EudaeSense

\ˌyüdēˈsens\

sensing happiness and well-being

Richard Chen, April Dawn Kester & Audrey Leung
Advisor, Professor John Chuang

Capstone Project 2016 - Final Report
Master of Information Management & Systems

Berkeley School of Information

Abstract

The vast majority of mood tracking applications today have a significant drawback, they require too much user input and effort. Our solution is a system that passively collects user data, which is used to infer early signs of potential depressed mood, and suggests behavioral interventions.

We use wearable biometric sensors to measure and track physiological signals that reflect emotion such as heart rate, sweat, temperature, muscle tension, and breathing rate. We also use smartphone and wearable sensors to monitor behavior associated with mood changes such as sleep quality, activity level, mobility and social interaction.

Our smartphone/smartwatch application collects this data and transmits it to a back-end system that uses generalized machine learning models to predict possible mood changes.

Lastly, our application suggests behavioral interventions. Interventions were derived from existing research on cognitive behavioral therapy, in tandem with expert interviews and user studies which uncovered new and novel approaches to digital mental health management. These contextually relevant, just-in-time interventions are pushed to users discreetly on their smartwatch.

Acknowledgments

This project would not have been possible without the help and support of several other individuals and their time.

We would like to thank our advisor, Professor John Chuang for his steadfast mentorship and keeping us grounded and sane throughout the process. We would also like to extend our thanks to David Bamman, Steve Fadden, and Steve Weber for their time and input.

Pablo Paredes inspired us with his PopTherapy research, which informed some of our design decisions, and we greatly appreciated his keen insights.

Would would also like to thank Orianna DeMasi, who was instrumental in guiding us through the pitfalls of data collection, data analysis and mental health research.

We had an amazing team of students from our Data Mining and Deconstructing Data Science courses who were instrumental in helping us develop our mood predictive algorithm -- Anubhav Gupta, Byung Song, Heidi Huang, Laura Montini, and Yiwen Tang.

We received significant help for our formative user research from a team of students in Needs and Usability Assessment -- Hannah Gogel, Jamie She, and Pavel Vanegas.

To our family and friends, and significant others, thank you for your unwavering encouragement and patience, which was invaluable in keeping us going through the long days and late nights.

Table of Content

Table of Figures

1 Introduction

According to the World Health Organization, “1 in 4 people in the world will be affected by mental or neurological disorders at some point in their lives” [1]. On the rise is one such mental health disorder, depression, “currently ranked fourth among the 10 leading causes of the global burden of disease, it is predicted that by the year 2020, it will have jumped to second place” [2]. Although most depression is treatable, fewer than 50% of those affected in the world (and fewer than 10% in many countries) receive such treatments. Barriers to effective treatment include a lack of resources, lack of trained health care providers, social stigma associated with mental disorders, and inaccurate assessment [3].

Our project leverages current research on computational algorithms used to detect and assess mental well-being, by building an application on a wearable device to passively collect biometric user data, evaluate risk for depressive mood, and introduce subtle micro-interventions through a wearable interface. Our goal is to help those suffering from depression self-adjust their mental states.

Through our research, we identified five indicators of depression detectable through wearable sensors: social interaction, mobility, sleep patterns [4], phone usage [5], and nocturnal temperature [6]. Current applications that aim to improve mood or mental health suffer from low adherence and engagement rates due to the obstructive and time-consuming nature of experience sampling methods [7].

EudaeSense passively gathers sensor data using open source APIs or platforms and apply researched-backed algorithms to assess depression risk based on these indicators. Wearable devices can interpret these indicators to prompt users at the time they are slipping into depression, so they can actively self-adjust to overcome a depressed state of mind.

The vast majority of mood tracking applications today have a significant drawback: they require significant user input and effort. The idea of the application assuming responsibility and prompting the user for attention -- as opposed to the other way around -- is our approach that we believe will be more effective and have greater retention and usage rates.

There are three distinct parts to our system: First, we used Fitbit to passively collect user data. This was a substitute for the currently insufficient smartwatch algorithms. As the Fitbit passively collects user data, we will be running algorithms analyzing the Fitbit data to predict the user’s mood. By relying heavily on sensor data for analysis as opposed to user reported data, we developed relatively unobtrusive system that seamlessly integrates into users' daily routines and can be run in the background without their conscious exertion or effort. The second part of our solution was an Android application to collect and analyze the sensing data and assess the mood of the user. The third part of our solution is behavioral intervention beginning on the smartwatch,, which will be crucial to the success of our product. The specific interventions will be derived from existing research on cognitive behavioral therapy. This work will ultimately contribute to further research in interventions for mental wellness, keeping in mind recent research surrounding neuroplasticity.

2 Background

2.1 About Depression

Depression is an internal struggle that often times cannot be see through action, behaviors, or physical manifestations. It is inextricably tied to cultural, relational, and environmental factors. It extends beyond our current state of mind, and often times inappropriately juxtaposes past events onto our present, or even paints a toxic forecast on our future. It is not a simple linear transition from a healthy mind to an unhealthy one, but rather, a very complex network of issues that can attack us at any time.

We realize that the very idea of using wearables and smartphone sensors to monitor and detect signals of depression is in itself a very behaviorist approach. Traditional behaviorists focused solely on external and observable behavior rather than thoughts and perceptions, because they believed that what goes on inside someone's mind is difficult to measure and irrelevant in the process of influencing behavior [8]. We came to realize that using only biosignals from wearables and the behavioral patterns from smartphones to determine the existence of depression is essentially minimizing it to only an observable behavior. The comparison to behaviorism naturally guided us towards learning about the cognitive behavioral theories of depression. Cognitive behavioral theories focus on what sort of ways of thinking people with depression have that may influence or bring out depressed mood. There are four notable cognitive theories of depression: Aaron Beck's cognitive theory of depression, Albert Ellis' cognitive theory of depression, Albert Bandura's social cognitive theory of depression, and Martin Seligman's learned helplessness [8]. One of the common themes across all of these theories is that feelings and thoughts concerning powerlessness and usefulness shape people's self-concept and mood. And a lot of this is greatly influenced by culture. Looking into the sociology of depression has turned out to be very illuminating. A branch of medicine, known as ethnomedicine, is defined as "a study or comparison of the traditional medicine practiced by various ethnic groups, and especially by indigenous peoples" [8]. In the context of depression, it focuses on the role of culture, perception, and context in shaping someone's physical and mental health. The formation of mental illnesses can be strongly influenced by different cultural perceptions, how one focuses on their place within a social hierarchy (such as an individualistic vs. a collectivistic orientation), the varying degrees of stigmatization of mental illness in different cultures, and rigid gender roles (to name a few) [8]. Other social and relational factors, such as the death of a loved one, loss of a job, abusive relationships, etc. also play a crucial role in the development and onset of mental health problems.

According to the DSM5, there are four major categories of depression: Major Depressive

Disorder (aka clinical depression), Persistent Depressive Disorder (aka Dysthymia), Disruptive Mood Dysregulation Disorder (DMDD), and Premenstrual Dysphoric Disorder [9]. We believe that our app would best fit the individuals who fall within the category of Persistent Depressive Disorder. Due to its high risk high reward nature, we believe that this would be the most appropriate entry point, but more importantly, our hope is that it can serve as a proper intervention to prevent the depression from getting worse. Although we will not be able to address or tackle the root of one's depression with our app, we strongly believe that it has the potential to redirect the user towards a more consistent positive mindset, and at the very least, put a temporary halt on negative attributional styles of thinking.

2.2 Interventions for Depression

The single biggest issue underlying mental health applications is the high attrition rates and the struggle to maintain engagement and adherence [10]. Through our readings, we discovered that there are 4 design strategies for helping online applications maintain engagement: interactive strategy, personal strategy, supportive strategy, and social strategy [10]. This has been proven to be necessary elements in intervention design, so regardless of what we build, we will need to include these 4 elements. We also know that are two general groups of interventions for mental health: in person one-on-one interventions, and online based interventions [11]. Within the online based group, there is a web-based group and a mobile-app based group. The application we are building has its own category -- it is a smartwatch to smartphone cross-platform application.

Despite all of the information we gathered from our research and expert interviews, it was difficult to conceptualize how we were going to design interventions for a cross-platform application. In our research following the expert interviews, we discovered two main sources of inspiration. The first came from an application we found online called EmotionSense. EmotionSense is an app with the closest resemblance to our project concept. It is a mood tracking app that passively tracks user data, such as calling and texting patterns, by using the inbuilt GPS, accelerometer, and microphone in the smartphone. Throughout the day, users are also prompted to rate emotional states via a brief survey [12]. The second source of inspiration came from a research paper written by Pablo Paredes, a PhD student from the EECS department here in UC Berkeley. In his paper, titled PopTherapy: Coping with Stress through Pop-Culture, Pablo discusses a study on a mobile application he built that is designed to aid with stress management [13]. His paper resonated with us, as his overarching research question was almost identical to ours: how can one design the "right" interventions (the WHAT) to be delivered at the "right" times (the WHEN)? Pablo's mobile app was driven by micro-interventions. These micro-interventions are formatted to a minimal expression and all have the following components: a text prompt that tells the user what to do and a URL that launches the appropriate tool to execute the micro-intervention. Each of these micro-interventions also fall under 1 of 4 psychotherapy groups: positive psychology, cognitive behavioral therapy, meta-cognitive therapy, and somatic therapy. They were further subdivided into ones that can be performed alone (individual) or with/for others (social) [13]. When a user is feeling particularly stressed, they would launch the app,

The EmotionSense app and Pablo's mobile app helped to springboard our own intervention design process. EmotionSense includes libraries to trigger notifications based on sensor events, and their code can be adapted for use in a wearable APK [12]. We were inspired by how Pablo organized his micro-interventions into different psychotherapy groups and we sought to create a similar organizational structure for our sleep interventions. We also realize that the smartphone sensors are limited, and our project aims to take EmotionSense and Pablo's mobile app one step further by incorporating wearable sensors. We believe that seamless data collection combined with an interface that is less obtrusive to the user will increase application usage and improve overall mental health.

