented it on my own terms I would have a different feeling about it, but I dont want a buzz to distract me while I am taking a test.*	**Negative**	**Negative**

Table 7. Examples of 'Positive', 'Neutral/Mixed', and 'Negative' text free writes from Study 1.

Study 1 Discussion

Results from Study 1 provide partial support for the main hypotheses. We see that there were no counterproductive increases in stress as a result of being given goals for stress reduction (i.e., well-being effects). This lack of well-being differences across conditions could be partly attributed to quantitative results showing participants across conditions did not internalize different beliefs about stress that were reflected by the system descriptions. However, there were key differences in reported engagement and in the hypothesized directions with Self-Knowledge participants reporting greater interest in viewing their stress data and adopting the stress-tracking tool for daily use. These effects for engagement appear to be mediated by differences across conditions in Perceived Success, with Self-Knowledge participants tending to consider themselves the most successful at the end of the study. The implications of Study 1 results are discussed in greater detail in the General Discussion section. I now turn to Study 2 which largely replicates and expands upon Study 1 results in examining potential well-being and engagement effects in a long-term deployment of a manual emotion-monitoring system.

Study 2: Impact of Positivity Goals on Emotional Well-being

Study 1 looked at the role of implicit goals in an application that is described as automatically monitoring physical signals of stress. While an in-lab study has the benefit of allowing for a broader range of experimental conditions, Study 2 allows for testing hypotheses in a real-world context by exploring the impact of implicit goals in a system that supports daily manual emotion tracking. The goal for this study was to understand how goals for emotion embedded within a manual affective PI system impact aspects of one's emotional well-being, beliefs about emotion, and engagement with the system. Compelling in-lab, experimental research has demonstrated counterproductive well-being effects when individuals are primed to value happiness (Ford & Mauss, 2014; Mauss et al., 2011), and such effects are tested in Study 2 in an applied setting.

The second core question of Study 2 is how participants engage with systems under the presence of different implicit system goals. I hypothesized that systems which present goals for improvement towards greater emotional positivity will have paradoxical effects for participant well-being. I also hypothesized that participants will be less likely to engage with the PI system as they will feel less successful when presented with an emotion standard. Similar to Study 1, Study 2 highlights the impact of implicit goals on engagement with self-relevant information. In contrast to an improvement-oriented system, I anticipated that a system which encourages more accurate self-knowledge and emphasizes no emotion standards will result in greater reported and actual engagement with the system.

As in prior research (Hollis et al., 2017), we deployed an application we have developed for daily mood tracking. Participants were randomly allocated to one of two experimental conditions or a survey-only control. Participants in the experimental conditions received different onboarding instructions for a manual, emotion-monitoring system. In addition, the app interface was be modified to make more salient particular system goals. For one experimental condition, the emotion-tracking system reflected goals for high positivity and low negativity (the 'Improvement' frame). For the other experimental condition ('Self-Knowledge' frame), participants instead were encouraged to accept emotions equally, and told that the goal of the system is for more accurate self-knowledge (see Appendices 4 and 5). The purpose of framing the Self-Knowledge condition in this way was to reduce apparent system goals for positivity, based on pilot results.

Participants submitted pre-post measures of well-being, and beliefs about emotion. In addition, we prompted participants for perceptions of accuracy for visualized forecasting analytics (Hollis et al., 2017) and in particular tested whether *Improvement Frame* participants consider the analytics to be less accurate compared to *Self-Knowledge* participants. We also gathered subjective and behavioral measures of engagement in the form of reported interest in continuing to use the system, reported frequency of viewing data visualizations, number of entries participants completed, and length of those entries. In optional follow-up interviews, participants discussed their experience and gave feedback on ideas that hypothetical emotion-monitoring systems could reflect.

Study 2: Research Questions

To restate, Study 2 looked at (RQ1): *What are the consequences on subjective reports of happiness and general well-being resulting from different system goals for emotion?* This was addressed in the form of daily mood reports and pre-post changes in scales to measure subjective happiness and mood frequencies. Following RQ1, I also tested for a potential mechanism of any possible well-being effects (RQ2): *To what extent users of these systems internalize the system goals they are presented with?* This was measured by pre-post surveys for (1) valuing happiness, and (2) prioritizing positive activities.

Finally, we could explore aspects of direct system use and perceived accuracy that may differ depending on system goal (RQ3). This is broken into multiple subcomponents such as *How is engagement with a manual emotion-tracking system impacted by the presence of different system goals for emotion? How is perceived accuracy of analytics impacted by system goals for emotion?* Engagement was measured as reported interest in adopting the system, ratings of likelihood for continued use, and reported rates of data engagement. Behaviorally, engagement was captured by number of entries submitted in the application and text length of entries. Given Study 2 involves a manual system, I also looked at potential issues around submitting greater positivity-correspondent information for the Improvement frame: *How are reports of logging bias impacted by system goals that emphasize positive emotion?*

Study 2: Hypotheses

Hypothesis 1: Participants will have the lowest engagement with the mood tracking system when they are in conditions which have high positivity goals (Improvement). This will be potentially mediated by condition differences in (1) pre-post well-being scores and the life stress inventory score or (2) perceived success at post-test. As predicted by OSA, with a self-to-standard discrepancy, they can be motivated to avoid information (e.g., Greenberg & Musham, 1981). If participants do not consider themselves unsuccessful, then data avoidance should not occur.

Hypothesis 2: Negative analytics about one's mood in terms of future forecasts will be considered less accurate if participants have used a system that advocates for increased positivity.

Hypothesis 3: Pre-post scale changes in subjective reports of happiness will be lowest in 'Improvement frame', where there is a high standard for positivity. This will include controlling for the Life Experiences Scale.

Hypothesis 4: Participants in the Improvement condition will report valuing happiness more at post-test. This will partially serve as a manipulation check for our experimental conditions. It would also explain a potential mechanism for any well-being findings and we can address the research question of whether users of these systems adopt the goals that are embedded within them.

Hypothesis 5: Submitting norm-correspondent information: Participants will report greater positive logging bias in the 'Improvement' condition.

The Mood Tracking System

Mood Monitoring: As used in our prior work (Hollis et al., 2017), the mood monitoring application used in Study 2 supports participants manually logging information about current mood, energy level and trigger activities contributing to the current mood. Participants were instructed to create at least 2 mood entries per day and were sent reminders through text message. Making a mood entry is lightweight and can typically be done in about 40 seconds.

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). To provide additional context, participants can also optionally log time, date and set location

(Home, Work, Other). After selecting mood, energy level, time and location, participants are prompted to identify possible factors that explain their mood and rate these factors on a scale of -2 (negatively impacted mood) to +2 (positively impacted mood). Participants can identify up to 14 factors 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). After identifying factors, participants submit a free-write description to reflect on how the factors impacted their mood.

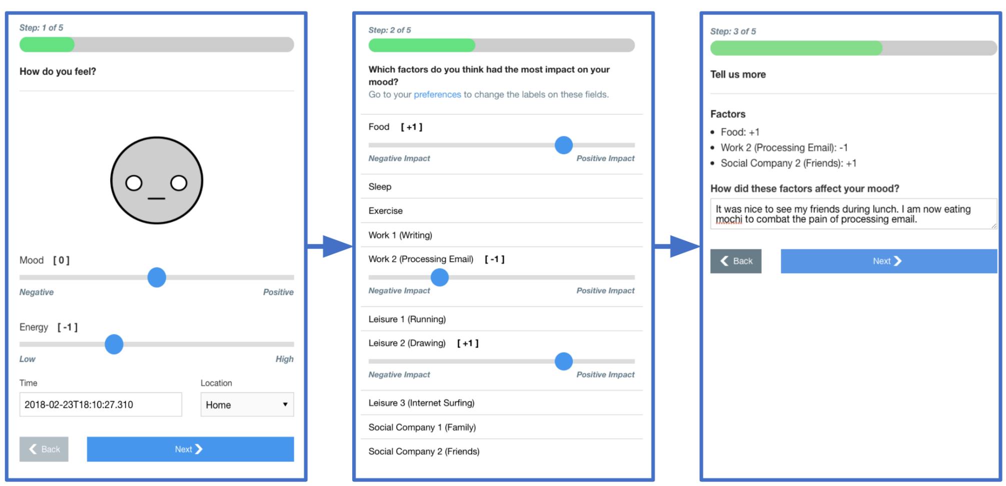


Figure 14. Mood Tracking System Components. The leftmost image shows the mood-monitoring interface with options to rate mood and energy level, as well as contextual information, e.g., time and location. The centermost image shows the UI for choosing factors that influenced their current mood (e.g., that food had a positive impact on current mood). The rightmost image shows a summary page of the entry and allows the user to submit a compliment text entry. There are a total of 14 possible factors the user might select as affecting mood, although not all are shown in this UI view. If custom labels were specified, these were displayed in addition to the factor – for example, 'Custom 1 (Teaching class)' or 'Social Company 2 (Partner)'.

Data Visualizations: The second part of the app is aimed at providing visualized summary statistics of the participant's mood data. We modified the landing page to display a long-term retrospective graph of all mood records made. The visual elements of the data visualizations were slightly different across conditions as piloting showed that the pie chart of mood distributions was associated with increased perceptions that the app valued happiness. For this reason, the pie chart of the Self-Knowledge condition instead shows a breakdown of energy ratings.

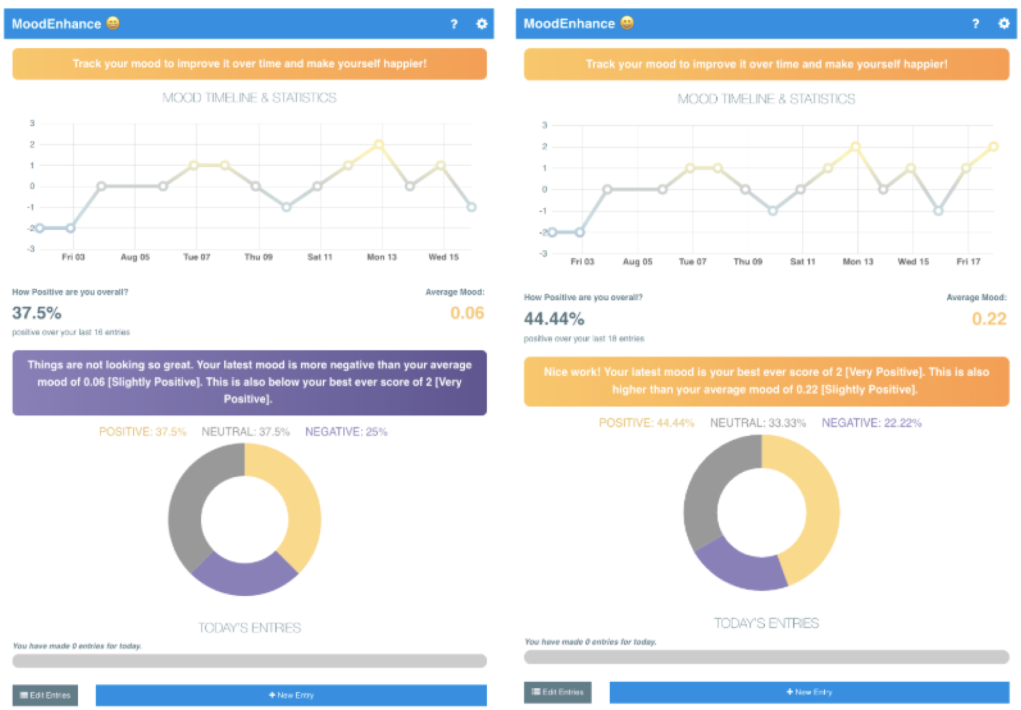


Figure 15. Example app landing page for the Improvement condition. The app landing page UI was designed to value happiness.

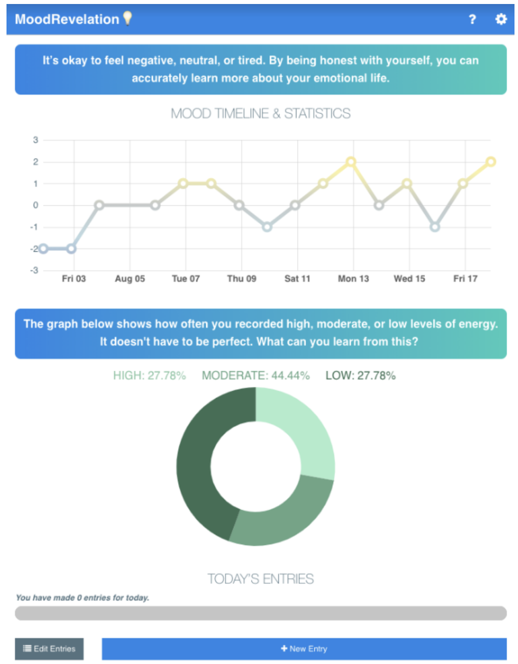


Figure 16. Example app landing page for the Self-Knowledge condition. The app landing page UI was designed to encourage accurate self-knowledge and downplay valuing happiness.

Participant Recruitment

Participants were recruited externally through online forums, social media, and classified ads. All participants were required to be at least 18 years of age or older. Recruitment materials explicitly stated that participants must not have a current life-threatening physical or mental health complications and this was additionally checked through the consent form.

Participants received compensation in the form of Amazon.com gift cards up to their level of study participation. They were offered $5 worth of gift card credits per week of using the mood tracking application, $5 worth of credits for each pre-post survey, and $10 worth of credits for an optional post-study interview. Total study compensation could be up to $35 worth of Amazon.com gift card credits. We also separately hosted a raffle for several $50 Amazon.com gift cards that anyone could opt-in to join.

Measures

Pre-Test Questionnaire

Measures for Emotion Standards and Motives to Track:

Valuing happiness (Mauss et al., 2011): 7-item scale to measure the extent to which participants value happiness (e.g., "*To have a meaningful life, I need to feel happy most of the time.*"). Responses are given on a 7-point Likert scale from 'strongly disagree' (1) to 'strongly agree' (7).

Prioritizing Positivity (Catalino et al., 2014): Participants were asked to rate the extent to which they agree with a series of 6 statements. These statements are intended to capture a range of activities and decision making processes that respondents engage in to increase their happiness. For example "*I structure my day to maximize my happiness*". Ratings are given on a 9-point Likert scale ranging from 'disagree strongly' (1) to 'agree strongly' (9).

Well-Being Measures:

Modified Differential Emotions Scale (mDES; Fredrickson et al., 2003): The mDES measures the reported frequency with which people have experienced different positive (10 items) and negative emotions (10 items) over the past three weeks

(Fredrickson et al., 2003). Participants can give responses on a 5-point scale, ranging from 'not at all' (0) to 'most of the time' (4). While the PANAS is popular in well-being research, the PANAS tends to be biased towards more high-arousal emotions (Harmon-Jones et al., 2016).

Subjective Happiness Scale (SHS; Lyubomirksy & Lepper, 1999): The SHS consists of 4 items to assess global subjective happiness using absolute ratings, as well as ratings of self compared to the perception of others (Lyubomirsky & Lepper, 1999). Participants evaluate their general happiness levels rather than how happy they have been across any specific time period. An example item is, "*Compared to most of my peers, I consider myself…*" which has response categories ranging from "*less happy*" (1) to "*more happy*" (7).

Situational Life Stressors

Life Experiences Survey (LES; Sarason et al., 1978): The LES (Sarason et al., 1978) is a scale consisting of 45 life experiences (e.g., 'Marriage', 'Serious Illness', 'Divorce'). Respondents are instructed to for each event "*…indicate the extent to which you viewed the event as having either a positive or negative impact on your life…*". Ratings are given on a scale of 'Extremely Negative' (-3) to 'Extremely Positive) (+3). To reduce participant burden, they checked which of these events had occurred to them in the past three weeks and that they considered to have impacted them negatively. They then provided ratings only for the events they had selected. As in prior related work (Mauss et al., 2011), we only used the negative impact of events and used sum impact rating across events for a single life stress score.

Demographics and Participant Characteristics

Demographics: Participants submitted age, self-identified gender, income, ethnicity (optional), their highest level of education, and selected their U.S. time zone.

Quality Check: Given the experimental manipulations involve minor wording changes and the study is conducted entirely remotely, ensuring that participants read instructions was crucial to the experiment design. The end of the pre-test survey included an Instructions Manipulation Check (IMC; Oppenheimer et al., 2009) to assess whether the participant read a paragraph of text provided in the web form. This ICM involved reading a paragraph-formatted question with a list of sports activities below. The paragraph instructed participants to select the open-ended option and enter in the text field '*I read the text*'. Participants failing this check in the pre-test survey were not included in the study.

Post-Test Questionnaire

At post-test, participants again completed measures for emotion beliefs (Prioritizing Positivity, Valuing Happiness), and well-being (mDES, SHS). The post-test survey additionally included questions for perceived system accuracy, system use, and system adoption interest.

Perceived Success: Again as with Study 1, participants in the experimental framing conditions were asked to rate on a 5-point scale the extent to which they considered themselves successful or unsuccessful at achieving the aim outlined in the application (i.e., *Perceived Success*). Ratings were given on a 5-point scale ranging from 'Not at all successful' (1) to 'Extremely successful' (5). This allows for a potential mediation analysis of Perceived Success to explain any possible condition differences in reported or behavioral measures of engagement (Davis & Brock, 1975; Duval, Wicklund & Fine, 1972; Greenberg & Musham, 1981).

Study Experience: Similar to Study 1, participants provided a free-write to the prompt "*How was your experience with using the app to track your mood?*". They were then presented with two separate text boxes to list what they liked or disliked.

Analytics accuracy: Participants were presented with a visualized forecast of their expected mood over the upcoming two days and asked to judge accuracy. This visualized analytic allows us to see whether users discount an aspect of system accuracy when it conflicts with a goal to be more positive. Participants first provided ratings of their expectations for what the graph will show allowing us to record their original beliefs. The web form then displayed the forecasted moods. Participants rated the extent to which they considered the forecast accurate on a 7-point scale (1: 'Not at all accurate' to 7: 'Extremely accurate'). They then completed a free-written explanation for their accuracy judgment.

Logging Bias: Given participants could bias record-keeping in the face of system goal, we also explored self-reports of this bias. Participants were asked "*Were* *there any types of entries you avoided recording?*" with responses options for '*Yes*', '*No*', '*Don't know/unsure*'. If the respondent selects '*Yes*', they can then multiselect the types of mood ratings and/or energy ratings they avoided recording, an 'other' option, and a free-written explanation of their answer. They were also asked a parallel question "*Were there any types of entries you emphasized recording more often?*" as my prior work has shown each of these questions yield slightly different information typically with respondents saying that they avoided negative records and emphasized positive (Hollis et al., under review).

Reported Data Engagement: Given there were no behavioral measures of data engagement to record passively in the application, participants instead completed items for how often they viewed visualizations of their data. Participants were reminded that "*The landing page is the screen that would display after you signed in, sometimes showing a visualization of your entries or stats of how you are doing. This is the page where you could click the 'Make an Entry' button*.". They were then asked for the frequency at which they viewed visualizations or statistics on the app landing with response options ranging from 'Daily', '4-6 times a week', '2-3 times a week', 'Once a week', or 'Never'. A similar question was asked for how often or not they deliberately avoided looking visualizations or statistics on the app landing page with options again ranging from 'Daily' to 'Never' and was reverse coded. They were also asked with respect to individual entries: "*When making an entry, how often or not often did you look at visualizations or statistics on the app landing page?*". Options for the relative frequency of viewing data ranged from 'Always looked', 'Looked most of the time', 'Looked about half of the time', 'Sometimes looked', and 'Never looked'. A single composite score of these three questions was used for a reported *Data Engagement* score.

System Engagement: Finally, participants were asked to rate Adoption Interest: "*Rate your interest or lack of interest to use this mood-tracking application in your daily life:*". Ratings for Adoption Interest were given on a 5-point scale ranging from 'Not interested at all' (1) to 'Extremely Interested' (5). They were then shown the text "*You have the option to continue using the app to track your mood. Any additional use is voluntary and you will not be offered additional compensation for this use.*" and asked to rate "*How likely or unlikely is it that you will continue to use this mood-tracking app?*" with ratings given on a 7-point scale ranging from 'Extremely unlikely' (1) to 'Extremely likely' (7). These two items were used as measures of reported *System Engagement*.

Suspicion Check: Finally, participants were prompted with an open-ended question to determine if they knew the study design or hypotheses "*What do you think this study was about?*".

Post-Study Interview

In the post-study survey, participants could indicate a preference to be remotely interviewed. Interviews lasted approximately 45 to 60 minutes and were conducted over web conferencing software (GoToMeeting). These interviews enabled us to better understand the participants' subjective experiences during their research participation. Furthermore, we could learn more about (1) what they liked or disliked about their study experiences, and (2) what other types of ideas about emotions they would want to be reflected in an emotion-monitoring system and why.

Conditions

All participants in experimental conditions had the same retrospective data visualizations and mood-tracking interface. However, they were presented with different onboarding procedures and the app landing page UIs were modified to emphasize different emotion goals.

Condition 1: Improvement

In onboarding materials (see Appendices 4 and 5), participants were instructed of the system goal as the following: "*The goal of this app is to help you monitor and increase your happiness on a day-to-day basis. Being more positive corresponds to a range of other benefits and we want to help you to reach your emotional goals. By tracking your happiness, you can be more likely to increase it!*". The application they would use to track their mood was titled 'MoodEnhance' and the landing page UI elements aimed to convey a goal for increased positivity. The Improvement frame displayed summary statistics for how positive the participant's mood records were, the average mood from their entries, and also a summary chart of mood distributions.

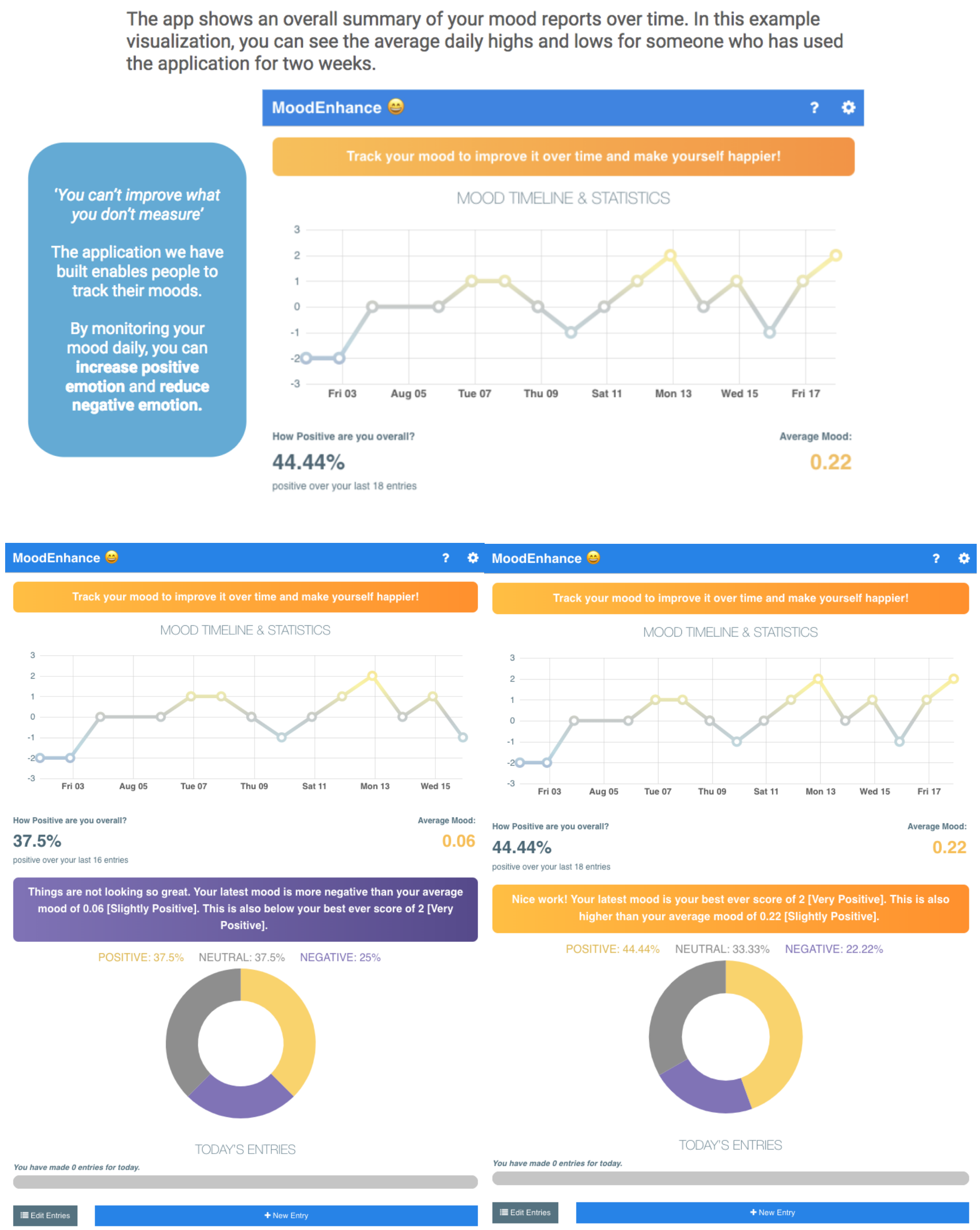
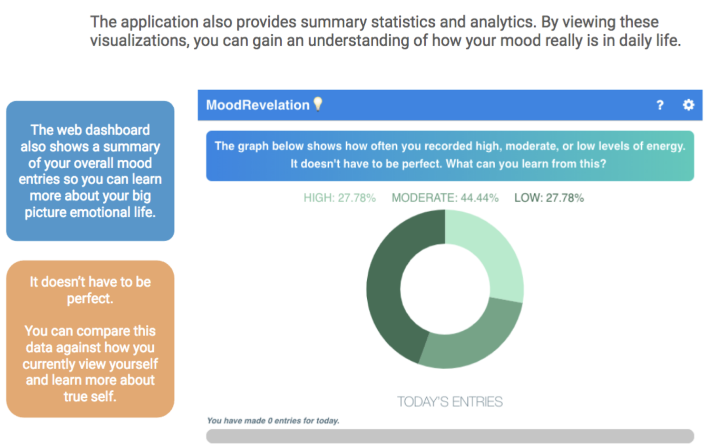


Figure 17. The topmost image shows an example of onboarding materials for the Improvement condition. The bottommost images show app landing pages that participants view if they are in the *Improvement* condition. The landing page emphasizes increased positivity and provides summary statistics for progress.

Condition 2: Self-Knowledge

In onboarding materials (see Appendices 4 and 5), participants were instructed of the system goal as the following: "*The goal of this app is to help you learn more about yourself. Negative and positive emotions are okay and both have value. By being honest with what you are experiencing, you can learn more about what impacts your mood negatively or positively, and by how much.*" The application they would use to track their mood was titled 'MoodRevelation' and the landing page UI elements aimed to convey a goal for increased self-knowledge, downplaying specific emotion standards. The Self-Knowledge frame did not include summary statistics for how positive the participant's mood records were. Based on pilot results, the summary pie chart instead showed a distribution of energy levels recorded in previous entries.



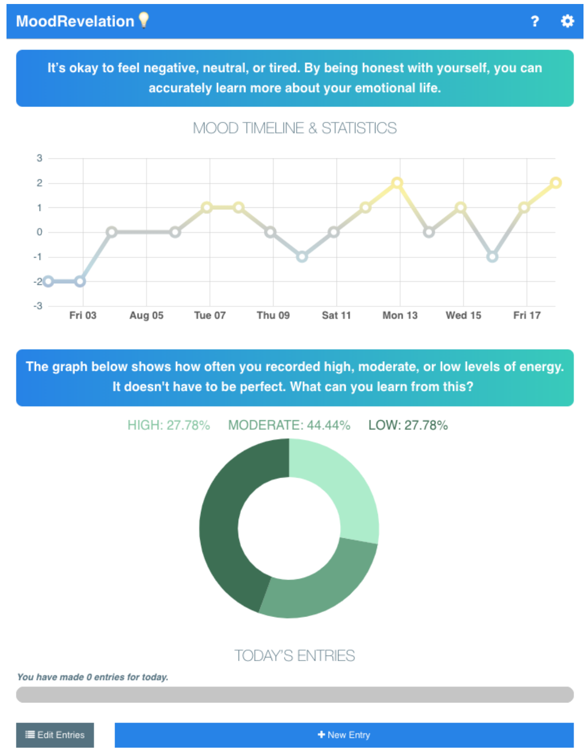


Figure 18. The topmost image shows an example of onboarding materials for the Self-Knowledge condition. The bottommost image shows the app landing page that participants view if they are in the Self-Knowledge condition. The landing page emphasizes more accurate self-knowledge and purposefully downplays specific emotion outcomes.

Condition 3: Survey-Only Control

A subset of participants were randomly allocated to a *Survey-Only* control in which they did not use an emotion-tracking application and only answered pre-post surveys.

Procedure

For the study advertisement, all participants received the same advertisement that the study is about trying an emotion-tracking web application. No specific language related to improvement, regulation, or positivity was mentioned in the advertisement.

Participants who complete the online pretest and satisfied study requirements (e.g., no known mental health issues, over 18 years of age, passed attention checks) were then emailed a link to instructions for the study. These instructions varied depending on the experimental condition that participants are randomly allocated to. Some randomly received instructions related to improvement (C1), others will receive instructions pertaining to self-knowledge (C2), and some were allocated to a survey-only control.

Instructions were presented as a Qualtrics web form. All instructions included the same descriptions for how to successfully make an entry, and troubleshoot any issues. The instructions form included a free-write from the participant to restate the apparent goals of the system. Participant logs were additionally checked one to two times per day by Victoria Hollis for any entries that may raise concern (e.g., self-harm). Participants also received two SMS notifications per day as reminders to make an entry.

After the three week logging period, participants were invited to complete a post-test survey and given the option to interview. The post-test survey re-administered well-being scales, measures of reported engagement, and interest in continuing to use the system. Participants were also presented with forecasting analytics and asked to rate perceived accuracy for these.

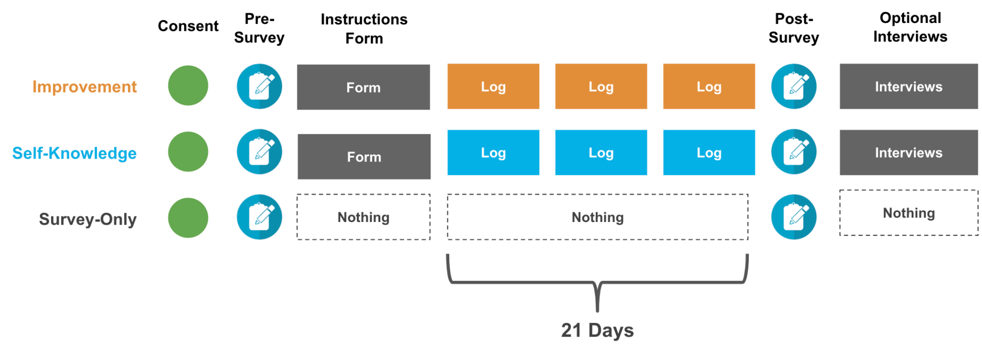


Figure 19. Design of Study 2 showing three conditions: Improvement, Self-Knowledge, and a Survey-Only control.

Dataset and Analysis

The dependent variables for this study were (1) pre-post changes in emotional well-being scores (mDES, SHS), (2) changes in measures of beliefs about emotion (Valuing Happiness, Prioritizing Positive Activities), (3) text analysis of free-written descriptions of the study experience, (4) ratings of perceived success, and (5) measures of data engagement and interest in adopting the system for daily use.

Processing Qualitative Responses: Participants could submit free-written descriptions for how they felt during the study period and what they liked or disliked about the experience. As in Study 1, two researchers who were blind to conditions individually coded the written participant descriptions of the study experience, categorizing descriptions by emotional valence as 0 ("Negative"), 1 ("Neutral or Mixed Valence"), or 2 ("Positive"). If the inter-rater agreement is good (Cohen's kappa over 0.61), then an average of the two ratings is used to assess valence differences across framings.

Logfile Analysis: The web application for tracking mood produces a log of daily reports of mood. Daily mood scores were analyzed for frequency of positive mood, frequency of negative mood, frequency of neutral moods given prior work has shown that conceptually subjective well-being is more closely linked to the relative frequency of positive emotion rather than intensity (Diener et al., 1991).

Dealing with incomplete data: Given this study was conducted remotely and online, there are risks of participants submitting nonsense responses to our surveys or in the daily mood entries. There is also a risk that participants must be removed from the study if they show concerning behaviors. To be kept in the dataset for final analyses, participants must have satisfied the following minimum conditions. All response filtering occurred before the quantitative analysis step.

- Filtering Sign-ups: Participants were only included if they passed the IMC in the pre-study survey.
- Risk to participant: If a participant must be removed from a study because of possible health risks, then their data would not be included in the dataset. For example, reporting self-harm or eating disordered behavior in the mood tracker entries. This did not occur for any participants in the study.
- Minimum Compliance: Those in the experimental conditions must have completed at least 14 entries.

- Multiple Survey Entries: In cases where a participant submitted multiple survey responses, only the first complete survey response was used for data analysis.

Study 2: Onboarding Materials Pilot

Similar to the web-based pilot conducted for Study 1, materials per condition were randomly presented to 76 mechanical turk respondents (n=34 for *Improvement* and n=42 for *Self-Knowledge*). Participants viewed the materials and then submitted scales for *valuing happiness*, the importance of engaging in *positive activities*, and 3 revised questions from Feldman et al. (2007) for *acceptance* (e.g., "*People can tolerate negative emotions*."). These scales were revised to be taken from the perspective of the phone application, similar to what was done for the Study 1 pilot. Similar to the Study 1 pilot, participants also rated the perceived *likeability* and *usefulness* of the system on 5-point scales.