2.3 Tracking Depression

2.3.1 Autonomic Nervous System - Measuring Stress and Depression

The autonomic nervous system (ANS) regulates functions such as heart rate, respiration, and perspiration. There are two branches in the ANS, the sympathetic nervous system most often associated with "fight or flight", and the parasympathetic nervous system considered the "feed and breed" system. There is growing evidence demonstrating that stress and depression can influence autonomic nervous system responses [14]. It then posits then that emotion can be inferred through ANS activity. "Numerous psychophysiological measurements are shown to be differentially sensitive when testing correlations with stress and depression [14]."

Such psychophysiological responses include: heart rate and blood volume which INCREASE when nervous. Skin conductance INCREASES when encountering stress, reflecting sympathetic nervous system activity. Temperature DECREASES when stressed, blood constricts. Respiration INCREASES when sympathetic nervous system is activated, but DECREASES when parasympathetic nervous system is activated [14].

Skin conductance represents autonomic arousal during the day, responses during sleep are highly likely to occur in either non-REM Stage 2 sleep or Slow Wave Sleep, this helps to characterize sleep better than using only acceleration data. Skin temperature also helps to understand sleep/wake patterns [15].

High nocturnal temperature significantly higher in depressed patients, compared to controls, and decreased significantly with recovery [6].

2.3.2 Physiological Signals and Wearable Biosensors

ANS activity can be measured through various biometric sensors, many of which are becoming available on the consumer market. Along with biometric data these devices can also measure activity and behavior such as exercise and sleep.

FIGURE 2.3.1 TABLE OF DEVICE SENSORS - CONSUMER MARKET

Sensor	Signal/Measurement
Optical Sensor - Photoplethysmography (PPG)	Heart Rate and Respiration
Electrodermal activity (EDA), Galvanic skin response (GSR) sensor	Perspiration
Infrared Thermopile	Temperature
Electrocardiography (ECG)	Heart
Electroencephalography (EEG)	Brain, Event-Related Potential (ERP)
Electromyography (EMG)	Muscle Activation
Bio-impedance Plethysmogram (IPG)	Heart Rate, BMI, GSR, Respiration
Nasal/Oral Thermocouple, Respiratory Effort Belt	Respiration, Spirometry
Pulse Oximeter	Blood Oxygen Saturation (%SpO2) and Pulse Rate
Electrooculography (EOG)	Eye movement
Sphygmomanometer	Blood Pressure
Gyroscope	Orientation
3-Axis Accelerometer	Proper Acceleration
Barometer	Atmospheric Pressure
Altimeter	Altitude

Sensors are routinely used in clinical settings to monitor patient health, these body sensor networks (BSNs) incorporate sensors such as photoplethysmography (PPG), and electrocardiogram (ECG) and relays these physiological measurements wirelessly and in real time for processing [16]. As wearable technologies become more prevalent, and the cost of biometric sensors decrease, it is probable BSN's will be used by individuals at home or in the workplace.

Patients have also used Plethysmograph to train themselves to follow targeted respiration rhythm. "Various reports suggested that respiratory trainings have assisted patients who suffer stutters, panic attacks, asthmatic conditions, depression, and post-traumatic stress disorder" [16].

2.3.3 Affective Phenomenon - Mood, Affective State, Emotion

Affective computing is a field where machine learning methods are used around something very subjective -- our emotions. Affective computing has the goal of designing interactive computing systems that that take into account the users' emotions. For example, Affectiva is a company that spun out of research at the MIT Media lab that takes in video of a person's face and detects different parts of their facial expression, such as the shape of their mouth or their eyes, and uses machine learning to map combinations of those to different aspects related to emotion such as valence (positive or negative feelings), surprise, smiling, etc.

Within affective computing, there are two different approaches to emotion -- an "information"-centric approach which emphasizes symbolic representation and an "interaction"-centric approach which focuses more on interaction. In the Information Model, emotions are assumed to be individually experienced, measured at the level of the individual, and are symbolically encoded. The goal is to have computers detect emotions and put emotions into discrete categories such as "happy", "stressed", or positive or negative valence. If the user's emotion is happy, then that information can be transmitted from the user to the computer and then transmitted somewhere else to another user, and the meaning of happiness is taken as being largely independent of context. In the Interaction Model, the focus is less on detecting and categorizing emotions and more on helping other humans interpret emotion in the context of interaction. So happiness is not assumed to be a discrete emotion that can be interpreted outside of context. Rather, the emphasis is on emotions as socially situated -- how we think about our own emotions, even when we're alone, is inextricable from our culture and our interactions with others. In terms of designing interactive systems with the information model one might simply display or report the emotion detected, whereas the interaction model is still being explored as to the different ways to design a system to allow a user to interact or respond to their own emotions.

One example of a hybrid information and interaction model is "Freaky", an alien larvae-like creature worn in a baby carrier [17] that uses machine learning to detect fear in its user. When it detects fear, it "freaks out" by making noises and vibrating. Freaky serves as a machine interpretation rather than claiming that this is the "true value" for the human's current emotions. The system accommodates both machine interpretation and human interpretation, rather than claiming they are the same. Rather than assigning the information-centric label of "fear" to the wearer, Freaky enacts that fear itself. Much of the meaning and interpretation comes out in the wearer's interaction with Freaky, and with other people who are around Freaky.

Similarly, EudaeSense also aims to employ the interaction model of affective computing in the implementation of its interventions. Users will be prompted to interact with the smartwatch device that will recommend activities based on the algorithmic interpretations of the user's activity and movement. Rather than simply displaying the mood that is detected by the algorithm, EudaeSense would go one step further to notify the user to interact with the system through an activity, intervention,, or notification.

2.4 Related Work

Prior research has been conducted on inferring mood or depression by leveraging sensor data from smartphones. We have identified five research-backed indicators of depression detectable through wearable sensors:

1. Social interaction: To determine the level of social interaction, researchers used number of calls, ambient noise, text [5], speech duration [4], and conversation frequency and duration [18].

2. Mobility: Spending most of the time at home of in fewer locations as measured by GPS were linked to depression and having a less-regular day-to-day schedule has also been linked to depression [19]. GPS location and wi-fi features that record location have also been shown to predict mood [5]. Increased geospatial activity was associated with a better change in PHQ-9 scores [4].

3. Smartphone phone usage patterns: Greater app usage and frequently on-state status of phone screen was linked with depression [5]. Studies also showed that the more time participants spent using their phone, the more likely they were depressed [19]. Bipolar or manic depression can be detected using a combination of the accelerometer and GPS as those patients tended to make more fast-talking, frequent phone contact with others [20].

4. Sleep patterns: An inability to sleep has been linked to depression with lower sleep duration predictive of a negative change in PHQ-9 scores [4, 18].

5. Nocturnal temperature: The nocturnal temperature was significantly higher in depressed patients compared to controls, and decreased significantly with recovery. The nocturnal sweat rates of depressed patients did not differ significantly from

those of controls, but decreased significantly with recovery. The nocturnal sweating rates in the depressed patients suggest that impaired sweating is not the cause of the high nocturnal temperature commonly found in depressed patients [5].

Current applications that aim to improve mood or mental health suffer from low adherence and engagement rates due to the obstructive and time-consuming nature of experience sampling methods [7].

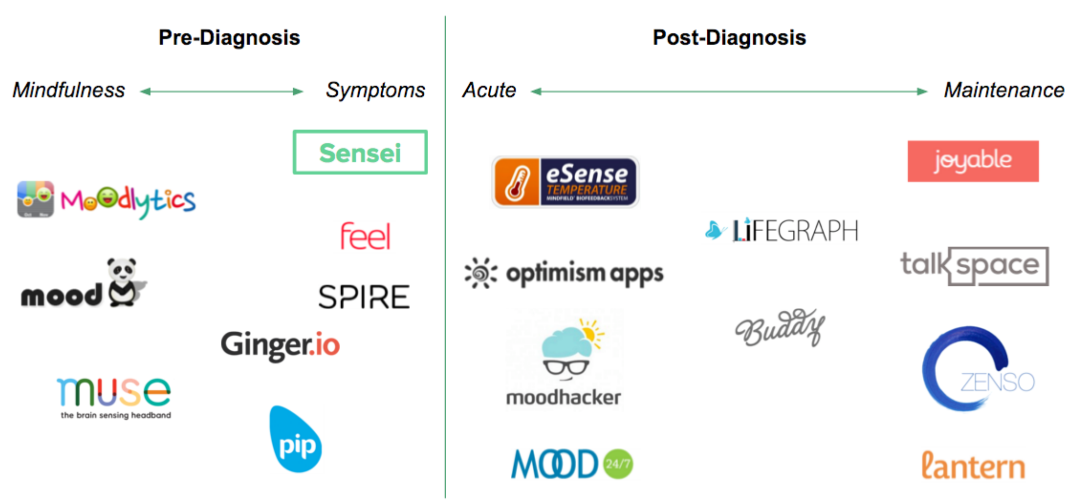
2.5 Competitive Analysis

2.5.1 Industry Landscape and Differentiation

There exists a very large industry with respect to the treatment of depression and the symptoms that lead up to it, so we had to carefully consider where our solution would best fit itself in this landscape. The tools available to address the symptoms of depression run the gamut from consumer mood trackers and wearable sensors all the way to medical tools such as telemedicine (helplines) and behavioral tracking and intervention applications.