A series of either *t*-tests or Mann-Whitney U tests were conducted to compare ratings for the dependent variables listed above. These tests confirm that materials are perceived as significantly different and in the expected directions. The Improvement condition corresponded to greater ratings of *valuing happiness* ($t(74)=3.139$, 95% CI [2.18, 9.78], $p=.002$) and *positive activities* ($U = 450.5$, $z = -2.756$, $p = .006$). The Self-Knowledge condition had higher ratings for *acceptance* ($t(74)=-2.273$, 95% CI [-3.51, -0.23], $p=.026$) and importantly for ratings that people can *tolerate negative emotion* ($t(74)=-2.773$, 95% CI [-1.49, -0.22], $p=.007$).

Ratings of likeability and perceived usefulness were taken on 5-point scales with higher scores indicating greater liking or perceived usefulness. There were also no significant differences in likeability of frames or perceived usefulness (*p*s .748 to .994). Overall liking and perceived usefulness of the frames were between neutral and moderately positive.

	Positive Activities*	Values Happiness*	Acceptance*	Perceived Usefulness	Likeable
Improvement	41.38 (6.53)	35.03 (7.45)	13.44 (4.08)	3.29 (1.27)	4.12 (0.73)
Self-Knowledge	35.55 (9.41)	29.05 (8.86)	15.31 (3.08)	3.38 (1.08)	4.12 (0.80)

Table 8. Means and standard deviations for an online pilot (n=76) for onboarding study materials. Participants viewed content for either the *Improvement* or *Self-Knowledge* frame and provided scale responses from the perspective of what the phone application believed about mood. The '*' indicates differences across the two frames was significant (p<.05).

Study 2: Mood Tracking Results

Participants

100 participants were enrolled in the study after completing the pre-test survey. The final sample consisted of 79 participants who completed the pre-test survey and, if part of the experimental conditions, made at least 14 entries in the application (Control 24, Improvement: 27, Self-Knowledge: 28). Participant ages in the final sample ranged from 18 to 52 (Mean = 24.84, SD = 6.28). 66 participants self-identified as female, 11 self-identified as male, and 2 identified as Genderqueer. Of the final sample, 76 participants also completed the post-test surveys (Control: 24, Improvement: 24, Self-Knowledge: 28). The app entries of the 3 participants in the

Improvement condition who did not complete the post-test survey were retained for logs analysis.

A series of *t*-tests and chi-squared tests were used to compare demographics of participants who dropped out (n=19) against those retained (n=79). Between the participants who were retained and those who dropped out there were no differences in age (t(.825)=21.883, p=.419), gender ($X^2(13)=16.352$, p=.231), or condition assignment ($X^2(2)=.350$, p=840). For the suspicion check question in the post-test survey, only two participants described that there might be multiple conditions and no participants guessed the study design or hypotheses.

No Differences in Well-Being Scores Across Conditions

Dispositionally, participant individual characteristics did show the expected relationship between scores for *Values Happiness* and well-being outcomes. A series of Pearson's correlations showed that higher scores on the *Values Happiness* scale were associated with greater reported frequency of *Negative Emotions* at pre-test (r=.424, p<.001), lower reported frequency of *Positive Emotions* at pre-test (r=-.359, p<.001), and lower overall perceived *Subjective Happiness Scale* scores (r=-.398, p<.001). This supports prior work that greater reports of valuing happiness are associated with lower well-being (Ford & Mauss, 2014; Mauss et al., 2011).

Three separate repeated measures ANOVAs were then conducted with Condition as a between variable and pre-post well-being scales (SWLS, MDES, SHS) as the within-subjects variable. Post-test scores of Life Stress were included as a covariate in each of the tests. The test showed no main effects for Time (ps .207 to

.897) and no Condition by Time interactions for any well-being measure (*ps* .158 to .649). Logs analyses similarly showed no differences in participant well-being. There were no differences across the conditions in terms of frequency of reported positive emotion or types of affective language used across entries (ps .099 to .919).

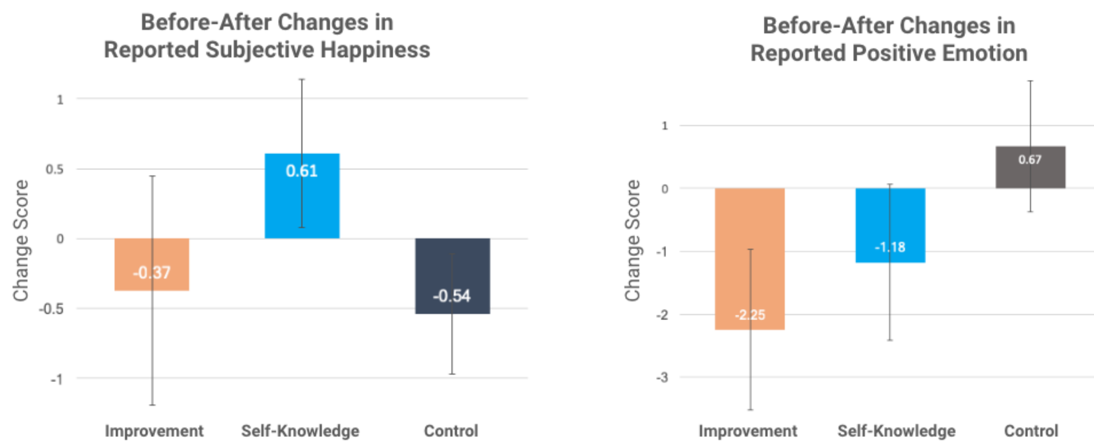


Figure 20. Average differences in pre-post scores for the Subjective Happiness Scale (SHS) and reported frequency of positive emotions from the Modified Differential Emotions Scale (mDES). A series of repeated measures ANOVA showed no Time by Condition interactions for any well-being measures. Error bars represent ± 1 standard error.

A repeated measures ANOVA was also conducted to look at any changes in terms of *Beliefs* about one's emotions. There was a main effect of Time, with participants across all conditions generally reporting a decrease in *Valuing Happiness* over the study period ($F(1, 73)=4.661$, $p=.034$, $\eta_p^2=.060$). However, there were no Time by Condition interactions for other *Belief* scales (ps .624 to .938). Given the lack of differences across *Well-being* and *Belief* measures between conditions, the planned mediation analyses for *Beliefs* were no longer necessary.

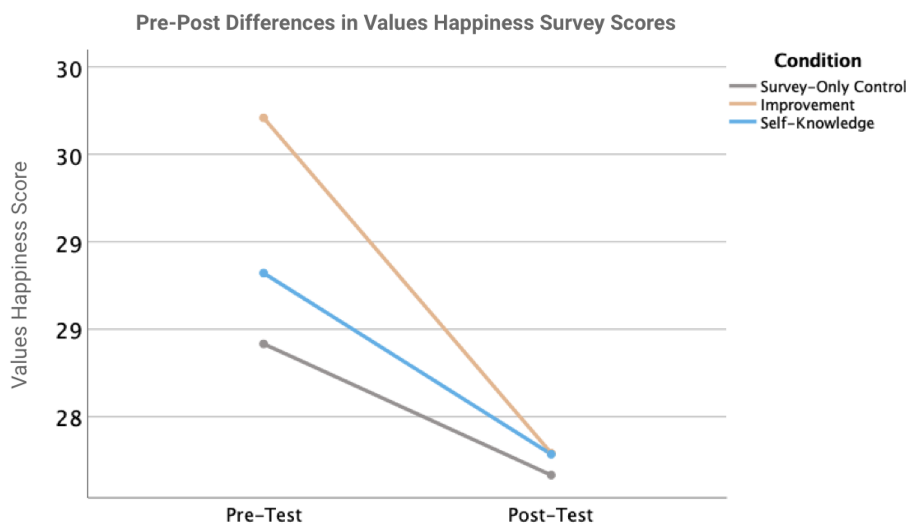


Figure 21. Average pre-post scores for the Values Happiness scale. Scores on the Values Happiness scale range from 7 to 49. A repeated measures ANOVA showed a main effect for time with participants on average reporting that they less valued happiness at the end of the study.

Subset Across Conditions Emphasized Positive and Avoided Negative Entries

One question in this study pertained to issues around logging behaviors, and that participants may be inclined to emphasize positive and avoid negative mood entries (i.e., *logging bias*). There was no differences across conditions in reports of avoiding ($X^2(2)=3.909$, $p=.142$) or emphasizing particular types of entries ($X^2(2)=1.155$, $p=.561$). Across all participants, 13 reported avoiding submitting certain types of entries and 18 reported emphasizing particular types of entries. Of the 13 participants who described some aspect of avoiding types of entries, privacy concerns seemed to be low. On a 5-point scale with higher scores indicating greater privacy concerns, the average score was 2.15 (SD: 1.14, SE: .317).

Seeing that there were no differences between experimental conditions in terms of well-being effects or how participants reported making entries, I now turn to the second key question in this study with respect to user engagement for systems that reflect different implicit goals.

Self-Knowledge Condition Reported Greatest Engagement

Engagement was measured through a series of post-test measures on self-reported frequency of viewing their data visualizations (3-items) and interest in continued system use (2-items, i.e., *System Engagement*). A single sum of the 3-items to measure frequency of viewing data visualizations was used for a single *Data Engagement* score which had good internal reliability (Cronbach's $\alpha = .807$). We also captured behavioral measures of engagement in terms of number of participant entries and text length of entries.

A Mann-Whitney U test shows that the composite *Data Engagement* score was significantly different across the experimental conditions ($U=231.5$, $z=-2.024$, $p=.043$). With regards to *System Engagement*, Self-Knowledge participants gave higher ratings of interest in adopting the system for daily use ($U=209.0$, $z=-2.427$, $p=.015$). However, there was no difference across conditions in ratings of hypothetical likelihood to continue using the system ($U=241.00$, $z=-1.767$, $p=.077$).

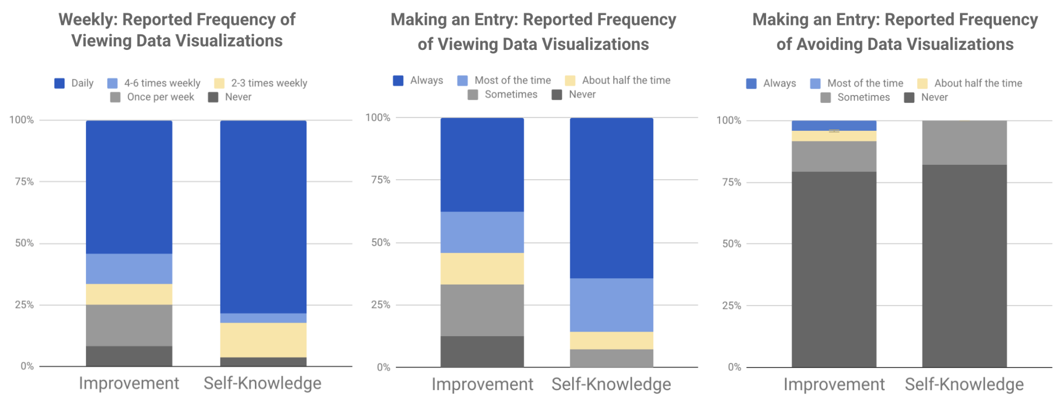


Figure 22. Responses to reported engagement and avoidance with data visualizations. Participants in the Self-Knowledge condition reported viewing their data visualizations more often, and there was no difference between the two conditions in reported avoidance of visualizations.

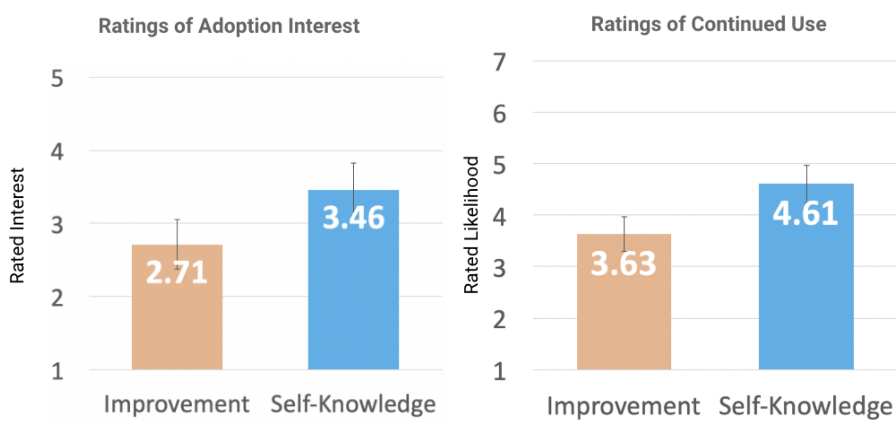


Figure 23. Ratings of *Adoption Interest* and *Continued Use* from the post-test survey.

Turning next to analysis of the daily entries, a one-way ANOVA across all Self-Knowledge and Improvement participants was done to assess whether there were differences in baseline compliance due to condition. This test showed no difference in number of entries between conditions ($F(1, 67)=.170$, $p=.682$, $\eta_p^2=.003$). One potential reason for this lack of difference across conditions could be due to the format and frequency of reminder notifications, as participants were sent two reminders per day through text messages to their personal devices.

Subsequent logs analysis on text length of entries was conducted only for compliant participants who made at least 14 entries. We see that while there was no difference in number of entries made across participants in the Self-Knowledge and Improvement conditions, there was a significant difference in length of text in entries. Self-Knowledge participants wrote on average 181.61 character entries (SD: 95.33, Median: 128) compared to 118.26 characters on average (SD: 92.69, Median: 107) for Improvement participants ($U=254.0$, $z=-2.088$, $p=.037$). Given the script to process text entries was developed by the author, as a precaution differences in text length were additionally assessed with the word count score computed by the LIWC software. This similarly showed that participants in the Self-Knowledge condition wrote more words per entry compared to those in the Improvement condition ($U=252.5$, $z=-2.113$, $p=.035$). Note that text analyses violated Shapiro-Wilk tests for normality (ps <.005 to .033) so non-parametric tests were used here instead.

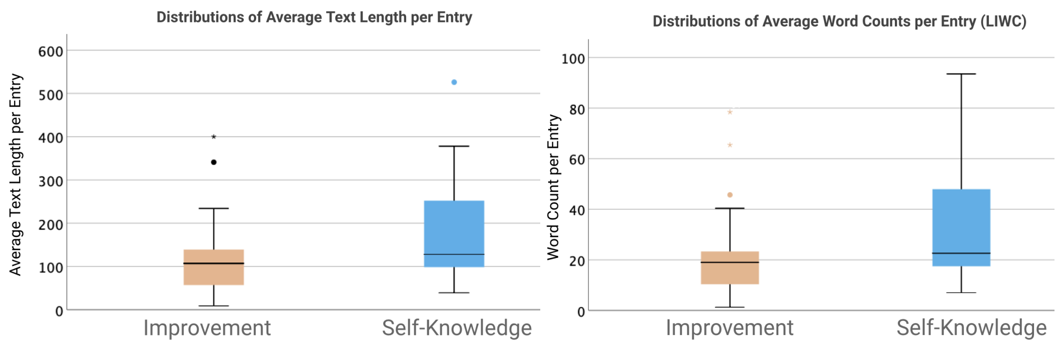


Figure 24. Distribution of average word count and character length of text produced per entry. Participants in the Self-Knowledge condition wrote significantly more text and complete words in their daily entries.

No Difference in Perceptions of Mood Prediction Accuracy: Given the presence of an extreme outlier, the Mann-Whitney U test was used to compare perceived accuracy ratings across the two experimental conditions. The test showed no significant differences in perceived prediction accuracy across the two conditions ($U=220.5$, $z=-1.609$, $p=.108$).

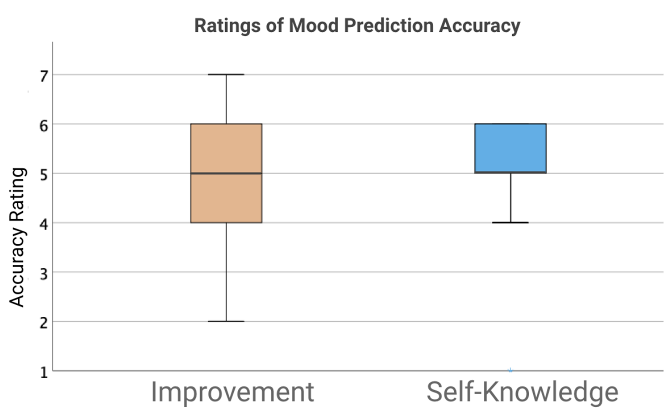


Figure 25. Distribution of ratings for how accurate participants considered their mood predictions to be with higher scores indicating greater perceived accuracy. A Mann-Whitney U test showed that there were no significant differences in accuracy perceptions across the two conditions.

Explanatory Mechanism for Engagement: Self-Knowledge Participants Rated Themselves More Successful

As predicted by Objective Self-Awareness (OSA) Theory, people may be more likely to engage with self-relevant information if they consider themselves close to or exceeding a standard (Davis & Brock, 1975; Duval, Wicklund & Fine, 1972; Greenberg & Musham, 1981). To account for this as an explanatory factor for condition differences in engagement, in the post-test survey, Self-Knowledge and Improvement participants were asked to rate on a 5-point scale the extent to which they considered themselves successful or unsuccessful at achieving the aim outlined in the application (i.e., *Perceived Success*). A Mann-Whitney U test shows that participants in the Self-Knowledge condition considered themselves largely more successful (Median: 4, Mean: 3.68, SD: .67) than Improvement participants (Median: 2, Mean: 2.29, SD: .91), $U = 85.5$, $z = -4.785$, $p < .001$.

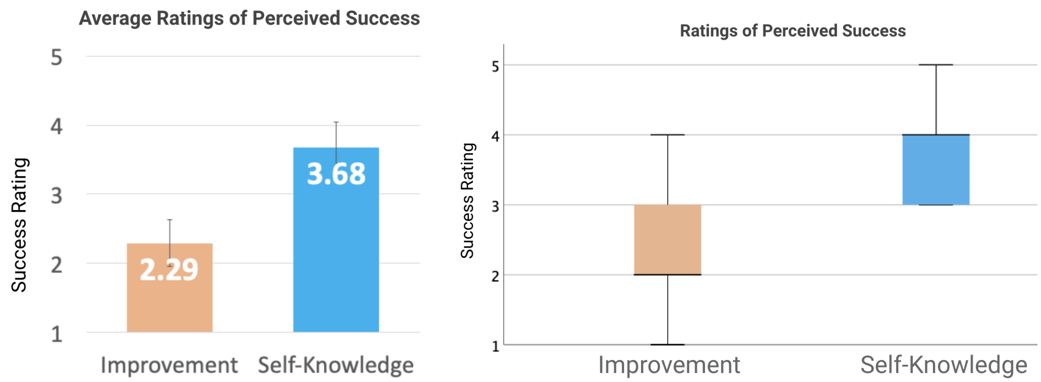


Figure 26. Ratings of *Perceived Success* were taken in the final survey on 5-point scales with higher scores indicated greater self-perceived success at achieving the aim of the system. Participants in the Self-Knowledge condition rated themselves as more successful at the end of the study. Error bars in the leftmost image represent ± 1 standard error.

A mediation analysis was performed using an ordinal logistic regression to assess impact of condition on *Reported Data Engagement*, when controlling for *Perceived Success*. Results show that differences in *Perceived Success* fully mediate these condition differences in *Reported Engagement* (Wald X^2 = .339, b=-.401, 95% CI [-1.753, .950], p = .561). We see this mediation also occur for impact of condition on the ratings of system *Adoption Interest* (Wald X^2 = .099, b=-.214, 95% CI [-1.551, 1.123], p = .754) and likelihood of *Continued Use* (Wald X^2 = .062, b=.163, 95% CI [-1.126, 1.453], p = .804). A similar test was conducted for *Text Length* of entries again showing that differences across conditions in this behavioral measure of engagement were mediated fully by *Perceived Success* (Wald X^2 = .323, b=-.368, 95% CI [-1.639, .902], p = .570).

Experience Freewrites: Self-Knowledge Participants Most Positive

Participants in the post-study survey completed a free-written response to the question "*How was your experience with using this app to track your mood?*". Two separate researchers who were not the author and were blind to the conditions individually rated the valence of participant freewrites as 'Positive', 'Neutral/Mixed', or 'Negative'. There was good agreement between the raters (Cohen's kappa =.732, 95% CI [0.571, 0.893], $p<.001$). An average of the two ratings was computed and these averages were used in a Mann-Whitney U test to compare differences in valence of free writes (i.e., *Experience Freewrites*) between the Self-Knowledge and Improvement conditions. Participants in the Self-Knowledge condition were found to be significantly more positive in *Experience Freewrites* than participants in the Improvement condition ($U = 250.5$, $z = -2.192$, $p = .028$).

A mediation analysis was performed again using an ordinal logistic regression to assess impact of condition on *Experience Freewrites*, when controlling for Perceived Success. Results show that differences in *Perceived Success* fully mediate these condition differences in *Experience Freewrites* (Wald $X^2 = .366$, $b=.491$, 95% CI [-1.099, 2.081], $p = .545$).

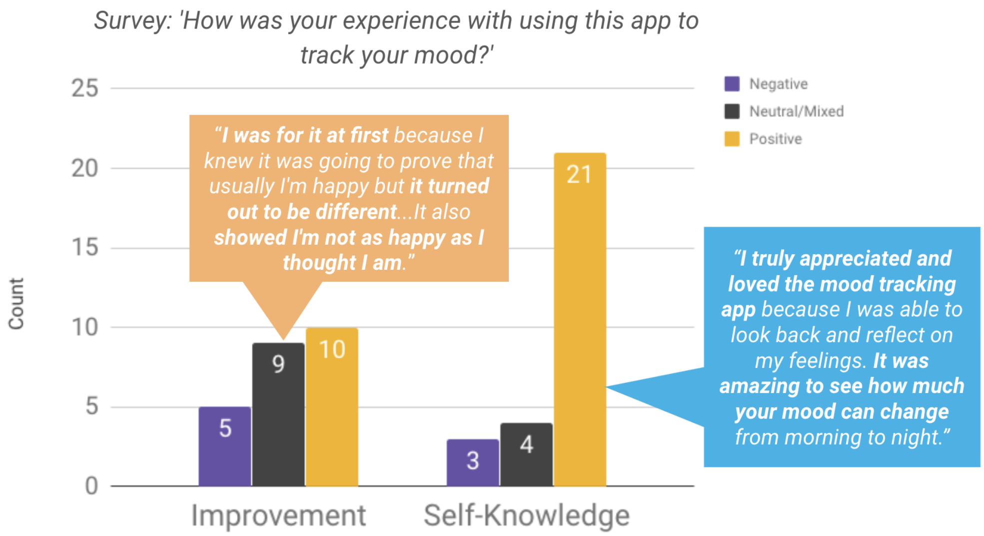


Figure 27. Two human coders who were not the author and were blind to conditions individually coded participant freewrites as 'Positive', 'Neutral/Mixed', or 'Negative'. An average score between two coders was used for statistical tests, showing that participants in the Self-Knowledge condition were significantly more positive about their experiences in using the application to track their moods. For ease of interpretability, results from one human coder are shown here to display the distribution of 'Positive', 'Neutral/Mixed', and 'Negative' freewrites of participants across the conditions.

ID	Quote	Rater 1	Rater 2
P71	*I think it's fantastic. I can go back and remember things because I wrote them while I was feeling a certain way. I think the app is very clear and easy to use. I'd like to have my friends use it.*	**Positive**	**Positive**
P46	*I never usually reflect on my mood but the mood tracker forced me to do so. I realized I am a much more negative person than I thought I was. I also realized the people I am surrounded by have an extreme impact on my mood.*	**Neutral/ Mixed**	**Neutral/ Mixed**
P45	*Tracking my mood twice a day felt more like a chore than anything. I only felt happy about tracking my mood when I was in a good mood. The app was difficult to use on iOS mobile.*	**Negative**	**Negative**

Table 9. Examples of 'Positive', 'Neutral/Mixed', and 'Negative' text free writes from Study 2. Conditions are denoted with participant ID with "I" standing for those in the Improvement condition, and 'S' for those in the Self-knowledge condition.

Qualitative Dataset and Analyses

Fifty-two participants submitted free-written responses in the closing survey about their experiences in using the mood-tracking applications. They also described anything they particularly liked or disliked about their experiences in two subsequent text fields. Participants from both the Self-knowledge (n=7) and Improvement (n=9) conditions also volunteered to additionally complete remote interviews spanning approximately 45 to 60 minutes. Interviews were conducted over video conferencing software with support for screen sharing. The interviews covered a range of questions regarding participants' experiences in using the application to track their moods, and their expectations in using such systems.

A bottom-up coding of survey freewrites was conducted by the author. Coding was first done line-by-line and themes were then related and collapsed into higher-level categories. To streamline the coding process for interviews, interviewees answered open-ended questions which were then followed with a mixture of closed- and open-ended choices in web forms that interviewees could view. This process enabled the author to code for themes in the process of interviews and gain agreement of interviewees in how their answers were coded. Themes are discussed in terms of the extent to which they are important to the overall research questions (Braun & Clarke, 2006). Prevalence of themes is noted by the number of participants who were reflected in the theme, but does not necessarily mean that the theme can be expected to be more prevalent in the general population.

These qualitative analyses allowed us to explore why these valence differences across *Experience Freewrites* might occur. Importantly, the qualitative analyses also allowed for a broader understanding of how participants viewed the applications, their described usage of app during the study, and general expectations for emotion-tracking systems.

Participant Survey Freewrites: Recollection of Study Experiences

To explore these differences further, the author inspected what aspects of their experiences participants described as liking or disliking in the post-study survey. These results showed a range of benefits such as considering the application motivating (n=4), an outlet to discuss negative emotions (n=3), feelings of reduced shame in the application (n=3), and encouraging behavior change (n=6). 11

participants also noted directly that they considered the app highly usable. There were also generally positive reactions to the data visualizations with both Self-knowledge (n=13) and Improvement (n=7) participants describing that they liked the data visualizations. Several participants discussed liking the app for providing an avenue to reflect (n=17) and a majority of respondents described moments of insight or increased awareness of one's emotions (n=33).

However, several issues were also noted in the free-written responses. In particular, multiple participants were critical about the inherent limitations of quantifying mood and limits of the scale (n=11). Participants across both conditions specifically voiced disliking that they discovered they are not as happy as they thought (n=7) while others were pleased to discover they were happier than expected (n=3).

<u>Benefits of Reflection and Insight:</u> A key difference across the two conditions is discussions around appreciating the opportunity to reflect. While 13 Self-Knowledge participants described liking the aspect of reflection, only 3 participants in the Improvement condition cited similar benefits. As P68 (Self-Knowledge) describes, they found value in reflection: "*It was good to write down and to seriously think about my day. **It helped me to reflect on the day**...I liked connecting all my experiences during the day to what my mood was at the moment*". Similarly, P59 (Self-Knowledge) discussed finding value in reflecting on all emotional states: "*...**to actually think about how I was feeling**. This **let me understand what kind of mood I was in**...Some experiences might be sad, but **you can see value in them**...*". This

noted benefit of reflection which seemed to occur with greater frequency for Self-Knowledge participants could explain these differences we see in their valence of free writes and also increased length of text entries.

Mixed Reactions to Implicit Goals: One key question is how participants responded to the goals reflected in the application, and potential for differing reactions to the same system descriptions. A key finding was that participants across both conditions had very different reactions to the goals outlined in the application. Some in the Improvement condition found the system motivating (n=3). As P51 (Improvement) describes: "*...[I] liked the percentage of my general happiness.* ***When it fell below 70%****... found myself making an effort...* ***I want to live a happy life, so seeing a lower number encouraged me*** *to bring it up.*".

However, eight participants across both conditions specifically described paradoxical well-being effects due to using the application such as P33 (Improvement) who stated "***I feel sad*** *that* ***I'm often in bad moods****...I didn't realize I was as unhappy as I am*". P76 of the Self-Knowledge condition similarly described negative reactions to attending to their emotions: "*...****I felt resentful****. I didn't realize how unhappy I was...*". Others described the app as evaluative "***I felt discouraged****...it feels like* ***I was letting the app down****...*" (P40, Improvement).

In contrast, some participants specifically described reduced feelings of shame or expectations to be positive (n=3): "*...great at identifying emotions without having any negative association…* ***without feeling shame at having felt 'negative' emotions***" (P53, Self-Knowledge). This highlights one important implication that we may not see counterproductive well-being effects due to the very different interpretations users have of the system designs. In addition, there were also cases where participants directly rejected the goals outlined in the application.

Rejection of System Goals: Similar to research in OSA showing that individuals can purposefully revise goals for which to compare themselves against (Dana, Lalwani & Duval, 1997), some participants described in the post-test survey revising the goals they had for using the system. For example, P35 of the Improvement condition described instead focusing on self-knowledge goals: *"...**I understand now why I feel the way I do and am not mad at myself** for feeling that way unreasonably...I found myself **using the app to understand myself** based on what I do, not really by comparing it to how I felt other days.*".

P49 of the Improvement condition similarly described later in the study shifting focus to personal well-being goals rather than the positivity ones prescribed by the system: "*Seeing the faces every day **made me want to try and improve my happiness score** or at least **allowed me to be a little more reflective** on the things going on around me that were decreasing my positivity. However, **towards the end of the study, I was actually happier in the sense that I appeared to be relatively balanced** in feeling positive, negative, and neutral and **feel way better about that then I do about whatever my average positive score was**.*"

This raises the interesting possibility that some participants nonetheless set their own goals in how they used the application rather than being heavily influenced by the prescribed system objectives.

Participant Interviews: Motivations, Goals, and Values that Should be Reflected in Emotion-Tracking Systems

Self-Understanding is a Top Motivation: Prior work has shown that a top primary motivations to adopt an automatic stress or emotion tracker is to understand one's self better (Hollis et al., 2018). We see this result replicated here in the context of discussing a manual mood-tracking system after respondents have had direct experience of using such a tool. When asked to identify up to three of the best uses of a technology to track emotional states, interviewees most often identified '*to understand myself better*' (e.g., perceived versus actual mood fluctuations, n=12) and '*to understand how this factor relates to others*' (e.g., to understand how sleep might impact mood, n=9).

P77 (Self-Knowledge) described that they found self-understanding a key use of emotion-tracking systems as they could provide a safe outlet to reflect on emotions: "*...it's something generally people don't do...especially I think for men, I'm a believer in toxic masculinity, and I feel like some people haven't had the space to reflect on their emotional states or have been discouraged to do so…*". In other cases, participants described self-understanding as a necessary requisite to achieving personal emotion goals "*... you can really understand what certain things might cause your moods. Like I was saying, the irritability and stuff like that. Sometimes you don't want to sit down and actually say what's actually causing it, and I think that something like this can really force you to do that, which is a good thing*." (P78, Self-Knowledge).

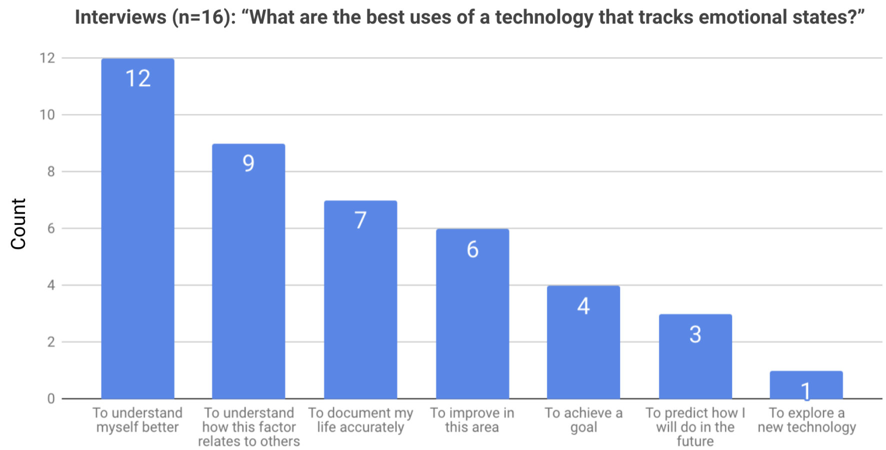


Figure 28. In a closing interview, 16 interviewees were asked to identify up to three of the best uses for an emotion tracking system. The majority of interviewees selected uses related to improved self-understanding or identifying relationships among data.

In the interview, interviewees were also asked to freely describe particular goals and system features they would want reflected in an application to track emotional states. Interviewees were then presented with a list of seventeen different closed-ended options and could include up to three open-ended option. They were allowed to skip options that they did not want to discuss. For each of the options, they could address whether the option should be reflected in an emotion-tracking system and why.

<u>Preference for Self-Understanding:</u> We again see here a large emphasis on expectations that an emotion tracking system should support identifying triggers of moods (n=14) and ensuring users have more accurate information about themselves (n=13). No participants thought that an emotion-tracking system should hide negative information, and many indicated instead that a system should share all information equally regardless of impact on mood (n=12). As P33 (Improvement) describes,

showing all information can lead to learning benefits and improve one's ability to regulate themselves: "*...it should show positive and negative ones...**only showing positive stuff. It seems like pretty avoidant**...praising somebody and then **they don't ever learn** how to deal with anything...*".

Split Views on Emphasizing the Positive and Goal-Setting: There was an interesting split on perspectives around whether applications should help users control their emotions (n=6), maximize pleasure (n=5), set personal emotional goals (n=9), help users have more balanced moods (n=8), or reflect specific goals to make users more positive (n=7). This difference in participant responses around goal-setting and positivity reflects a larger issue around the values embedded within self-tracking applications and the variety of users these tools serve.

Participants who considered positivity and goal-setting as concepts that should be reflected in these applications discussed how these tools could be motivating and encourage beneficial behavior change. As P43 of the Improvement condition explains with respect to the logging interface: "*...I would look forward to seeing 'How positive are you overall'...it would say 'today, you are feeling less positive than usual'....help me compare how I was today like usual then I would think...**what did I do today to make myself less happy** or more happy?...**Our goal should be to be as happy as possible***".