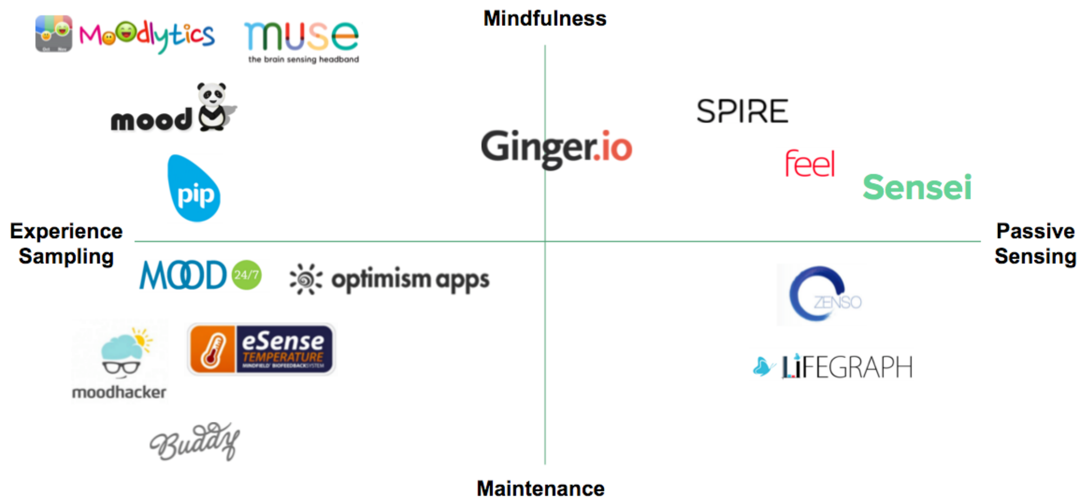
One of the first dimensions we considered was to assess the stage of depression where it would be best to introduce our kind of solution. As seen in the graphic below, you can classify existing applications based on the degree of severity of depression, from when a user is becoming mindful of their mood to when they start displaying the symptoms of depression up until they are diagnosed and require regular therapy and treatment to manage the depression. It can be seen that there are a representative number of applications for each of these levels, and we feel that EudaeSense would be well-positioned right before a user is diagnosed but after they start showing symptoms of depression that would be detected by EudaeSense's sensors and machine-learning algorithms.

FIGURE 2.5.1 COMPETITIVE LANDSCAPE - STAGE OF DEPRESSION



The next dimension we considered when it came to positioning EudaeSense was that of data collection. Many of the existing applications on the market rely on active collection of data from the user, prompting them to answer surveys and make subjective assessments of their mood at a given time of day, or for them to take a measurement of a certain biometrics such as heart rate variability or body temperature. One of the key issues we found with the active sampling approach was that users do not have a good baseline to assess their mood whenever subjective assessments are required, thus jeopardizing the accuracy of behavioral tracking. The same can be said for active collection of biometrics.

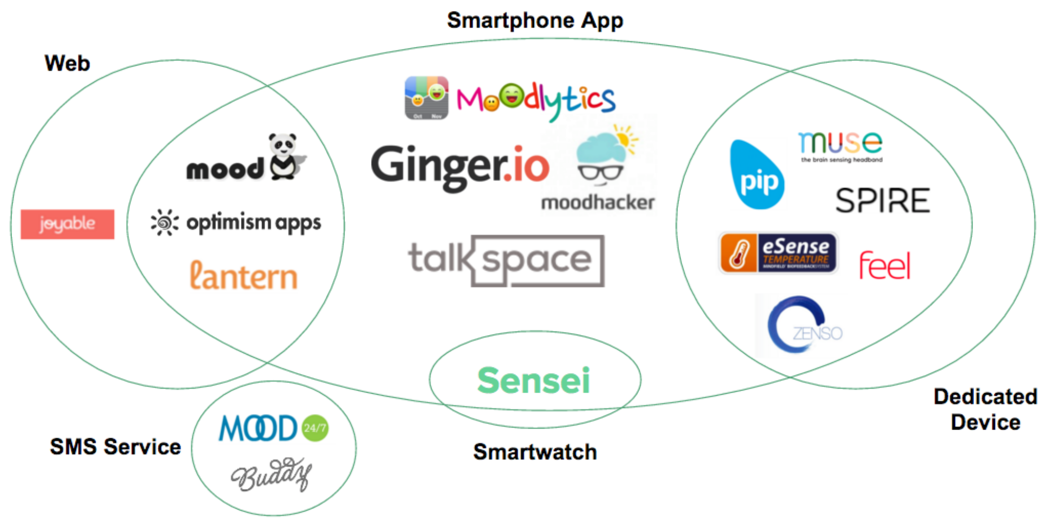
FIGURE 2.5.2 COMPETITIVE LANDSCAPE - DATA COLLECTION



Thus, the strength of our solution lies in the passive sensing of multiple indicators of depression coming from the different sensors available on the user's smartphone and smartwatch. This leads us to the next dimension we considered, which was the technology platform the solution would reside on.

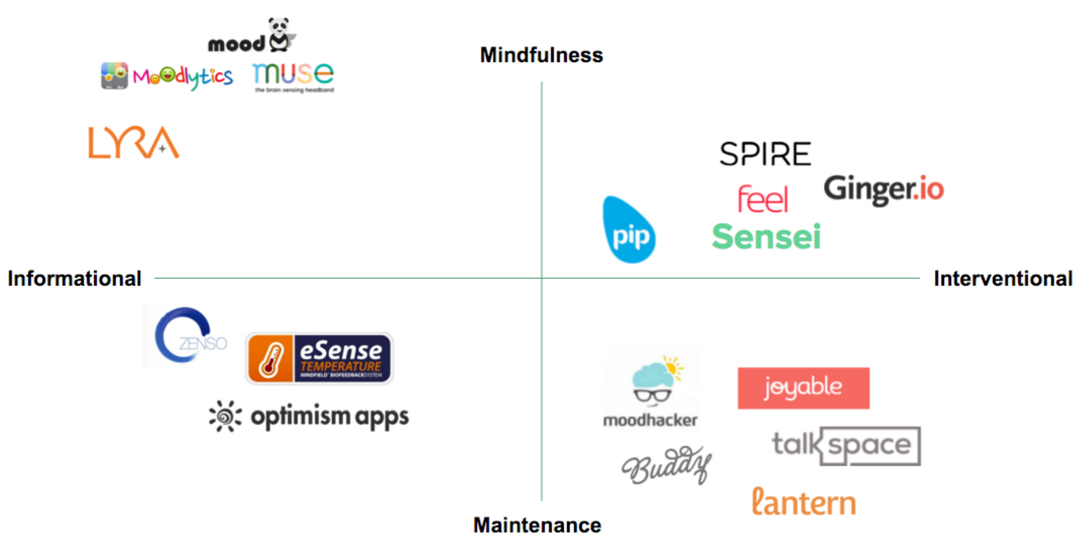
Almost all the competitors have some form of application installed on the smartphone, with some of them utilizing dedicated devices to enhance or collect more accurate sensor data. Our solution is the only one on the market that will be integrated into and will passively collect data from a user's smartwatch, thereby eliminating the need for a dedicated device.

FIGURE 2.5.3 COMPETITIVE LANDSCAPE - TECHNOLOGY PLATFORM



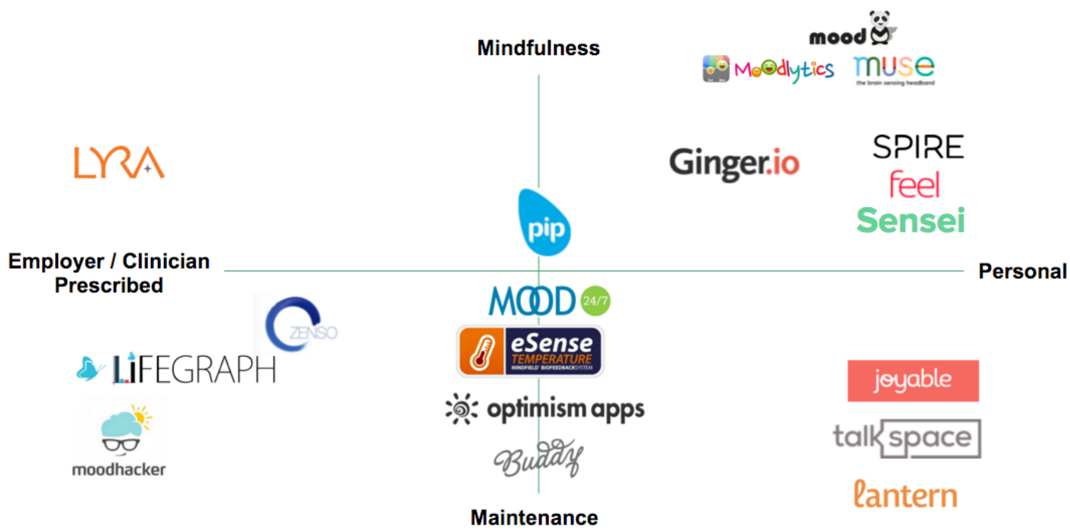
The next key dimension we looked at was behavior modification and at what level it would be acceptable to introduce interventions to alleviate the symptoms of depression the user was experiencing. In our survey we found that a large number of consumer-level applications already introduced some form of intervention and behavior modification, thus to maintain feature parity our solution would need to employ a similar system.

FIGURE 2.5.4 COMPETITIVE LANDSCAPE - LEVEL OF INTERVENTION



Finally, we looked at how the current solutions are distributed in the market. A large number of them can be easily downloaded from the application stores on the iOS and Android platforms, while a few are being used primarily for clinical trials or are distributed in partnership with employers and insurance companies.

FIGURE 2.5.5 COMPETITIVE LANDSCAPE - DISTRIBUTION CHANNEL



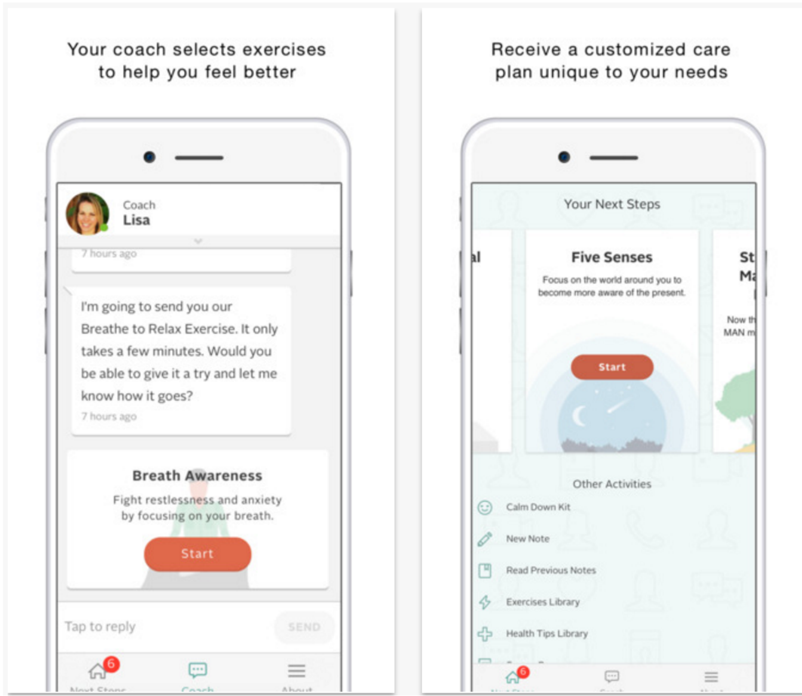
2.5.2 Key Competitors

After the analysis of the existing landscape of depression application across different dimensions, we narrowed down our focus to three key competitors in the market that employ the three key parts of our solution - passive sampling of sensor data, personalized analytics of user data and behavioral modification through micro-interventions. These three competitors are Ginger.io, Spire and Feel.

The first competitor we looked at is Ginger.io, which is a smartphone application that markets itself as providing personalized mental health care for anxiety and depression. Built on top of social sensing research done in the MIT Media Lab, the app analyzes data from the smartphone's location and communication sensors to understand user movement, smartphone usage and communication patterns and whether these indicate of depression. Ginger's business model hinges on using the data collected by the service and selling this to health care providers, while keeping it free for patients and users. The benefit for the healthcare provider is that they can tailor their care plans to the patient's diagnosis or risk profile.

From a user standpoint, the key advantage of our solution over Ginger's is that EudaeSense will be employing a larger number of sensors and will be tracking a greater number of depression indicators, thereby increasing the number of data points and accuracy of the machine learning algorithms EudaeSense will employ. Ginger also requests the user to occasionally provide subjective assessments of their mood (at least once a day) to supplement the data it collects from the mobile phone's sensors. Our solution will reduce the need for experience sampling due to its reliance on a larger number of sensors as indicators of depression.

FIGURE 2.5.6 KEY COMPETITOR - GINGER.IO



The second key competitor is Spire, which bills itself as a personal mindfulness coach that employs a smartphone application and a dedicated clip-on device that monitors a user's breathing patterns. The device, called a Spire Stone, retails for $149.95 and passively collects data on the frequency and magnitude of each of the user's exhaled and inhaled breaths. Based on this breath analysis, the application sends notifications to the user in the form of breathing exercises whenever it detects tension or stress in the user's breathing pattern.

FIGURE 2.5.7 KEY COMPETITOR - SPIRE

Spire's business model relies on revenue generated from sales of the Stones, which is the final monetary transaction they have with each customer. Spire does not sell the data to health care providers or insurance companies, unlike Ginger's approach.

EudaeSense plans to compete head on with Spire by not requiring a separate dedicated device to operate, while also looking at other indicators beyond the user's breathing patterns. According to a 2014 white paper from Endeavour Partners, it was found that roughly 57% of wearable users lose interest in wearing their devices within 12 months. EudaeSense avoids this by integrating into the existing smartphone and smartwatch setup of the user.

The final competitor we are looking at is Feel, which is a wristband that aims to track and measure a user's emotions throughout the day. The wristband comes with four integrated sensors that track biosignals such as galvanic skin response, blood volume pulse and skin temperature, which are sent to a smartphone application over Bluetooth which visualizes the data and provides the user with personalized recommendations similar to that of the first two competitors.

FIGURE 2.5.8 KEY COMPETITOR - FEEL

We believe that while this wristband does take into consideration a larger number of biosignals, it suffers from the same weakness as Spire in that it requires a dedicated device to operate. As of December 2015, the startup behind Feel, Sentio Solutions, is still raising the capital needed to mass produce their wristband. They have not launched their product in the market nor have they set any price point. All activity so far has been to promote their product at various industry conferences and events, while courting investors to back the company.

Our research found a similar product to Feel called Olive, which was also a wristband that attempted to assist users with stress management. The startup behind Olive took a different approach and launched with a crowdfunding campaign on Indiegogo in 2014, which raised $180,000 and allowed them to have some initial traction with users. However, as of May 2015, the company has shut down their operations due to the lack of a strong business model required to sustain the production of a dedicated device, and refunds have been issued to their backers on Indiegogo. The challenges faced by these startups to come up with robust revenue streams to fund the manufacturing of their devices strengthens EudaeSense's positioning in the market, as our solution does not require a separate device to operate and deliver value for the end user.

3 Proposed Solution

3.1 Summary

The value proposition of EudaeSense is simple: we offer an unobtrusive, accurate way for users to ward off symptoms of depression through personalized, timely micro-interventions that reduce stress and anxiety in their daily lives.

EudaeSense is unobtrusive - there is no additional device that the user needs to purchase, our solution leverages the user's existing smartphone and smartwatch to collect data. This passive sampling of data is invisible, the user does not need to disrupt their daily routines to input data into the system.

EudaeSense is accurate - our solution makes full use of the multiple sensors on their smart devices to collect multiple data points, which gives the platform a larger amount of data to train machine learning algorithms on compared to existing solutions.

EudaeSense is personal - through the usage of the user's personal data across multiple devices, they will receive individualized reports tailored to their profiles and personalized to meet their emotional needs.

EudaeSense is timely - our solution leverages machine learning algorithms to introduce micro-interventions at the most appropriate time for small, simple behavioral interventions to ward off depression.

3.2 Prototype Application and Hardware

The EudaeSense application was created for the Android platform, utilizing both the Android smartphone and Wear devices. Sensor data is collected from the smartphone using the ES Sensor Manager and ES Data Manager Libraries. These open source libraries for Android were developed as part of the EPSRC Ubhave project for the purpose of collecting sensor data from participants of human interaction or social psychology experiments. [http://emotionsense.org] The ES Background logger uses an asynchronous delay tolerant data transfer strategy which significantly reduces drain on the user's battery.

Wearable sensor data is collected through an additional wearable device API. For our initial application we are using data collected from Fitbit, but plan to transition to an open reference design platform such as Samsung Simband [https://www.simband.io] when it becomes available. As it stands now, due to hardware limitations we are constrained to using sensor devices with proprietary black boxed algorithms. These devices such as Fitbit and Spire can not be used for biofeedback interventions do to the required data sync and time delay of accessing remote servers.

Data from the smartphone and wearable device are transmitted to a centralized server for storage and processing. Currently, the server processes the data using simple algorithms designed to illustrate the capability of the application to trigger appropriate activities. Research outlined later in this paper will discuss general machine learning models that will be implemented in future iterations. Data suggests personalized models instead of general ones should be used [21], however, machine learners can become too “locked in” [13]. This area requires additional research and will be included in additional iterations of the application. It is our assumption that as we collect more user data and our machine learner models become more advanced we can better tailor our interventions to users, creating a truly contextually aware system.