Similar to findings from the survey free writes, some interviewees also considered the positivity statistics in the Improvement condition as encouraging and did not recall seeing negative statistics in the interface: "*I liked that it prompted you* *or **congratulated you** when you were sort of above average or **if your mood improved**...liked that it gave you that metric...**I don't recall if it did when you had a negative mood**...I like when you were positive and it encouraged you to stay positive...*" (P49, Improvement). In contrast to other participants in the Improvement condition, P49 instead interpreted the system as one that did not '*put you down*' for having negative moods and instead described how "*...it's really important to allow yourself to feel negative if you're feeling negative. It's important not to dwell on being negative but feeling negative or allowing yourself to feel negative occasionally*". Echoing results of the survey free written responses, this shows the very different reactions users had in interpreting the system.

In contrast, others were critical of assuming positivity as the goal of the system. A majority of interviewees thought instead that the application should help users be accepting of all emotional states (n=13). As P38 (Improvement) describes: "*...**all emotions have value**...We should take the good and the bad as they come....it is highly **popular to think about being positive**...you're not going to be happy every single day of your life...an app that specifically says, 'use this to improve your mood and to make yourself happier'...it's not realistic...**it's to record how I feel and all emotions are valid emotions**, not just the high and low ones…*".

In conclusion, the qualitative results highlight the wide range of expectations users have for emotion-monitoring technologies. Interviewees generally showed a high interest in using emotion-monitoring systems to better understand the self, and there was a clear split in preferences across whether systems should reflect particular emotion goals. They also suggest that at least a subset of users reinterpreted system goals to fit their own perspectives, or knowingly override the goals presented in the system.

Study 2 Discussion

Study 2 replicated and extended Study 1 in the context of examining well-being and engagement consequences of emotion-monitoring systems which reflect different implicit goals. Unlike prior work which has demonstrated lower reports of positive emotion when given primes to value happiness (Mauss et al., 2011), Study 2 shows instead no differences in daily and reported well-being measures across the experimental conditions. This result is illuminating as while there has been a wealth of compelling in-lab experimental research showing the risk of counterproductive well-being effects, yet we see that these effects are less likely to occur in this real-world application.

Similar to Study 1, there were also reported and behavioral differences in engagement with Self-Knowledge participants reporting greater interest in adopting the system, reported greater frequency of viewing their data, and submitting longer text free writes in daily entries. Self-Knowledge participants were also largely more positive about their experiences. Again as with Study 1, Self-Knowledge participants also considered themselves more successful at the end of the study which explains the significant differences across conditions in engagement and positivity of their experience freewrites. These results are discussed in detail in the General Discussions section below.

General Discussion

This thesis addresses whether implicit goals in affective personal informatics systems impact aspects of well-being and system engagement. There was considerable consistency in results across both experiments. Participants who used PI systems that conveyed Self-Knowledge goals reported greater perceived success at the conclusion of the studies. Furthermore, these differences in perceived success seem to explain why Self-Knowledge participants had greater reported engagement. Self-Knowledge participants of Study 2 also showed increased behavioral engagement through length of daily entries and were most positive about their study experiences. These findings are consistent with predictions from Objective Self-Awareness (OSA) theory (Duval et al., 1992; Duval & Wicklund, 1972; Silvia & Duval, 2001), that feeling close to or exceeding standards results in greater engagement with self-relevant information. For well-being, we replicated prior work showing that dispositional differences in valuing happiness (Study 2) or negative beliefs about stress (Study 1) are associated with negative emotional consequences (Catalino et al., 2014; Fischer et al., 2017). However, we did not replicate the result that experimental primes to value happiness would paradoxically result in decreases in positive emotion (Study 2).

These studies suggest several design and theoretical implications for Personal Informatics systems. One key issue for PI systems design concerns motivational barriers which can lead to participants avoiding self-tracking data. Persuasive design is popular in PI tools (e.g., goal-setting, social comparison, self-evaluations; Cheng, 2003; Consolvo et al., 2009; Lin et al., 2006). This thesis is the first to test whether implicit improvement goals, a persuasive component of some PI systems, can undermine system engagement and emotional well-being. These results also have especially clear design implications for affective PI systems. Study 2 survey freewrites showed that the positivity of participant's reactions to these systems differed based on the described goals for these tools. Interviewees of Study 2 articulated a diversity of perspectives on the types of ideas they would want to be reflected in technologies to track emotion with the majority discussing the value of using emotion-monitoring systems to understand the self better, yet there were clear differences in beliefs that these systems should reflect particular goals for emotion (such as positivity or personal goal setting).

Well-Being: Study 1 showed that negative pretest beliefs about stress were associated with greater reports of immediate and general stress. In Study 2, participants who valued happiness reported experiencing fewer positive emotions, a greater number of negative emotions, and lower happiness at pre-test. Although correlational, these relationships confirm prior work that particular values and beliefs about emotion have clear ties to well-being (Fischer, Nater & Laferton, 2016; Ford & Mauss, 2014; Mauss et al., 2011). Study 2 is the longest-running experiment to test whether the counterproductive effects of valuing happiness are demonstrated in the context of exposure to daily primes. In-lab research has demonstrated ironic decreases to well-being when participants are encouraged to value being positive (Ford & Mauss, 2014; Mauss et al., 2011; Schooler et al., 2003). However, this priming effect is not replicated here.

Why did we not replicate these prior findings? There are several potential reasons for why these results differ from prior work which could be attributed to the current research design choices to increase the ecological validity of Study 2. First, prior work has tended to restrict sampling to university populations and in some cases only sampled female participants (Mauss et al., 2011). Study 2 instead involved a range of self-identified genders and ages, with recruitment across the United States. The current study's well-being results could then depart from prior work due to simple differences in sampling. Participants in Mauss et al. (2011) and Schooler et al. (2003) may have been more susceptible to priming effects than those in Study 2, or more susceptible to primes due to direct researcher observations of an in-lab study. Second, well-being measures of Study 2 involved general assessments of subjective happiness and recall of mood frequencies over the past three weeks, rather than the short-term emotion ratings used in Mauss et al. (2011) and Schooler et al. (2003). These general well-being measures could be less sensitive to the influence of primes. This highlights the possibility that primes to encourage valuing happiness are short-lived or have relatively weak effects for general, rather than state, measures of emotion. Finally, given the condition differences in engagement across participants, one alternative explanation is that participants in the Improvement conditions disengaged with system information, reducing opportunities for them to be negatively impacted by goal priming.

The self-knowledge conditions involved system descriptions that suggested users accept all emotion outcomes equally (Study 2 pilot). By invoking acceptance, we intended to downplay specific emotion ideals and increase differences across the experimental conditions. Interviewees of Study 2 also discussed aspects of emotional acceptance in their use of the system, sometimes purposely rejecting goals for increased positivity in lieu of greater self-acceptance ("*...understand now why I feel the way I do and am not mad at myself for feeling that way unreasonably...understand myself based on what I do, not really by comparing it to how I felt other days..."*). Some interviewees later discussed hypothetically that systems should be designed to reflect emotional acceptance, speculating that such increased acceptance could aid self-understanding ("*if something is pushing for you to be more positive, it could bias you against understanding your negative emotions...if you're just trying to be positive all the time...you won't understand yourself as well*") or emotional health ("*they don't want to accept that emotion...people really need to be able to accept their moods so that they can really just work through it...they'll begin to accept what they're feeling at the moment and move on…*"). More work is needed to understand what acceptance-based designs for PI look like in practice and the tradeoffs of implementing such designs in contrast to more traditional focus on specific goal striving. Others have shown acceptance-based interventions can be valuable for improving well-being in therapeutic contexts (Hayes et al., 2002). We encourage future system designers to consider the role of acceptance-based designs in the implementation of PI tools.

Goal-Setting and Self-Monitoring: Across psychology and HCI there has been a wealth of work showing the benefits of monitoring (Burke et al., 2011; McFall et al., 1970; Whittaker et al., 2016), goal-setting (Konrad et al., 2015), and daily emotional reflection (Hollis et al., 2015) for goal success. Yet participants in the Improvement conditions of Study 1 and Study 2 showed no apparent benefits to achieving goals related to increased positivity or reduced stress. The lack of a benefit is peculiar, as behavior change systems are increasingly deployed with the promise of improvement through self-monitoring. The discrepant results of this thesis are consistent with a general observation that monitoring emotions alone does not result in well-being benefits (Conner & Reid, 2012; Faurholt-Jepsen et al., 2015; Hollis et al., 2017).

One reason this lack of well-being benefits might be due to how emotions are operationalized in PI systems and measures of well-being outcomes. Popular commercial PI systems tend to focus on concrete, discrete behaviors (e.g., steps walked, number of cigarettes smoked, calories consumed), and as a result can have more precise indicators of successful versus unsuccessful days (e.g., Hollis et al., 2015). However, there are recent theoretical concerns about whether emotions can be considered discrete categories (Feldman-Barrett, 2017). Some participants in Study 2 requested additional granularity in the app for tracking emotional states, and such preferences in how emotions are operationalized could vary across users. As a result, the tracking and categorization of emotions in a standardized way for a PI system is a difficult problem in and of itself. This is unlike other measures of PI system outcomes which may involve clearer tracking metrics.

Also unlike other common PI systems for health or productivity, participants of Study 2 discussed a variety of emotional outcomes they wanted to achieve such as more positive moods, accepting their negative emotions, or experiencing less irritability. Given there is no uniform definition of success for emotion outcomes, this may be why we do not see general well-being benefits of monitoring. This breadth and complexity of emotion outcomes users want to achieve in affective PI could starkly contrast against other PI interventions focused on simply goals of reducing (Hollis et al., 2015) or increasing single metrics (Konrad et al., 2015). We now explore theoretical implications concerning self-awareness, information avoidance, and system deferral:

Personal Informatics and Self-Awareness: This is the first study to explore whether objective self-awareness theory (Duval et al., 1992; Duval & Wicklund, 1972; Silvia & Duval, 2001) predicts how personal informatics users engage self-awareness and how such (dis)engagement might occur. Across both studies, when participants perceive a failure to meet a system standard, rather than reject system accuracy (Arkin & Maruyama, 1979), they instead tend to disengage with the data and system. Consistent with the predictions of OSA theory, there were multiple ways participants disengaged with the system when they felt unsuccessful: they reported lower engagement and interest in seeing their data, along with lower reported interest to adopt the system. That we do not see differences in perceptions of accuracy follows findings of prior work with emotion-sensing and analytics (Springer et al., 2017; Hollis et al., 2018; Hollis et al., under review) on the tendency to trust interferences from PI systems, leading to the next point of how results in this thesis connect with prior work on authority effects.

Authority effects: Rather than discredit system accuracy, our results instead indicate that users are more likely to disengage with the system. But how do these findings align with system deferral effects demonstrated in previous research (Hollis et al., 2018; Hollis et al., under review; Springer, Hollis, & Whittaker, 2017; Warshaw et al., 2015; Snyder et al., 2015)? Past work has shown that users are highly susceptible to rely on system outputs which can influence the perception of one's emotions (Hollis et al., 2018; Snyder et al., 2015), trust in the system (Springer, Hollis, & Whittaker, 2017), and beliefs about the self (Hollis et al., under review; Warshaw et al., 2015). However, when in consideration of systems assigning *goals* to users we do not see authority effects to the same extent. Together, these results suggest that users of PI systems may be more susceptible to deferring to PI systems for how they perceive the self (e.g., "*...could tell me about an emotion I don't know that I am feeling...*"; Hollis et al., 2018) with close ties to findings in self-perception theory (Bem, 1972). However, users of PI systems seem more skeptical about the appropriate responses to such data, as they showed different reactions to ideal levels of happiness or the types of goals that should be reflected in PI systems (Study 2). In short, this suggests that users engage with PI systems to understand their current state, but make their own judgments for what they should actually do with that information.

Implications for Design: A clear take away is that implicit goals can influence user feelings of success with subsequent consequences for reported interest in adopting the system (Study 1 and 2), reported engagement with data (Study 1 and 2), and how positive they are of their experiences in using such applications (Study 2). While seemingly intuitive, multiple real-world commercial applications suggest implicit goals for users to strive towards. One risk for such systems that emphasize improvement is that they can reduce engagement by setting expectations that users may not feel capable of achieving. If we relied only on qualitative data to gauge reactions to the stress- or emotion-tracking systems, it would have been easy to overlook the role of system goals in impacting user experiences. Only a fraction of participants explicitly described the goal of the system as having a positive or negative influence on their experiences, instead sometimes discussing superficial aspects of the UI in their judgments of what they liked or disliked. As a caution for practitioners, the impact of implicit goals on users may not be readily determined through qualitative results alone. This issue may explain why assumptions of user goals and idealized outcomes are pervasive in existing commercial systems (Spiel et al., 2018; Epstein et al., 2017).

A deeper understanding of user goals and the types of values reflected in self-tracking systems fits within a broader discussion of value-sensitive design (Friedman, Kahn, & Borning, 2002) and normative approaches to the development of self-tracking tools. The interviews in Study 2 showed that interviewees were negative about features associated with competition, while a majority of interviewees wanted emotion-monitoring tools to encourage greater understanding and acceptance. However, there was a distinct split across interviewees about whether they wanted systems to value positivity and promote specific goal seeking. An implication for PI design is therefore to ensure such systems align with user goals or support personalization, rather than being overly prescriptive about the types of goals users are expected to adopt as these could create feelings of low success.

Limitations: There are limitations of this work which provide opportunities for further understanding how implicit goals impact engagement. These systems were tested in a research context with external incentives for class credit (Study 1) or gift card credits (Study 2). Participants enrolled in these studies may also have had a low personal interest to use these tools compared to those who might elect to adopt these systems of their own free will. It is unclear to what extent such external pressure influenced results for engagement, user behavior, and perceived accuracy of the system. These issues bolster the argument for testing the impact of implicit goals in public commercial applications.

A consideration for these results also is that emotions are a somewhat specialized domain of self-tracking. While stress-monitoring is a top user interest for users of PI systems (Hollis et al., 2018), the results in this thesis may not entirely generalize to personal informatics as a whole. The issues of implicit goals were partly explored because we saw such striking differences to adopt tools for emotion-monitoring to better understand the self, in contrast to the common motivation of specific goal achievement in systems for health or productivity (Hollis et al., 2018). This thesis partly set out to contrast systems that reflect these alternative underlying motivations. However, it is unclear to what extent encouragement concerning self-knowledge, rather than specific ideals, in other domains would also produce changes in perceived success and engagement. OSA theory suggests we could see the effects of this thesis replicated in other domains of self-tracking, but there are multiple confounding factors such as goal commitment, trust in system accuracy, and perceived self-efficacy. Understanding how implicit goals impact feelings of perceived success and engagement is a crucial area for personal informatics as a whole so that users necessarily engage their data for helpful insights.

Future Directions: This work identifies different motivations that can be reflected in affective PI systems. Interviewees of Study 2 described self-understanding as a top motivation, replicating results of prior work (Hollis et al., 2018). An open question is *why* we see this difference in primary motivations for emotion-tracking in contrast to systems in the health or productivity domain (Hollis et al., 2018). Furthermore, this also leaves open questions of whether these primary motivations are static or a reflection of a larger progression of self-tracking motivations. One possibility is that users of emotion tracking systems are in the 'diagnostic' phase of self-tracking, determining their goals and how to achieve them (Li et al., 2011). In contrast, those who use tools for health or productivity may already consider themselves aware of the problems and relationships among their data, being more concerned with tracking goal progress (i.e., 'directive' phase; Li et al., 2011). Alternatively, stages of the transtheoretical model (Prochaska et al., 1997) could explain these differences in underlying motivations to track, again with consideration for those who are in the action to maintenance phases against those in earlier contemplation stages. If such motivations are a progression, systems could reflect the different needs at each of these stages for emotion monitoring. Future longitudinal research is needed to understand how motivations to track potentially evolve over time.

There are also interesting future questions about how goals prescribed by the system influenced feelings of intrinsic versus extrinsic motivation. Interviewees of Study 2 described both aspects of motivation with respect to the Improvement framing. Some consider the system overly prescriptive and controlling (e.g., "*I didn't wanna feel like I was working...*") while others interpreted the same system design as fostering the personal goals they had (e.g., "*...it was my personal goal. I wanted it to be a more positive person.*"). Similarly, a participant in Study 1 described the stress-tracking system as '*patronizing*', suggesting that '*...if i implemented it on my own terms I would have a different feeling about it...*'. This leaves open the question of the extent to which the prescribed goals of the system impacted feelings of autonomy and personal motivation. This is a key question to consider as some commercial PI applications promise external rewards to use the system (e.g., Mango Health). An area we encourage for future work is how intrinsic and extrinsic motivation are impacted by the goals reflected in PI system designs, and how such differences influence engagement.

Another area of potential future research is understanding how these results might differ if we give users the option to choose system goals. This would be a more ecologically valid study, because users of commercial PI systems have the ability to select among multiple applications. By allowing users to choose which goals or values they want to be reflected in the system, we might expect to see differences in perceived success. Users could choose goals that feel more achievable, or have greater engagement with the system due to opportunities for personalization (Kang et al., 2017).