The EudaeSense application accesses the server via a RESTful API built using the Node.js framework. In most cases the application is designed to display interventions to the wearable device which may prompt the user to open the application on their phone.

An example of an intervention activity: data pertaining to sleep is collected from the wearable device, the server maintains information regarding when the user typically goes to bed and wakes. Our algorithms calculate the optimal time that a user should go to bed based off their historical sleep data and calculated sleep efficiency. If a user is up past the time they should go to bed for optimal sleep, they will be sent a notification with an intervention designed to encourage them to wrap up their day.

3.3 Interventions

All of our brainstormed interventions fell into one of 3 psychotherapy groups: relaxation and mindfulness, positive psychology, and cognitive behavioral therapy. In total, we produced high fidelity prototypes for 4 of the interventions from our *realtimeboard* using the prototyping tools *Sketch* and *Invision*. Each of the interventions start as a notification on the smartwatch, and transitions over to the smartphone. Screenshots of the prototypes can be found in the Appendix section and descriptions for each intervention are provided below.

3.3.1 Three’s the Charm

This intervention is based from positive psychology and cognitive behavioral therapy. The goal is to help our user end the night on a positive note by going into the app and writing down three different types of positive thoughts: someone they are grateful for, a strength they possess, and what they are looking forward to tomorrow. The idea is that by doing this exercise, the user is forced to think of positive thoughts. The act of writing down the thoughts also serves to help reframe the user’s state of mind and to combat negative thoughts, which is a core element of cognitive behavioral therapy. This intervention requires the use of the smartphone -- the user receives a notification on their smartwatch to complete the Three's the Charm exercise, which then leads them to open the application on their smartphone with the "Open on Phone" action button.

Three's the Charm is an intervention that also makes use of behavioral psychology heuristic the of availability bias. The availability heuristic is a mental simplification in which one mistakenly refers to the immediate situations or examples that spring to the forefront of the mind. As a result, one can quickly judge that more readily available events are more frequent and possible than others, giving inordinate credence and overestimating the likelihood of similar events happening in the future. By bringing to mind more positive events from the past, one could be more apt to assume that these positive events would happen again in the future, according to the availability heuristic.

3.3.2 Positive Reminder

The positive reminder intervention collects ESM (experience sampling method) data from the user by asking them to rate their mood. This then goes towards our machine learning algorithm that matches biosensor data with ESM data to predict user's mood. If the user rates their mood on the lower end of the 5 point scale (3 or lower), they are presented with a positive note they had previously written in a Three's the Charm entry. We believe this is more powerful than simply presenting the user with an arbitrary positive quote, as reviewing something they had previously written is more personal and relevant to their lives. This is primarily a smartwatch intervention, meaning that the user does not have to open up the smartphone to complete this activity. However, the user can still open this intervention on the smartphone to view more positive notes that they've previously written.

This positive reminder exercise could help to balance the anchoring effect. Sometimes people may be 'anchored' into thinking that they are in a bad mood all the time. This exercise would hopefully correct that thinking by reminding them of prior moments when they experienced happiness or future moments when they expect to experience happiness.

3.3.3 Controlling Worry

This intervention serves to help our users control worry and feelings of guilt associated with the inability to fall asleep. This is a very context specific intervention -- it is designed to be used when the user is having trouble falling asleep. The smartwatch picks up on the user struggle via biosensing and application algorithms, and sends them alternative ways of thinking to combat their negative thoughts. The user can also open this intervention up on the smartphone to review more alternative thoughts.

This type of intervention utilizes reframing as a tool to encourage System 2 thinking and reduces negativity bias. System 2 thinking is a theory popularized by Daniel Kahneman purporting that system 2 thinking tends to be more conscious, effortful, and better for making complex decisions. System 1 thinking on the other hand is more unconscious, spontaneous, automatic, and error prone. Sometimes people may unconsciously start thinking negative thoughts with negative bias, which can cause anxiety and worry. The Controlling Worry intervention attempts to invoke System 2 so that one can think rationally about their thoughts about whether their negativity bias is actually rational or if there is another way of thinking that is less anxiety-inducing and emphasizes more of the positive rather than the negative.

3.3.4 Blowing on Dandelions

In this intervention, a soft spoken audio guides the user through a mindfulness exercise, which also invokes System 2 thinking. The analogy here is that each dandelion that appears on the screen represents a thought in your head. Blowing on the dandelion is a metaphor for setting free and allowing the participant to be an observer of their thoughts, as opposed to being confined into thinking that they are their thoughts. This intervention is prompted by biosensing data such as heart-rate and breathing but also using contextual data from the user's calendar, present and historical.

3.3.5 Surprise Hug

The 'surprise hug' intervention takes the form of 'surprise' notifications with pictures such as cute cats or other images associated with positive moods. The notifications are not actually 'surprise', but timed for when the user is feeling down or neutral and could use a pick-me-up during the day.

This intervention uses the behavioral psychology principles of surprise and adaptation. People get more pleasure out of a surprise gain than an expected gain. People also tend to adapt to change over time,, and surprise helps to maintain interest and keep up the additional gains.

4 Process

4.1 Exploratory User Research

4.1.1 Expert Interviews

Our first course of action was to conduct interviews with experts in the field of mental health. At this point in the semester, we had begun to realize just how large our project could be and decided that we needed narrow down our focus for the sake of meeting the deadline by May. We had three overarching goals in mind for conducting expert interviews. One, we wanted to gain a better understanding of our users, the realm of mental health and sleep, and the methods that have been most effective in treating depression. Two, we wanted to hear their thoughts and suggestions on our project, and gather some advice on the directions we can take. Third, we hoped to get some pointers and references to additional resources we could look into. This turned out to be a very fulfilling experience for us. Our discussions with experts played a crucial role in helping us narrow down our project -- one of the common themes derived from our interviews is that insomnia and sleep related issues remain one of the most prevalent symptoms of mood disorders like depression. Because of this, we were inspired to shift the weight of our project to concentrating fully on building interventions for sleep.

Since our experts have mostly worked with students, our questions were designed to understand mental health in the context of student life. Through our expert interviews, our goals were to gain a better understanding of the following:

- The type of mental health needs that students have
- The interventions and coping techniques that have worked well
- The interventions and coping techniques that have not worked well
- How they feel about incorporating potential technological solutions into their current practice
- What types of technology they would like to see or feel would be most beneficial
- What their thoughts are on our project

We interviewed 4 experts in total: Dr. Aaron Cohen, a senior staff psychologist who works closely with students on campus who require mental health aid and attention; Dr. Julio Ozores, a psychiatrist who works in the Alta Bates Summit Medical Center and on campus at UCSF, and specializes in the evaluation, psychotherapy and psychopharmacological treatment of young adults and college students; Dr. Nicole

Beasley, a clinical psychologist, who focuses on EFT (emotionally focused therapy) and couples and relationship therapy; and John Chuang, our final project adviser who is an expert in the field of biosensor wearables. In the next section, we will share some of the inspirations and insights we received from the interviews.

Findings

As noted earlier, common subject discussed was sleep. In our discussion of different methods and techniques for treating depression with Dr. Julio Ozores, we learned about CBT-i, Cognitive Behavioral Therapy for Insomnia. This is essentially a branch of the conventional cognitive behavioral therapy that focuses specifically on providing methods for treating insomnia. There are five main components for CBT-i: stimulus (arousal) control, establishing a good sleep hygiene, sleep restriction based on the person's TIB (time in bed), mindfulness training, and cognitive therapy (or sleep hygiene education). On a physiological level, we also learned that there are 3 primary indicators of insomnia: arousal levels (such as from heart rate or galvanic skin response), sleep deficit levels, and circadian rhythm. These pieces of information were revelations for us at the time -- the different categories from CBT-i was exactly the resource we were looking for to our sleep intervention designs in.

We also learned that in a conventional one-on-one therapy session, it is important for the therapist to communicate in a non-judgmental manner. Dr. Cohen cued us into some of the important elements of non-judgmental communication: asking open ended questions, paraphrasing what has been said, and giving the patient a voice. He reminded us that there is an emotional piece to speaking with a human that an application may or may not be able to capture. When we started mocking up our intervention prototypes, we were careful in our attempts to include language that was a) non-judgmental and b) that spoke to our users on an emotional level.

4.1.2 Diary Studies

Following up on our expert interviews, we conducted a one week long diary study. The goal was two-fold: get better insight into the internal and external factors that affect our user's sleep pattern, and gather unstructured and structured data to help us decide on which interventions will be best to create for our prototype. We felt that a diary study - where the user can keep a sleep diary and self-track - would provide us greater insight that other methods would not be able to. The primary focus for this study was to learn about our user's relationship with sleep, and to bring the findings into our intervention designs down the road.

Recruiting and Preparation

One of the early struggles we encountered was that we could not obtain IRB approval within the time frame given for our final project. As a result, we had to make adjustments into both who we recruited to participate in our user research, and also the type of questions we wanted to ask. For our diary study, we reached out to various health and wellness Facebook groups on campus, the I School noise channel, and convenience sampling. We recruited a total of 7 people that consisted mostly of I School students, family members, and friends. Out of the 7 people who agreed to participate, 6 went through the week long study to completion.

We decided to use a digital platform for our diary study, since we felt like it expedited the data entry process. We're aware that the participants are all very busy individuals, and we wanted to provide a platform for them that was easily accessible and easy to input. The vast majority of the time spent in the preparation phase was devoted to crafting the morning the evening surveys. As noted earlier, one of our biggest challenges obtaining IRB approval, so as a team we were more conscious (and at times, slightly paranoid) that the questions we asked were going to cross the line over to unethical territory. Because of this concern, we went through numerous revisions of these forms before launching.

Study Design

In preparation for the study, we created the following documents: an onboarding survey, a morning survey, an evening survey, and an informed consent form. All of these surveys were created using Google Forms, and the data was managed using Excel spreadsheet. On the day before the study began, we asked our participants to complete the onboarding survey, which consisted mostly of questions that were designed to give us insight into the struggles that they have with sleep and if they have more severe issues like insomnia. It is important to note that our goal for asking these questions was not to diagnose our participants -- rather, it gave us the opportunity to ask more in-depth questions regarding their struggles with sleep and what solutions they've tried in the interview following up from the study.

We asked the participants to fill out the morning survey as soon as they woke up in the morning, and to fill out the evening survey before they went to sleep at night. The types of questions in these surveys consisted of multiple choice, 5-point scale, free response, and binary (such as yes/no). Our morning survey was designed to ask questions specifically about the previous night's sleep. Questions such as "how many times did you wake up in the middle of the night" and "how would you rate the quality of your sleep" were asked to get a glimpse into how well they slept the night before. The evening survey on the other hand was used to describe the environment around them and what activities they engaged with a few hours leading up to bedtime. This served to help us key into their sleep hygiene practices and open up discussion in the follow-up interviews about these habits.

In addition, we also asked the participants to send us an SMS message whenever they felt drowsy or particularly tired during the day. The purpose of doing this was to help us key into when and how often they experienced daytime drowsiness. Participants were encouraged to provide a brief description in their text message about where they were and they were doing at the moment.

Follow-up interviews

Upon completion of the study, we immediately scheduled follow-up interviews with our participants. Each interview lasted about 20 to 30 minutes. In preparation for each interview, we wrote a script with specific set of questions for each participant. The interview had 3 phases. The first phase was introduction -- since this was the first time interacting with our participants in person since the start of the study, it was important for us to take some time to establish rapport and create a friendly environment. The second phase was asking questions. We went through each of the surveys they filled out prior to the interviews and highlighted sections that stood out or of which we wanted to dig deeper into. During the interview, we went through the onboarding survey, morning survey, evening survey, and any SMS text messages that they sent throughout the week. In the last phase, we wrapped up by asking them what sort of technological solutions they would like to have to improve their sleep quality. At the end, we offered cookies and snacks as compensation for their participation in the study.

Findings

The insights we received from our participants in the diary study and follow-up interviews were crucial in helping us begin designing sleep interventions for our application. This section summarizes the most important findings we uncovered from the study.

Issues related to sleep

One of the biggest issues we discovered was that participants had trouble staying asleep, not falling asleep. At the top of the list are the following: falling asleep, waking up in the middle of the night, difficulty falling back asleep after waking up, feelings of anxiety and stress associated with not being able to fall back asleep, and feelings of guilt associated with going to bed before finishing work. Although these are not uncommon sleep related problems, it was important for us to know which issues were most prominent as this serves to help us conceptualize the potential sleep interventions we want to include in our application. In addition to the list above, most of our participants also experienced daytime drowsiness throughout the week. Only four participants sent us SMS messages, and all of them stopped sending messages by day 3.

Quality of sleep

One of the things that the diary study could not tell us was what an ideal night sleep looks like to our participants. Our mental model of what “good quality” of sleep looks like has been largely made up of assumptions and guesses. Because of this, we made a note to ask them “how do you define good sleep quality?” in the follow-up interview. This question was particularly important, not only because it helps us understand what a good night sleep is like to our participants, but also because it allows to compare how a good model of sleep looks like next to a bad model of sleep. In retrospect, rephrasing this question to “how would you describe a good night’s sleep” or “what does an ideal good night's sleep look like to you?” may have been more accurate way of asking the question.

What we learned is that good sleep quality is not necessarily defined by being able to fall asleep, but rather how well-rested they feel after they wake up. And how well-rested they feel is dependent on various factors, which include the following: how many hours of sleep they get, how many times they woke up in the middle of the night, how long it took for them to fall back asleep, and their level of stress and anxiety associated with work that they need to do. Although there are other obvious factors such as noise levels, dietary habits, amount of caffeine consumption, etc., these were the ones that were highlighted by our participants.

Use of technology for improving sleep

We discovered that most of our participants have either never used any applications before to help them improve sleep quality, or they stopped using them after a while. Most of the applications that users have used before were meditation and breathing tools, but all of them noted that they have stopped using them. The reasons for this is unclear. Participants noted that these applications stopped interesting them as much after a while, and there seemed to be a greater inclination towards other hobbies and activities. For example, participants cited that reading, playing games on their smart devices, light exercises, and crochet were a part of their nightly routine.

Takeaways from Diary Study

One important takeaway we got from doing the diary study is that our participants experienced some annoyance and confusion in filling out our surveys. Participants noted that the evening survey felt too long and time consuming, and some questions like the ones that asked them what they ate or drank felt repetitive. The morning survey felt like a chore at times, especially in the mornings where they were in a rush to get somewhere. We also noticed that some participants had forgotten to fill out surveys and made up for it by filling it in the next day (leading some data to be not as accurate as others). Others forgot to fill out some surveys completely. This was a crucial finding in itself, because on a conceptual level our interventions will be similar to these surveys -- that is, there will be certain times during the day or night where we would send notifications to our users and ask them to engage with our application. We will have to design our interventions in such a way that it retains user engagement and interest in the long-term, and that which avoids evoking feelings of annoyance or confusion.

4.2 Intervention Design

Following the diary study, we started to put our design interventions on paper. To do this, our team went through a week long sprint brainstorming ideas and drawing them on paper. After we had drawn all of our ideas down, we uploaded them to an online tool called realtimeboard. Realtimeboard is a nifty tool that allows us to organize and view all of our drawings in one place. During this time, we held multiple meetings to pitch our ideas and consolidate designs. All of our ideas fell into one of 3 psychotherapy groups: relaxation and mindfulness, positive psychology, and cognitive behavioral therapy. In total, we produced high fidelity prototypes for 4 of the interventions from our realtimeboard using the prototyping tools Sketch and Invision. Each of the interventions start as a notification on the smartwatch, and transitions over to the smartphone. Screenshots of the prototypes can be found in the Appendix section and descriptions for each intervention are provided below.

4.3 Formative User Research

For our project, we drew from three different research methods: (1) survey, (2) interviews, and (3) usability tests. We decided upon these three methods for several reasons. First, we were interested in using both quantitative (survey) and qualitative (survey, interviews, and usability tests) methods. The survey allowed us to get responses from a greater sample of people than we would have otherwise been able to. It also allows us to obtain with some baseline information about users which we can then compare against via the other methods. The interviews allowed us to dig further into the content of our research questions. With regards to usability tests, we felt that this method would allow for efficient (i.e., the ability to uncover usability issues with a limited number of participants) discovery of potential usability issues. In the section below we describe in detail the process by which each method was carried out. The section ends with a brief discussion of limitations.

4.3.1 Surveys

We developed a survey with about 20 questions using Qualtrics, a web-based research platform that aimed to address our research questions and also provide generative insight for the EudaeSense Prototype. As suggested by Goodman, Kuniavsky, and Moed (2012), we aligned our questions with our research goals. The questions were divided broadly into two sections, each addressing one research question: one focused on the user's relationship with sleep and one focussed on the user's relationship with technology. We prioritized keeping the survey to a reasonable length, with mostly multiple choice questions and no question asking for more than a three word response.

The survey was open for about one week. We had a targeted having 25 respondents and ended the survey administration with 48 participants.

Survey Findings

After conducting our survey and analyzing our data we were developed three main findings:

1. People reported their sleep habits in different ways due to varying definitions of sleep and sleep quality.
2. Respondents seemed to have basic understanding of the relationship between mood and sleep.
3. The majority of respondents had morning and evening routines and indicated general awareness of good and bad sleep habits.

With regards to the variability in reporting (Finding #1), it is important to note that our survey design did not allow for a statistically significant analysis of our data. Achieving significant results, to a degree, wasn't our focus. However, we did want to be able to obtain enough responses such that we felt comfortable utilizing our descriptive results for more general insights.

Our survey results indicated that our respondents were aware of the important (cyclical) role that sleep plays in affecting one's mood (Finding #2). For the prompt, "Sleep patterns affects one's mood," the average rating was 1.33 (with 1 = Strongly Agree, 2 = Agree, 3 = Neither Agree nor Disagree, 4 = Disagree, 5 = Strongly Disagree). Somewhat similarly, for the prompt, "Mood affects sleep patterns," the average rating was 1.76.

For the third survey finding, the majority of respondents indicated that they had both morning and nighttime routines which they undertook prior to going to sleep.