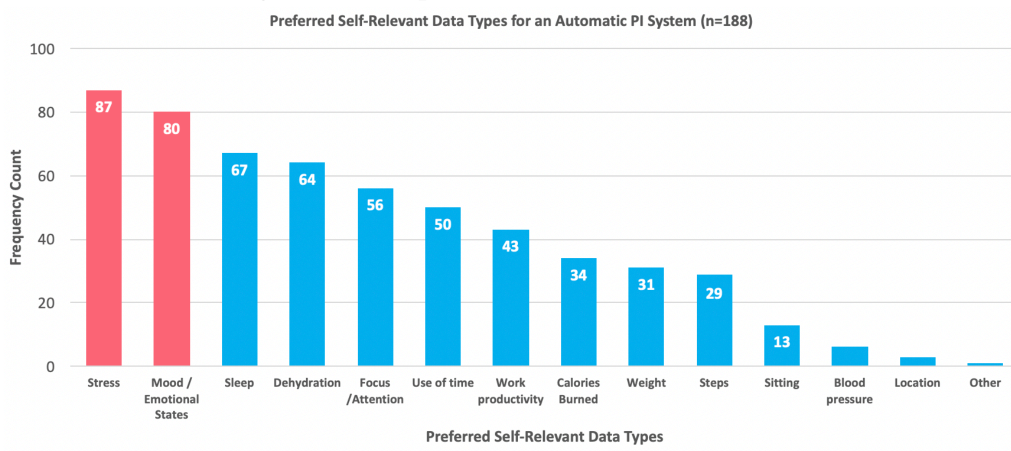
Conclusion

This thesis details two experimental studies for how goals embedded within stress- and emotion-tracking systems can have consequences for user well-being and engagement. We conducted a multi-stage design process to develop alternative versions of working affective PI systems. These systems were piloted to convey system goals for improvement towards particular emotion ideals or instead more accurate self-knowledge. We see across both studies that participants who used systems that reflected self-knowledge goals showed greater ratings of perceived success at the conclusion of the studies. These participants also showed increases in some measures of engagement and were more interested in adopting these the PI systems for daily use. We encourage system designers to consider whether the goals embedded in PI systems match those of the users, and to what extent such implicit goals can impact feelings of success. More work is needed in addressing the types of goals and values subtly conveyed by self-tracking systems and the impacts they have on users. Only recently has there been discussion in HCI of normative ideals conveyed in the context of health tracking, such as definitions of fitness as steps walked, or assuming that all users desire weight loss. Studies 1 and 2 add to this growing dialogue of the types of values and goals currently conveyed in self-tracking systems. We encourage conscious decisions about the types of values and goals that should be reflected in affective PI systems.

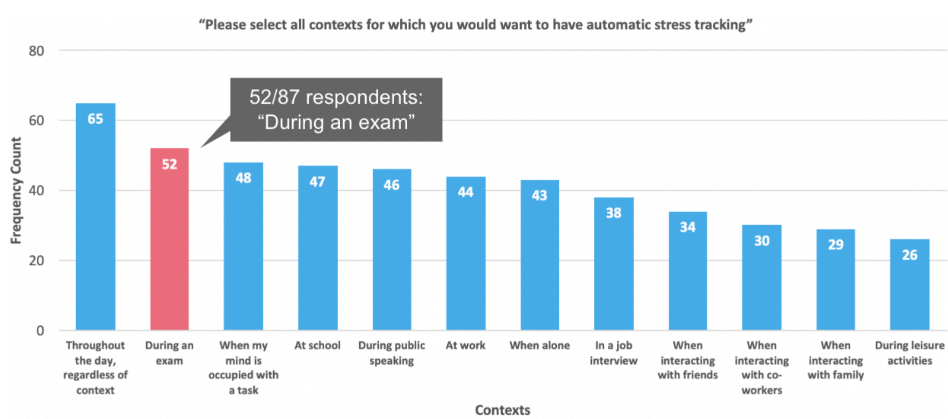
Appendices

Appendix 1. Motivations and expectations for automatic PI (Hollis et al., 2018).

1. Survey results from (n=188) UCSC student respondents on the self-relevant data types respondents want automatically collected by a personal informatics system. Each respondent could select up to 3 choices. The top interests for automatic tracking in this sample were for stress and emotional states.



2. Top contexts in which UCSC student survey respondents would want automatic stress tracking



Appendix 2: Study 1: Test Instructions and Overview

In this study, you will answer 20 math problems.

These problems may include some questions that will be potentially more challenging. Please remember that you are free to stop the study at any point and you will still receive full credit if you do so.

1. For the test, you will see a series of multiple choice or multi-select options labeled A, B, C, D and so forth.
2. You will have the option to use scratch paper to answer each question. Once you have decided your final answer, select that answer or set of answer options on the screen.
3. When you have decided your final answer, you must announce your final answer aloud to the experimenter. **You cannot change your answer after announcing it.**
4. The experimenter will give you verbal feedback on whether that answer was correct or incorrect so that you can keep track of your performance
5. You must then click the Next button ('>>') on the screen
6. There is a strict time limit to answer the questions. The amount of time remaining will be displayed on each page.
7. You must do calculations by hand and will not have access to a calculator

Please try your best to answer these questions correctly and quickly.

Appendix 3. Study 2: Online Advertisement.

TITLE:
REMOTE Participants invited to an online research study about daily emotions & behaviors

BODY:
Interested in learning more about your daily emotions & behaviors? Want to try a novel self-tracking technology?

University of California Researchers are looking for participants to try a new web application that is designed to provide analytics about a person's emotions and how those relate to their behavior. In this study, you will use a phone and web application to track your mood twice a day for 21 days. You will also be asked to submit two surveys (approximately 15 minutes each), and optionally participate in a remote interview (up to 60 minutes).

Receive compensation for time spent (up to a $25 Amazon gift card and entry to a $50 Amazon gift card raffle), and the opportunity to learn more about your emotions.

Participation in the study is not required to join the raffle.

To participate you must satisfying the following:
- 18 years of age or older
- Resides in the United States
- Have a laptop or smartphone
- Do not have a history of serious depression, mental illness, or life-threatening disease.

To join this study or address any questions you have, contact Victoria Hollis (lead researcher) at hollis@ucsc.edu.

IRB Number: HS3130

Appendix 4. Study 2: Onboarding Instructions.

Improvement Condition

Thanks again for joining the beta-test. Please enter your participant ID below. You will complete some surveys about yourself and then review instructions for the app.

You should have received an email with your ID and app password. If you are having trouble finding this, please contact Victoria at hollis@ucsc.edu

This form typically takes between 5-7 minutes to complete.

Enter your participant ID:

HCI Research Participation - Welcome to the MoodEnhance beta-test!

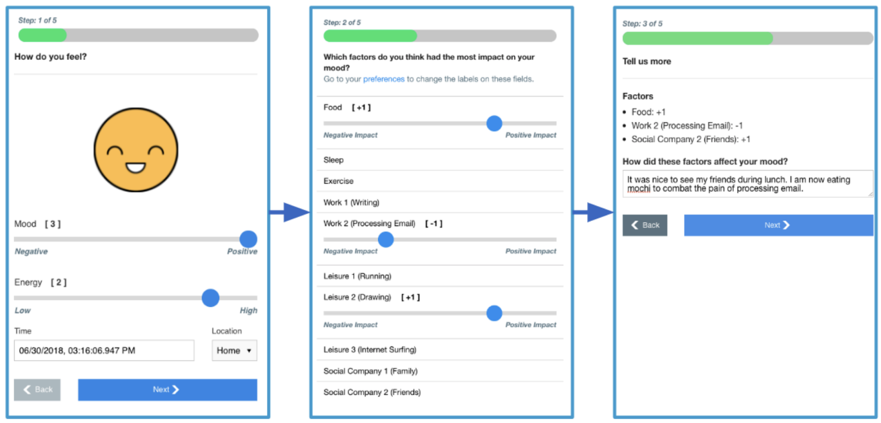
Please read through this text and the following screens for app instructions.

Thank you for signing up to try the MoodEnhance web application! We are researchers from the University of California, Santa Cruz who are researching new ways to improve health and well-being. We have designed the MoodEnhance web application to help you improve your mood and ultimately become much happier. Below is a brief description of what to expect and the attached instructions document has more detailed information about MoodEnhance.

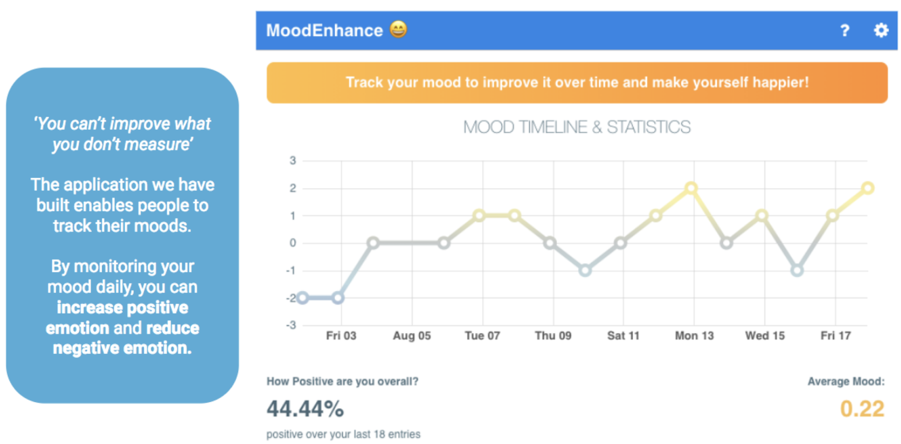
- **What you will do:** You can access MoodEnhance on either a laptop or phone and use the application to track your mood and factors that seem to trigger your mood.
- **Length of Participation:** To participate, please record 2 entries per day for 21 consecutive days.
- **Technical Issues?** MoodEnhance is supported on both mobile phone and web browser. We recommend using either the Chrome or Safari web browser to access MoodEnhance.
- Please review instructions and confirm that you have read these instructions at the end of this form.

We hope you enjoy MoodEnhance and please don't hesitate to contact me at hollis@ucsc.edu if you have any questions.

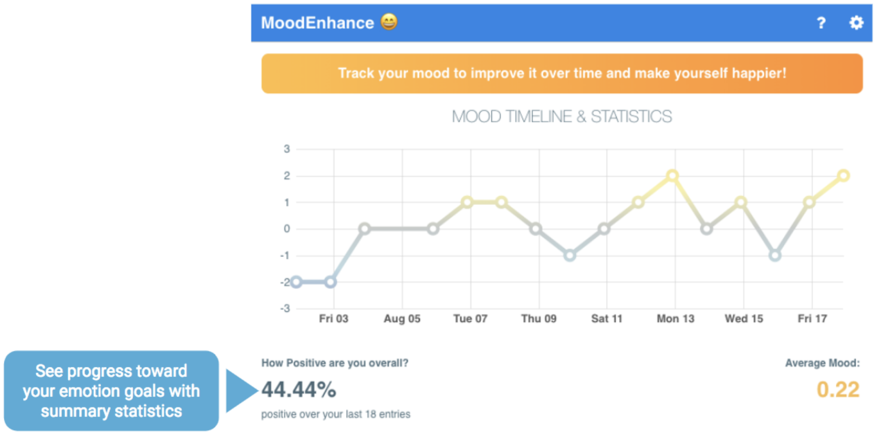
The mobile application we are building will let people track mood in their daily lives. Below is the logging interface. Here you can track your mood, energy level, location, and different factors that impacted your mood positively or negatively.



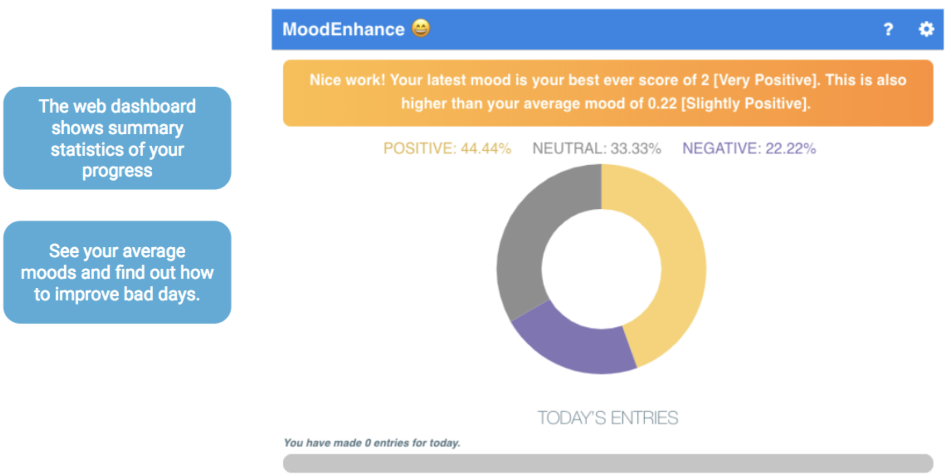
The app shows an overall summary of your mood reports over time. In this example visualization, you can see the average daily highs and lows for someone who has used the application for two weeks.



The app shows an overall summary of your mood reports over time. In this example visualization, you can see the average daily highs and lows for someone who has used the application for two weeks.



The application also provides summary statistics and analytics. By viewing these visualizations, you can track progress and find what makes you happy.



Please paraphrase, what do you think are the goals of this system?

____

Self-Knowledge Condition

HCI Research Participation - Welcome to the MoodRev beta-test!

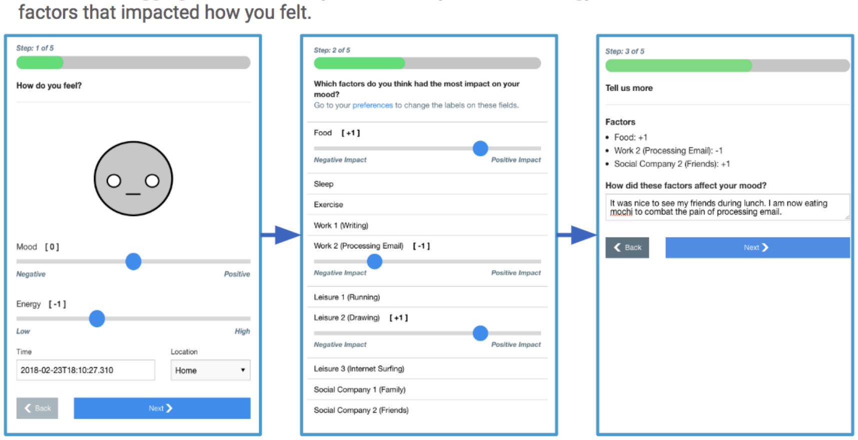
Please read through this text and the following screens for app instructions.

Thank you for signing up to try the MoodRev (Mood Revelation) web application! We are researchers from the University of California, Santa Cruz who are researching new ways to improve health and well-being. We have designed the MoodRev web application to help you learn more about your emotional life and better understand yourself. Below is a brief description of what to expect and the attached instructions document has more detailed information about MoodRev.

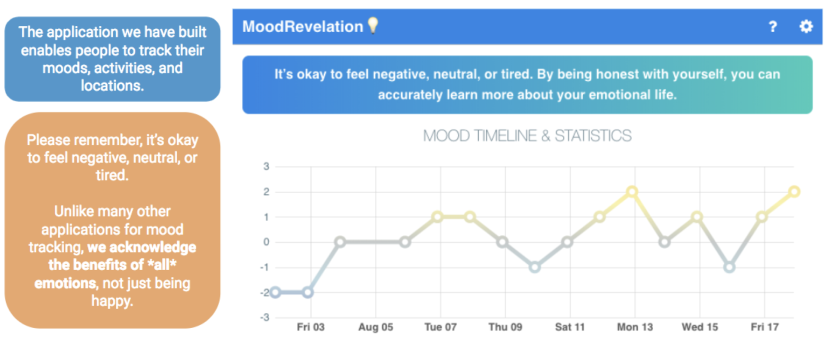
- **What you will do:** You can access MoodRev on either a laptop or phone and use the application to track your mood and factors that seem to trigger your mood.
- **Length of Participation:** To participate, please record 2 entries per day for 21 consecutive days.
- **Technical Issues?** MoodRev is supported on both mobile phone and web browser. We recommend using either the Chrome or Safari web browser to access MoodRev.
- Please review instructions and confirm that you have read these instructions at the end of this form.

We hope you enjoy MoodRev and please don't hesitate to contact me at hollis@ucsc.edu if you have any questions.

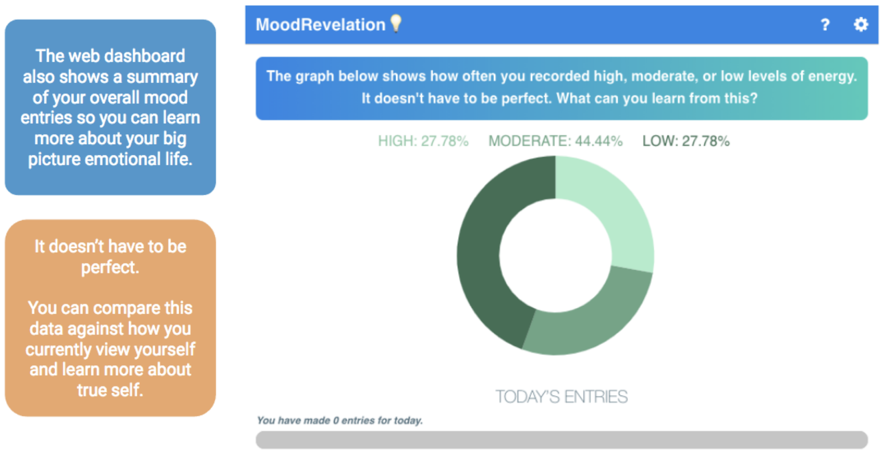
The mobile application we are building will let people track mood in their daily lives.
Below is the logging interface. Here you can track your mood, energy level, location, and different factors that impacted how you felt.