4.3.2 Interviews

After carrying out the survey and conducting the ensuing analysis, we drafted questions for our directed interview protocol which contained roughly 20 questions. These questions focused on four areas: morning and evening routines, ideal sleep, technology in general, and wearables.

In our survey, we had asked participants if they would be interested in a follow-up interview. We e-mailed all interested survey-respondents and scheduled the first to respond for interviews of about 30 minutes. Seven of the 48 respondents to the survey indicated that they were interested in participating in a follow-up interview. After dealing with responses to follow-up contact and scheduling, we conducted four interviews in all, with each taking about 30 minutes. For each interview we provided participants with a $5 Amazon gift card. Three of the interviews were conducted on the UC Berkeley campus (South Hall). One interview was done via Google Hangouts. For reference, we have included our interview protocol in the supplementary document.

For each in-person interview that was conducted we had two members of the team attend, one to take notes and one to ask the interview questions.

Interview Findings

As Lee (2000) discusses, participants like to be seen in a positive light [23]. Within the surveys and surface-level interview questions, participants talked about having strict morning and evening routines, not using technology before trying to sleep, and other positive sleep habits.

When asked in more detail, participants revealed that although they have these sleep habits in mind, in reality, they don't often abide strictly by them. Things come up: school, work, last-minute deadlines. Then things like morning and evening routines fall to the bottom of a participant's priority list. Lee (2000) argues that interviews are not the best way to get past these pressures to impress an interviewer, but our interviews seem to have gotten us at least a bit past these obstacles [23].

Even when participants aren't claiming to have healthy sleep habits, they seem to know the right answers: not to use technology right before bed, to stick to a routine, etc. Although it may be due to the topic at hand, there was a particular emphasis on knowing not to use technology before going to bed.

Overall, participants know what to do, they just haven't figured out how to make all the pieces fit together in their day-to-day lives.

One of the more interesting findings from the interviews, was that participants didn't mind being bothered by multiple notifications as much as they minded the notifications distracting others they were interacting with. One interviewee reported that when he receives notifications on his apple watch, people think he isn't paying attention to the current conversation. Of course participants also didn't love the idea of being bombarded with multiple notifications, but it was interesting to hear about the impact it has on other people's perception of them.

Participants interested in well-being technology also showed a preference for smartwatches as opposed to fitness trackers. Participants were interested in the connectedness of different technologies (phone, tracker, etc.) and measures rather than just one metric (i.e. Fitbit). Multiple users expressed concerns about the accuracy of Fitbit specifically and about the discomfort of wearing trackers while sleeping.

4.3.3 Usability Tests

For our final user research method, we wanted to tie all of our findings back to our product. To do this, we conducted usability tests on high fidelity prototypes created using Sketch and Invision. The findings from our surveys and interviews served to guide and define which interventions we wanted to prototype. Our goal for this phase of the project was to work with our users to initiate the iterative design process for these interventions. We wanted to gather preliminary findings about 2 specific things: which features captivated our users the most, and what the most appropriate times are to send these interventions to our users.

In preparation for the usability tests, we assembled the following documents: a script that included a list of instructions and questions, two sets of interactive prototypes on Invision that were linked to each other (one for the smartwatch screens and one for the smartphone screens) to emulate the cross-platform interactivity of our application, and an audio/video recording consent form. The recruiting for our usability study was not difficult -- we reached out to people who participated in the survey and/or the interview. We also conducted two pilot tests - one within our team, and one with an actual participant - and continued to revise our script throughout the study. Each session took between 45 minutes to an hour, and afterwards participants were rewarded with a $5 Amazon gift card.

Our protocol followed 3 main parts. The first part was devoted to introduction and establishing rapport -- we welcomed our participants, introduced ourselves and the project, requested their consent for audio and video recording, and asked them to apply the think-aloud protocol when completing the upcoming tasks. We were careful with our wording and delivery for this part, as we wanted to create an environment that is friendly and as comfortable as possible for our participants. In the second part, we turned on video recording using an application called Camtasia and asked the participant to go through each intervention. As we presented each screen, we would ask questions to gauge their initial reactions. For example, we would ask them "what are some things your eyes are drawn to when you see this screen?" and "what are some thoughts and feelings evoked when you look at this screen?". As the participant stepped through the intervention, we would ask questions along the way, and towards the end of each task, we would ask them questions related to timing and feelings evoked, such as “at what time during the day or night would you most likely engage with this intervention?” and “how did you feel going through this intervention?”. In the third and final part, we wrapped up by asking them about their usability test experience and if they had any feedback or ideas for us about the product.

Usability Test Findings

The responses and feedback we received from our participants were illuminating and insightful. In general, they exhibited strong emotional responses to the prototypes. The findings for each of the interventions are presented below.

Three’s the Charm

The Three’s the Charm was the first intervention presented to our participant’s. One common struggle that all participants had was with navigation -- basic navigational elements such as saving, going to next page, returning to previous page, returning back to the main menu, were difficult to find and not intuitive. This is a simple but crucial design fix that we will incorporate into our next iteration of designs. Participants expressed mixed reviews for both the length and content of the instructions text provided within each of the three entries -- some felt that they were too long and tedious, while others showed delight in reading them. Participants suggested adding the ability to hop through entries, and only write entries that they want to write. In similar vein, all of the participants also expressed a great interest in adding a feature that allows them to view all the entries they had written.

As for timing, users noted that they would typically engage with this exercise towards of the end of their day, as they are winding down and getting ready for bed. One thought-provoking suggestion was made by a participant, who suggested that we can ask the user to write in a “looking forward” entry, and then send what they wrote as a notification back to them in the morning the next day as a quick boost. We’re excited about incorporating this idea into our next set of designs, as it can potentially lead to greater adherence and engagement with our app.

Note: The issue of privacy arose as we were discussing our usability findings with John Chuang. Adding a filtering mechanism into our app to validate the content of our user’s entries and ensure that what they write is positive is an attractive feature to have, but doing so sacrifices user agency and also is an invasion of their privacy.

Positive Reminder

There was a strong positive response to this intervention. Most of the user’s expressed delight when they saw the “Remember this?” screen, and all of the user’s instantly made the connection that this was something they had previously written in the Three’s the Charm exercise. Although they really liked the positive reminder feature, participant’s also expressed an interest in being able to self-set the type of reminder they receive in this particular exercise. Participants were concerned that if they were presented with something that was no longer relevant or brought up bad memories (i.e., “what if Maries and I are no longer talking?”), that this could have a negative effect on mood. It was also suggested that the ESM question could be better worded (i.e. what does “bad” and “good” mean?). When the participant’s reached the last screen for this exercise on their smartwatch, they were surprised to see the “Open on Phone” action button, as they were not expecting to continue this activity on the smartphone.

Controlling Worry

This was perhaps our most debatable intervention amongst our participants. One participant was very passionate about taking this exercise out of our app, noting on multiple occasions that this was a “huge no-no for sleep hygiene”. Her argument was that interacting with a bright screen on the smartwatch - let alone opening up the smartphone to resume this exercise - when laying in bed and already struggling to sleep has greater detrimental effects than anything positive this intervention may have to offer. A couple of the participant’s expressed concern for the color scheme for this particular exercise -- they felt that the colors were too bright for the context this intervention was meant for. While one participant did find the alternative thoughts helpful, most of them expressed frustration with the texts. One participant described the experience by noting that “I’m reading this and getting more pissed off”.

Blowing on Dandelions

This intervention was mainly theoretical -- we did not have much user interaction components in this activity because the actual audio and the ability to blow on a dandelion through the phone are features we were unable to integrate into our prototypes. However, we received strong positive responses to this intervention. All of the participant’s expressed great interest in the name of the intervention and concept. However, they also noted that this is an activity that they would mostly engage with in private and not in public, due to the potentially “awkward” act of blowing into the phone and the requirement for a quiet surrounding.

4.3.4 Reflections

Data for this research study was gathered specifically for the purpose of exploring usability research techniques for UC Berkeley School of Information class INFO 214: Needs and Usability Assessment. This research is not intended for publication.

Although CPHS/OPHS approval was not required, we decided to tread carefully around how specific we wanted to ask our research participants about their sleep habits and mood. Because we didn't want to elicit negative memories around sleep and didn't want our participants to feel uncomfortable we made the choice to ask more general questions about sleep in both the survey and interview.

Another constraint was the timeline we had to work with. Because of this timeframe we weren't able to carry out more longitudinal methods and instead had to limit the amount of time we spent on recruitment and screening of participants.

Lastly, had we had a budget to work with we would have potentially been able to carry out a few additional research methods in order to better triangulate answers to our research questions. For example, we did not conduct any heuristic evaluations of the EudaeSense prototype nor did we carry out any competitive reviews. Both of these research methods could have bolstered and enhanced our findings.

Survey Reflections

The time and energy necessary to carry out a high quality survey is much more than we anticipated. For our survey we spent a couple team meetings in which we developed and discussed possible questions. While we attempted to construct questions aimed at our research goals, in some cases we also wanted to simply gather some baseline data about our respondents (particularly in this case because we were also using the survey as tool to recruit interviewees).

Interview Reflections

With our interviews, it was clear that rapport is a key element. Upon listening to the audio recordings, it seemed clear that sometimes we jumped into the interview head-on and didn't spend much time establishing a connection. Part of this might have been due to the fact that sometimes the interviewee and interviewer knew each other from the MIMS program or simply from being on campus.