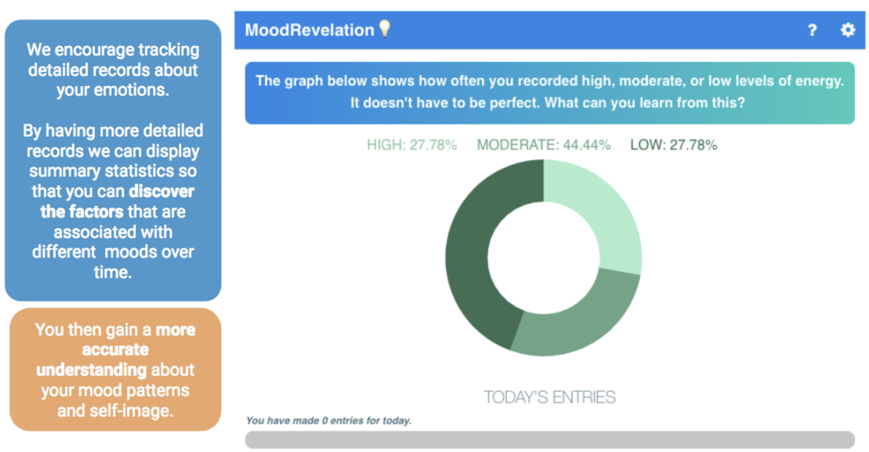


The app shows an overall summary of your mood reports over time. In this example visualization, you can see the average daily highs and lows for someone who has used the application for two weeks.



The application also provides summary statistics and analytics. By viewing these visualizations, you can gain an understanding of how your mood really is in daily life.





Please paraphrase, what do you think are the goals of this system?

___

Survey-Only Control Condition

Thank you for your interest in joining the app beta test.

Unfortunately, we cannot include you at this time. This is decided by chance. It is not based on anything about you personally, or the answers you provided.

We request that you complete an additional survey (similar to this one) in about 21 days. For completing both surveys, you will receive a $10 Amazon.com gift card ($5 card for this survey, and a $5 card for the final survey).

We hope that you are open to taking this survey again in 21 days and greatly appreciate your help.

Appendix 5. Study 2 Onboarding App Instructions (Self-Knowledge and Improvement Participants)

You can customize different factors to track over time.

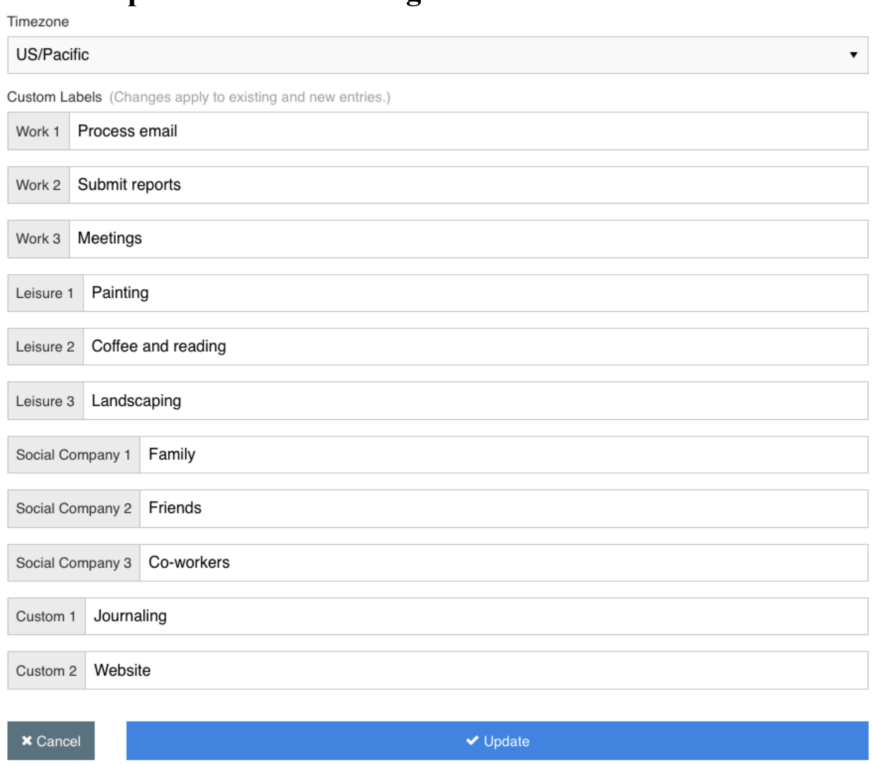
Once you have set the factors **please do not change them**. We need these factors to stay consistent so that the application works for you.

Steps:

1. In the upper right corner of the homepage, click on the settings icon.
2. Enter text for different factors that you might encounter on a regular basis.

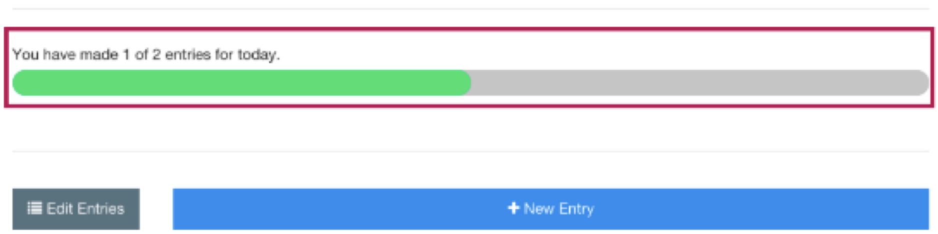
3. You can enter a specific name, or a category of activities/people. For example, with 'Social 1' you could enter a name 'Alex' or group of people 'Family'.
4. After you have set factors, please do not change these.

Screenshot Example of Custom Settings:



Create a new entry

Twice per day, you will need to create an entry about how you feel and what factors triggered your mood. The landing page shows how many entries you've made per day.



To make a new entry:

1. Click on 'New Entry'
2. Set the following fields:
    a. Mood - How you are feeling. You can set mood from +3 (very positive, happy) to 0 (neutral) to -3 (very sad, angry or negative)

    b. Energy Level - You can set energy level as separate from mood. You can rank your energy level from +3 (very high energy) to -3 (very low energy)
    c. Time & Date - You can change the time and date if you want to make an entry for a past event.
    d. Location - Select where you were when you felt this way

Recommended Platforms:
Android (4.4.2 or later) - We recommend using Chrome on Android. Versions of Android earlier than 4.4.1 may have missing features. If you have an Android 4.4.1 or earlier, you may prefer using Chrome on your laptop instead.
iPhone (iOS 7.1 or later) - You can use Chrome or Safari on your phone without any problems. Earlier versions may encounter issues.
Laptop/Computer - Chrome or Safari work best. Firefox, Internet Explorer and Opera will not display all of the app's features.

Go to [app link] to get started!

Contact: If you have any questions, feel free to contact Victoria Hollis, at [author's email].

Login Information: You can find your username and password in an email from Victoria [author's email].

(Optional) Please provide any feedback to improve this form:

Appendix 6. Study 2: Interview Questions.

Warm-up/Personal Informatics:

- To get started, I'm going to ask a few questions about your technology use and any self--tracking practices. So in general, are there any other apps you use to log your mood, aspects of health, work or anything else?
- Okay, and do you use any fitness trackers?
- Is there anything you do to track yourself manually? Such as writing down calories in food content, journaling your mood, or anything else that comes to mind?

General Experiences:

- Okay, thanks. Now I am going to ask some questions about your general experience. To start, why did you decide to join this study?
- Okay and in general, how was your experience with using the app to track your mood? We are specifically interested in the process of you tracking and reflecting on your mood, rather than details of the app itself.
- Did your experience of tracking your mood change over time?
- Was there anything you expected to gain out of tracking your mood? It is okay to say you didn't expect to gain anything.
- To what extent do you consider yourself successful or unsuccessful at achieving that over the course of the study?
- Okay and did you have any questions about yourself you wanted to answer? Why is that?
- Did you answer those questions?
- Okay, now if you could use a mood tracking tool again, perhaps with even more visualizations, are there any questions about yourself that you would want to answer? Why is that?
- Did you ever view your past entries? If so, what would trigger you to look? Why? [Note: These could be viewed by clicking 'edit entries' in the bottom left corner of the landing page]
- Did you ever view your graphs? If so, what would trigger you to look? Why?
- What, if anything, did you like about using this application to track your mood?
- What, if anything, did you dislike about using this application to track your mood?
- Okay, and in general, were there events or moods you tended to avoid recording in the app? [If yes, which and why]
- Alright, and were there particular events or moods your prioritized recording more often than others? [If yes, which and why]

Visualization Avoidance/Engagement:

- I am going to ask you a few questions about the the graph and pie chart you could see in the app. Do you remember looking at these? What did you think of them in general?
- Thanks, now I am going to pull up an example of the homepage and visualizations that you saw. This is example data and not from your records personally. [Click next to display an image with the landing page]
- What, if anything, did you like about the visualizations?
- What, if anything, did you dislike about the visualizations?
- Okay, and what did you think of the text prompts? For example, this text up here (top bar) and here (bar in middle of page).
- [If in the improvement condition] Okay, and what did you think of the statistics on how positive you are and your mood distribution?
- Were there any times you would **want** to look at anything on the landing page? This includes visualizations, summary statistics, or the text prompts. (If so, ask why is that?). Under what context would you want to look at that?
- Were there any times you **avoided** looking at anything on the landing page? This includes visualizations, summary statistics, or the text prompts. (If so, ask why is that?). Under what context would you avoid looking at that?

Mindsets and Beliefs

- What do you think was the designers goal in making this app?
- And what goals did you have in using this app, if any?
- Now imagine you could create your own technology to automatically or manually track your mood. What do you think would be the best uses of such a technology. You can select up to three options.
- Okay, so you selected [OPTION] Why is that? [Repeat for each option they selected, up to 3]
- Thanks! Now technologies for tracking mood can reflect different ideas about mood and emotions. For example, some technologies advocate for being more positive and avoiding negativity, others suggest that all emotions have value. What ideas about emotion should such a tool reflect?
- Okay, I am going to show you a couple different ideas about emotion these apps could reflect. [Click next and pull up list]
- Please tell me which ones you think SHOULD be reflected in these apps. Why is that? [Repeat for each option, click 'Yes' for each one they say]

- Alright, and now which ideas do you think should NOT be reflected in these apps? Why is that? [Repeat for each option, click 'No' for each one they say]
- And of these options which is the most important idea that SHOULD be reflected in the app?
- And of these options which is the most important idea that SHOULD NOT be reflected in the app?

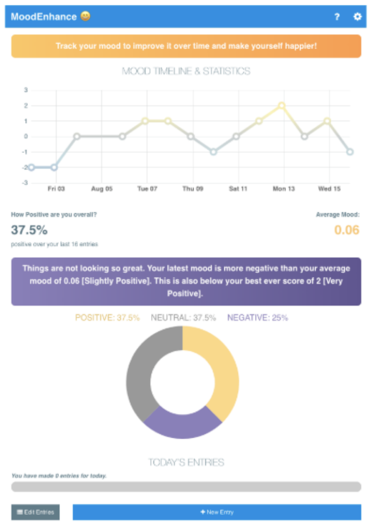
Closing

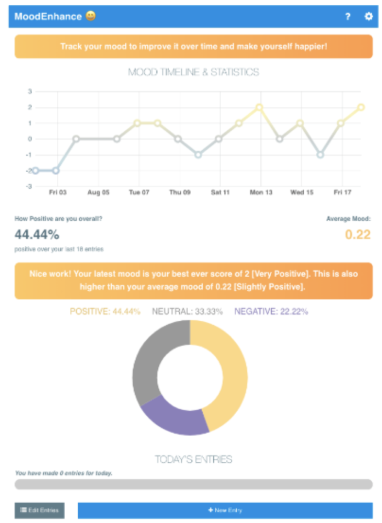
- Thanks for that! I am going to ask a few closing questions and then we are done for today!
- What was the most interesting thing we discussed today in relation to mood tracking?
- In terms of generally tracking your mood, what if anything did you find most useful?
- In terms of generally tracking your mood, what if anything did you find least useful?
- [insight] Did mood tracker or seeing your data today help you to discover something new about yourself that you were not aware of? It's okay to say you didn't discover anything.
- Okay, and on a 5-point scale with 5 being extremely likely and 1 being not likely at all, how likely is it that you would continue using this app to track your mood? I am referring to the app you used originally, not the one with the forecast. Again please be honest.
- Okay, and why is that?
- Alright, and what is the most important takeaway we should remember from our discussion?
- That's it! Are there any other comments you would like to share?
- Thanks again for taking the time to speak with me today and for trying out the app. You are welcome to continue using the app as you like, though you will not receive additional compensation for any voluntary use.

Appendix 7. Study 2: Participant view of web form during interview.

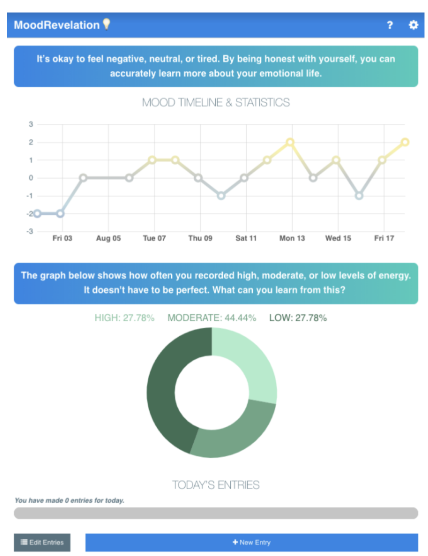
Enter participant ID: ____

[Author’s Note: If participant was in the improvement condition, they would see the following image]





[Author’s Note: If participant was in the self-knowledge condition, they would see the following image]



(Select up to 3) What are the best uses of a technology that tracks emotional states?
▢ To achieve a goal (1)
▢ To improve in this area, without a specific goal (11)
▢ To document my life accurately (2)
▢ To predict how I will do in the future (12)
▢ To understand myself better (13)
▢ To understand how this factor relates to others (for example, the impact of mood on exercise) (3)
▢ To explore a new technology (4)
▢ Other: (8) ____
▢ Other: (14) ____
▢ Don't know/unsure (6)

[Author's Note: Participants could select 'Yes' or 'No' for each of the concepts they thought should be reflected in an emotion-monitoring system. The options were presented in a randomized order.]

"The mood/emotion tracker should…"

… help you control your emotions
… help you identify triggers for different moods
… help you maximize pleasure and reduce discomfort
… have you set your own emotional goals
… help you have more accurate information about yourself
… hide negative information
… share all positive and negative information equally, regardless of impact on mood
... *not* have specific goals for making you more positive
… have specific goals for making you more positive
… help you accept your emotions
… help you have more frequent positive moods
… help you have more frequent negative moods
… help you have more balanced moods
… remind you to be positive
… compare how well you are doing against other people

… compare how well you are doing against your past self
… compare you against your 'best ever' positivity score
Other:
Other:
Other:

Appendix 8. Study 2 Interview: Responses to question about a hypothetical mood-tracking systems.

Study 2 interviewee responses to the question "*The mood/emotion tracker should…*"

Concept	Yes	No	Total
… help you control your emotions	6	8	14
… help you identify triggers for different moods	14	1	15
… help you maximize pleasure and reduce discomfort	5	9	14
… have you set your own emotional goals	9	4	13
… help you have more accurate information about yourself	13	2	15
… hide negative information	0	15	15
… share all positive and negative information equally, regardless of impact on mood	12	3	15
... *not* have specific goals for making you more positive	7	7	14
… have specific goals for making you more positive	7	8	15
… help you accept your emotions	13	1	14
… help you have more frequent positive moods	7	7	14
… help you have more frequent negative moods	0	15	15
… help you have more balanced moods	8	6	14
… remind you to be positive	6	6	14
… compare how well you are doing against other people	1	14	15
...compare how well you are doing against your past self	11	2	13
...compare you against your 'best ever' positivity score	2	11	13

References

Aitken, M., & Lyle, J. (2015). Patient adoption of mHealth: use, evidence and remaining barriers to mainstream acceptance. *Parsippany, NJ: IMS Institute for Healthcare Informatics*.

Arkin, R. M., & Maruyama, G. M. (1979). Attribution, affect, and college exam performance. Journal of Educational Psychology, 71(1), 85.

Ayobi, A., Marshall, P., Cox, A. L., & Chen, Y. (2017, May). Quantifying the body and caring for the mind: self-tracking in multiple sclerosis. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 6889-6901). ACM.

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. Social cognitive and affective neuroscience, 12(1), 1-23.

Bem, D. J. (1972). Self-perception theory. In *Advances in experimental social psychology* (Vol. 6, pp. 1-62). Academic Press.

Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.

Bradley, G. W. (1978). Self-serving biases in the attribution process: A reexamination of the fact or fiction question. Journal of personality and social psychology, 36(1), 56.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.

Burke, L. E., et al. (2011). Self-monitoring in weight loss: a systematic review of the literature. Journal of the American Dietetic Association,111(1), 92-102.

Carlson, E. N. (2013). Overcoming the barriers to self-knowledge: Mindfulness as a path to seeing yourself as you really are. *Perspectives on Psychological Science*, *8*(2), 173-186.

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological bulletin*, *92*(1), 111.

Catalino, L. I., Algoe, S. B., & Fredrickson, B. L. (2014). Prioritizing positivity: An effective approach to pursuing happiness?. *Emotion*, *14*(6), 1155.

CCS Insight. (2016). CCS Insight Wearables End-User Survey. 2016. Retrieved August 11, 2016 from http://www.ccsinsight.com/

Cheng, R. (2003). Persuasion strategies for computers as persuasive technologies. Department of Computer Science, University of Saskatchewan, 10(2010), 1-7.

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. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1143-1152). ACM.

Conner, T. S., & Reid, K. A. (2012). Effects of intensive mobile happiness reporting in daily life. *Social Psychological and Personality Science*, *3*(3), 315-323.

Consolvo, S., Klasnja, P., McDonald, D. W., & Landay, J. A. (2009, April). Goal-setting considerations for persuasive technologies that encourage physical activity. In *Proceedings of the 4th international Conference on Persuasive Technology*(p. 8). ACM.

Costa, J., Adams, A. T., Jung, M. F., Guimbretière, F., & Choudhury, T. (2016, September). EmotionCheck: leveraging bodily signals and false feedback to regulate our emotions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 758-769). ACM.

Davis, D., & Brock, T. C. (1975). Use of first person pronouns as a function of increased objective self-awareness and performance feedback. Journal of Experimental Social Psychology, 11(4), 381-388.

De Choudhury, M., Counts, S., & Horvitz, E. (2013, April). Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3267-3276). ACM.

Diener, E., Colvin, C. R., Pavot, W. G., & Allman, A. (1991). The psychic costs of intense positive affect. Journal of personality and social psychology, 61(3), 492.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, *97*(2), 143-156.

Dillon, A., Kelly, M., Robertson, I. H., & Robertson, D. A. (2016). Smartphone applications utilizing biofeedback can aid stress reduction. *Frontiers in psychology*, *7*, 832.

Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS one*, *6*(12), e26752.

Dunn, O. J. (1964). Multiple comparisons using rank sums. Technometrics, 6(3), 241-252.

Duval, S., & Wicklund, R. A. (1972). A theory of objective self awareness.

Duval, S., Wicklund, R. A., & Fine, R. L. (1972). Avoidance of objective self-awareness under conditions of high and low intra-self discrepancy. A theory of objective self-awareness, 15-22.

Duval, T. S., Duval, V. H., & Mulilis, J. P. (1992). Effects of self-focus, discrepancy between self and standard, and outcome expectancy favorability on the tendency to match self to standard or to withdraw. Journal of Personality and Social Psychology, 62(2), 340.

Elsden, C., & Kirk, D. S. (2014, June). A quantified past: remembering with personal informatics. In *Proceedings of the 2014 companion publication on Designing interactive systems*(pp. 45-48). ACM.

Elsden, C., Kirk, D., Selby, M., & Speed, C. (2015, April). Beyond personal informatics: Designing for experiences with data. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*(pp. 2341-2344). ACM.

Epstein, D. A., Lee, N. B., Kang, J. H., Agapie, E., Schroeder, J., Pina, L. R., ... & Munson, S. (2017, May). Examining menstrual tracking to inform the design of personal informatics tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6876-6888). ACM.

Epstein, D. A., Ping, A., Fogarty, J., & Munson, S. A. (2015, September). A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 731-742). ACM.

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*(13), 2691-2704.

Faurholt- Jepsen, M., Vinberg, M., Frost, M., Christensen, E. M., Bardram, J. E., & Kessing, L. V. (2015). Smartphone data as an electronic biomarker of illness activity in bipolar disorder. *Bipolar disorders*, *17*(7), 715-728.

Feldman, G., Hayes, A., Kumar, S., Greeson, J., & Laurenceau, J. P. (2007). Mindfulness and emotion regulation: The development and initial validation of the Cognitive and Affective Mindfulness Scale-Revised (CAMS-R). Journal of Psychopathology and Behavioral Assessment, 29(3), 177.

Fischer, S., Nater, U. M., & Laferton, J. A. (2016). Negative stress beliefs predict somatic symptoms in students under academic stress. *International journal of behavioral medicine*, *23*(6), 746-751.

Fisher, A. J., & Newman, M. G. (2013). Heart rate and autonomic response to stress after experimental induction of worry versus relaxation in healthy, high-worry, and generalized anxiety disorder individuals. *Biological psychology*, *93*(1), 65-74.

Ford, B., & Mauss, I. (2014). The paradoxical effects of pursuing positive emotion. *Positive emotion: Integrating the light sides and dark sides*, 363-382.

Fox, S., & Duggan, M. (2013). Health online 2013. *Health*, *2013*, 1-55.

Fredrickson, B. L., Tugade, M. M., Waugh, C. E., & Larkin, G. R. (2003). What good are positive emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. *Journal of personality and social psychology*, *84*(2), 365.

Friedman, B., Kahn, P., & Borning, A. (2002). Value sensitive design: Theory and methods. *University of Washington technical report*, 02-12.

Galvanic Ltd. 2016. Pip Stress Tracker. https://thepip.com/research-partnership/

Gibbons, F. X. (1990). Self-attention and behavior: A review and theoretical update. In *Advances in experimental social psychology* (Vol. 23, pp. 249-303). Academic Press.

Gimpel, H., Nißen, M., & Görlitz, R. (2013). Quantifying the quantified self: A study on the motivations of patients to track their own health.

Golbeck, J. (2016, November). Detecting Coping Style from Twitter. In *International Conference on Social Informatics* (pp. 454-467). Springer, Cham.

Goldsmith, K., Gal, D., Raghunathan, R., & Cheatham, L. (2013). The Pursuit of Happiness: Can It Make You Happy?. *ACR North American Advances*.

Greenberg, J., & Musham, C. (1981). Avoiding and seeking self-focused attention. *Journal of Research in Personality*, *15*(2), 191-200.

Gruber, J., Mauss, I. B., & Tamir, M. (2011). A dark side of happiness? How, when, and why happiness is not always good. *Perspectives on psychological science*, *6*(3), 222-233.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2002). Review of Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change.

Hollis, V., Konrad, A., & Whittaker, S. (2015). Change of heart: emotion tracking to promote behavior change. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 2643-2652). ACM.

Hollis, V., Konrad, A., Springer, A., Antoun, M., Antoun, C., Martin, R., & Whittaker, S. (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.

Hollis, V., Pekurovsky, A., Wu, E., & Whittaker, S. (2018). On Being Told How We Feel: How Algorithmic Sensor Feedback Influences Emotion Perception. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(3), 114.

Hollis, V., Springer, A., & Whittaker, S. (Under Review). Insight or Insult? Reasoning Strategies When Personal Informatics Contradict Self Beliefs. Submitted to Proceedings of the 2019 CHI Conference on Computer-Human Interaction (CHI). ACM.

Hua, J., Le Scanff, C., Larue, J., José, F., Martin, J. C., Devillers, L., & Filaire, E. (2014). Global stress response during a social stress test: impact of alexithymia and its subfactors. Psychoneuroendocrinology, 50, 53-61.

Hull, J. G., & Levy, A. S. (1979). The organizational functions of the self: An alternative to the Duval and Wicklund Model of self-awareness. *Journal of personality and social psychology*, *37*(5), 756.

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. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1071-1080). ACM.

Jamieson, J. P., Mendes, W. B., & Nock, M. K. (2013). Improving acute stress responses: The power of reappraisal. Current Directions in Psychological Science, 22(1), 51-56.

Jamieson, J. P., Mendes, W. B., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. Journal of Experimental Social Psychology, 46(1), 208-212.

Kang, J., Binda, J., Agarwal, P., Saconi, B., & Choe, E. K. (2017, May). Fostering user engagement: improving sense of identity through cosmetic customization in wearable trackers. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 11-20). ACM.

Kersten-van Dijk, E. T. (2018). Quantified stress: toward data-driven stress awareness.

Kersten-van Dijk, E. T., Westerink, J. H., Beute, F., & IJsselsteijn, W. A. (2017). Personal informatics, self-insight, and behavior change: A critical review of current literature. *Human–Computer Interaction*, *32*(5-6), 268-296.

Kluger, A. N., & Adler, S. (1993). Person-versus computer-mediated feedback. *Computers in human behavior*, *9*(1), 1-16.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, *119*(2), 254.

Konrad, A., Bellotti, V., Crenshaw, N., Tucker, S., Nelson, L., Du, H., ... & Whittaker, S. (2015, April). Finding the adaptive sweet spot: Balancing compliance and achievement in automated stress reduction. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 3829-3838). ACM.

Laferton, J. A. C., Stenzel, N. M., & Fischer, S. (2016, October 10). The Beliefs About Stress Scale (BASS): Development, Reliability, and Validity. International Journal of Stress Management. Advance online publication. http://dx.doi.org/10.1037/str0000047

Li, I., Dey, A., & Forlizzi, J. (2010, April). A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*(pp. 557-566). ACM.

Li, I., Dey, A. K., & Forlizzi, J. (2011, September). Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing* (pp. 405-414). ACM.

Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G., & Strub, H. B. (2006, September). Fish'n'Steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing* (pp. 261-278). Springer, Berlin, Heidelberg.

Lupton, D. (2014). Self-tracking cultures: towards a sociology of personal informatics. In Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design (pp. 77-86). ACM.

Lupton, D. (2014). Self-tracking modes: Reflexive self-monitoring and data practices. Available at SSRN 2483549.

Lupton, D. (2015). Quantified sex: a critical analysis of sexual and reproductive self-tracking using apps. Culture, health & sexuality, 17(4), 440-453.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. Social indicators research, 46(2), 137-155.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social indicators research*, *46*(2), 137-155.

Lyubomirsky, S. (2008). *The how of happiness: A scientific approach to getting the life you want*. Penguin.

MacLean, D., Roseway, A., & Czerwinski, M. (2013, May). MoodWings: a wearable biofeedback device for real-time stress intervention. In *Proceedings of the 6th international conference on pervasive technologies related to assistive environments* (p. 66). ACM.

Matthews, M., Murnane, E., & Snyder, J. (2017). Quantifying the Changeable Self: The role of self-tracking in coming to terms with and managing bipolar disorder. Human–Computer Interaction, 32(5-6), 413-446.

Matthews, M., Snyder, J., Reynolds, L., Chien, J. T., Shih, A., Lee, J. W., & Gay, G. (2015, April). Real-time representation versus response elicitation in biosensor data. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 605-608). ACM.

Mauss, I. B., Tamir, M., Anderson, C. L., & Savino, N. S. (2011). Can seeking happiness make people unhappy? Paradoxical effects of valuing happiness. *Emotion*, *11*(4), 807.

McDonald, P. J. (1980). Reactions to objective self-awareness. Journal of Research in Personality, 14(2), 250-260.

McFall, R. M. (1970). Effects of self-monitoring on normal smoking behavior. Journal of Consulting and Clinical Psychology, 35(2), 135.

McFarland, C., Ross, M., & DeCourville, N. (1989). Women's theories of menstruation and biases in recall of menstrual symptoms. Journal of Personality and Social Psychology, 57(3), 522.

Nielsen, C. (2014). Tech-styles: Are consumers really interested in wearing tech on their sleeves. *Retrieved from Nielsen Newswire website: http://www. nielsen. com/us/en/newswire/2014/tech-styles-are-consumers-really-interested-in-wearing-tech-on-their-sleeves. Html*.

Oishi, S. (2002). The experiencing and remembering of well-being: A cross-cultural analysis. *Personality and Social Psychology Bulletin*, *28*(10), 1398-1406.

Onnela, J. P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology, 41(7), 1691.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867-872.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, *71*(2001), 2001.

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. American journal of health promotion, 12(1), 38-48.

Rapp, A., & Tirassa, M. (2017). Know thyself: a theory of the self for personal informatics. *Human–Computer Interaction*, *32*(5-6), 335-380.

Raven, J. C. (1938). *Raven's progressive matrices*. Western Psychological Services.

Rimes, K. A., & Chalder, T. (2010). The Beliefs about Emotions Scale: validity, reliability and sensitivity to change. *Journal of psychosomatic research*, *68*(3), 285-292.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: evidence for an accessibility model of emotional self-report. Psychological bulletin, 128(6), 934.

Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. C. (2014, April). Personal tracking as lived informatics. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1163-1172). ACM.

Ruckenstein, M. (2014). Visualized and interacted life: Personal analytics and engagements with data doubles. *Societies*, *4*(1), 68-84.

Sarason, I. G., Johnson, J. H., & Siegel, J. M. (1978). Assessing the impact of life changes: development of the Life Experiences Survey. Journal of consulting and clinical psychology, 46(5), 932.

Scheier, M. F., & Carver, C. S. (1977). Self-focused attention and the experience of emotion: attraction, repulsion, elation, and depression. Journal of personality and social psychology, 35(9), 625.

Schooler, J. W., Ariely, D., & Loewenstein, G. (2003). The pursuit and assessment of happiness can be self-defeating. *The psychology of economic decisions*, *1*, 41-70.

Scollon, C. N., Howard, A. H., Caldwell, A. E., & Ito, S. (2009). The role of ideal affect in the experience and memory of emotions. *Journal of Happiness Studies*, *10*(3), 257-269.

Sealed Envelope Ltd. 2012. Power calculator for continuous outcome superiority trial. [Online] Available: sealedenvelope.com/power/continuous-superiority

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. In *Advances in experimental social psychology* (Vol. 29, pp. 209-269). Academic Press.

Silvia, P. J., & Duval, T. S. (2001). Objective self-awareness theory: Recent progress and enduring problems. *Personality and Social Psychology Review*,*5*(3), 230-241.

Simpson, C. C., & Mazzeo, S. E. (2017). Calorie counting and fitness tracking technology: associations with eating disorder symptomatology. *Eating behaviors*, *26*, 89-92.

Sjöklint, M., Constantiou, I. D., & Trier, M. (2015). The complexities of self-tracking-An inquiry into user reactions and goal attainment. Available at SSRN.

Sjöklint, M. (2014, September). The measurable me: the influence of self-quantification on the online user's decision-making process. In Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program (pp. 131-137). ACM.

Snyder, J., Matthews, M., Chien, J., Chang, P. F., Sun, E., Abdullah, S., & Gay, G. (2015, February). Moodlight: Exploring personal and social implications of ambient display of biosensor data. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 143-153). ACM.

Spiel, K., Kayali, F., Horvath, L., Penkler, M., Harrer, S., Sicart, M., & Hammer, J. (2018, April). Fitter, Happier, More Productive?: The Normative Ontology of Fitness Trackers. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (p. alt08). ACM.

Spielberger, C. D., & Gorsuch, R. L. (1983). *State-trait anxiety inventory for adults: Manual, instrument, and scoring guide*. Mind Garden, Incorporated.

Springer, A., Hollis, V., & Whittaker, S. (2017). Dice in the Black Box: User Experiences with an Inscrutable Algorithm. In AAAI Spring Symposium Series.

Swann, W. B., & Read, S. J. (1981). Acquiring self-knowledge: The search for feedback that fits. *Journal of Personality and Social Psychology*, *41*(6), 1119.

Swann, W. B., Wenzlaff, R. M., Krull, D. S., & Pelham, B. W. (1992). Allure of negative feedback: Self-verification strivings among depressed persons. Journal of Abnormal Psychology, 101(2), 293.

Swann Jr, W. B., Pelham, B. W., & Krull, D. S. (1989). Agreeable fancy or disagreeable truth? Reconciling self-enhancement and self-verification. Journal of personality and social psychology, 57(5), 782.

Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of personality and social psychology*, *90*(2), 288.

Van Boven, L., & Robinson, M. D. (2012). Boys don't cry: Cognitive load and priming increase stereotypic sex differences in emotion memory. Journal of Experimental Social Psychology, 48(1), 303-309.

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of personality and social psychology*, *98*(2), 281.

Vhaduri, S., Ali, A., Sharmin, M., Hovsepian, K., & Kumar, S. (2014, September). Estimating Drivers' Stress from GPS Traces. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1-8). ACM.

Warshaw, J., Matthews, T., Whittaker, S., Kau, C., Bengualid, M., & Smith, B. A. (2015, April). Can an Algorithm Know the Real You?: Understanding People's Reactions to Hyper-personal Analytics Systems. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 797-806). ACM.

Whittaker, S., Kalnikaite, V., Hollis, V., & Guydish, A. (2016, May). 'Don't Waste My Time': Use of Time Information Improves Focus. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1729-1738). ACM.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, *39*(2-3), 165-210.

Wilson, T. D., Laser, P. S., & Stone, J. I. (1982). Judging the predictors of one's own mood: Accuracy and the use of shared theories. Journal of Experimental Social Psychology, 18(6), 537-556.