For the most part, the person carrying out the interview had hard copies of our questions (while the designated note taker took notes on their laptop). Not only was this more personal, it also just made eye contact easier. Here, the importance of visual engagement can't be overstated and having a laptop between the interviewer and interviewee simply made rapport harder.

When conducting interviews our participants really responded to the times when we provided them with neutral confirmation of their answers. Some participants more than others seemed to be more affected by the lack of confirmation (rapport) and instead seemed to crave more expressive and personable feedback.

Usability Reflections

A reflection from our usability testing involves extracting and making sense of the observation data we collected from the usability tests. To the extent possible, after each test, we tried to discuss our main observations and made general comparisons about how the user progressed through the tasks. At times, this was straightforward. However, the real challenge was formalizing these discussions in a way that allowed us to get at the deeper causes as to why users completed tasks in certain ways. Trying to extract insights from our four usability tests was at times fairly challenging given that we were working with rough prototypes (i.e., limited data).

4.4 Data Collection & Predictive Modeling

4.4.1 Introduction

Many of today's mood tracking mobile apps have major drawback: they frequently require users to manually enter data. We surmised that there was a better way. Instead of using self-reported data, could we instead take data from activity tracking devices -- which have seen increasing adoption in recent years -- and infer a user's mood?

The results of a machine learning application like this could potentially have a large positive impact for an individual who suffers with depression. Such individuals who use a mobile app to track their mood will usually find that they are constantly asked to assess and report how they are feeling. With this data, apps will deliver advice, words of encouragement and/or behavioral interventions. We see a potential opportunity for improvement. If it is possible to accurately use passive data collected via a wearable device, we could deliver such interventions without requiring the manual data entry that is largely the cause of high dropout user rates with similar apps today.

This brings us to our research question: can we use activity data to accurately predict an individual's mood? There is a growing body of research on measuring affect using smartphone sensors. Successful models have been developed to predict mood and sudden mood changes from smartphone usage patterns [22], and inference of depressive states from mobility patterns [21].

There is reason to believe that we could have similar or better success deriving models using wearable sensor data. Again, we saw an opportunity to improve upon the work that has already been done. The relatively accurate (70-80 percent) mood prediction models that employed mobile devices had inherent flaws. For example, women tend to carry their smartphones in their purses, not their pockets. As a result, in many cases, activity data was not accurately collected.

A non-intrusive wearable device seemed like a more natural fit for this kind of experiment. Because of the high adoption of the Fitbit activity tracker, we decided to recruit participants who owned or were able to obtain and wear this specific device. Our dataset received a huge benefit from the fact that Fitbits track user activity 24/7. This includes during sleep, which is a time that could provide valuable data for our experiment. For example, a future of application of an accurate algorithm might involve measuring a user's sleep quality over a period of time. If the model is able to infer that a user's mood suffers when he or she does not get enough sleep, a mobile app can deliver an intervention that reminds the user to get plenty of rest. We hope that there are a number of similar use cases that could benefit from perfecting the machine learning techniques we tested here.

4.4.2 Data

Our raw dataset was obtained by conducting a two-week study that tracked user activity via a Fitbit device and recorded mood through a mobile app.

Participants

There were 18 participants for the complete duration of the research study. All 18 participants are included in the mood data set. One Fitbit malfunctioned and the data had to be discarded. Different models of devices yielded different data, depending on what kinds of sensors the tracker was equipped with. Fifteen participants are included in the sleep data set, 17 participants are included in the daily activity data set (without elevation data), 13 participants are included in the daily activity data set (with elevation data), and 8 participants are included in the heart rate dataset.

Participants were recruited from the general population over a one week period using Craigslist, UC Berkeley iSchool mailing lists, and Facebook. There were 27 total respondents to the prescreen survey which required that potential participants be at least 18 years old, own or have access to a Fitbit activity tracker, and own an iOS or Android smartphone. Of those respondents contacted 24 signed the informed consent to participate in the study. Compensation for the study was provided in the form of a lottery for a $100 Visa gift card.

Data for this research study was gathered specifically for the purpose of exploring machine learning models for UC Berkeley School of Information classes INFO 290: Deconstructing Data Science and INFO 290T: Data Mining and Analytics. This research is not intended for publication. Although CPHS/OPHS approval was not required, researchers completed the training course for social-behavioral human research. All data collected was labeled with an anonymous participant ID, transferred using SSL and retained on a secure server. Any personally identifying information included with the response from Fitbit was stripped from the dataset.

FIGURE 4.4.1 PARTICIPANT DEMOGRAPHICS - GENDER

#	Answer		Response	%
1	Male		5	28%
2	Female		13	72%
	Total		18	100%

FIGURE 4.4.2 PARTICIPANT DEMOGRAPHICS - AGE

#	Answer		Response	%
1	18-24		2	11%
2	25-30		7	39%
3	31-36		3	17%
4	37-42		4	22%
5	43-48		1	6%
6	49-54		0	0%
7	55-60		0	0%
8	60+		1	6%
	Total		18	100%

Employment:

Seven participants were employed full-time, five part-time and one retired. Seven participants identified as full-time students, two part-time students. One participant identified as self-employed, and one participant identified as a homemaker.

Location:

15 participants resided in California, 1 North Carolina, 1 New Mexico, 1 Wisconsin

Fitbit Data

Fitbit time series data was collected for the two weeks that each individual participated. The original Fitbit dataset consisted of 466,561 instances of time series data corresponding to activity and heart rate. It was represented the following way in our raw dataset [Appendix B].

Mood Data

The research study application was developed using Apache Cordova, an open-source mobile development cross-platform framework. This allowed the development team to design and implement the research application in a short time frame, and target potential participants that use either Android or iOS smartphones. Participants were given detailed instructions on how to access and download the research application via a secure dropbox. Of the 18 participants that successfully installed the application, 8 downloaded the Android APK, the other 10 downloaded the iOS IPA.

Mood data was collected using Experience Sampling Method (ESM) also referred to as Ecological Momentary Assessment (EMA). As opposed to retrospective reporting, ESM/EMA is collected in real time to reduce the effect of recall bias [24]. Participants that downloaded our application were prompted via a notification on their smartphone, randomly four times per day to complete a mood survey designed to assess their current mood. The questions were derived from the two dimensional Circumplex Mood Model [13, 25] which corresponds to dimensions of pleasure and arousal. We collected responses using two UI variations: multiple choice and a continuous slider [Appendix C].

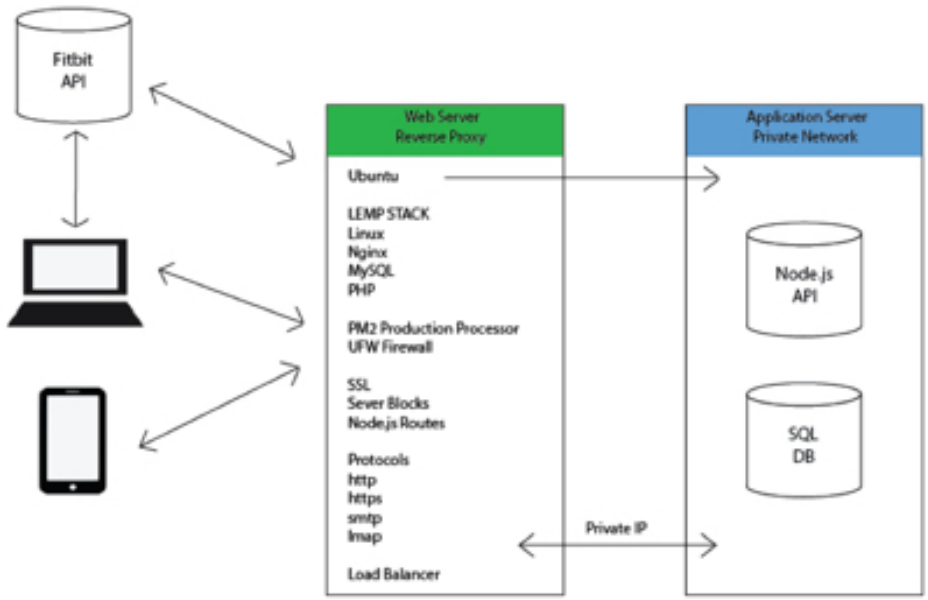
Two Datasets

As previously mentioned, different Fitbit models resulted in the reporting of different kinds of data. Each individual on the team was responsible for sub-setting the data in a way that would help her to best analyze it. In general though, the models were derived from two main datasets: one that included intra-day activity data — which was represented in our dataset each time a user completed a mood survey — and another that aggregated activity data and mood metrics for an entire day. The intraday dataset included 15 columns and 852 rows, and the aggregated dataset 54 columns and 259 rows.

Additional Data Collection

Surveys were designed and data collected through Qualtrics. Participants were required to fill out an entrance survey that consisted of questions pertaining to general demographics, Beck's Depression Inventory (BDI), Personal Health Questionnaire Depression Scale (PHQ), and Perceived Stress Scale (PSS). The survey was repeated at the end of the study adding the Sleep Quality Assessment (PSQI).

FIGURE 4.4.3 DATA COLLECTION - SERVER CONFIGURATION

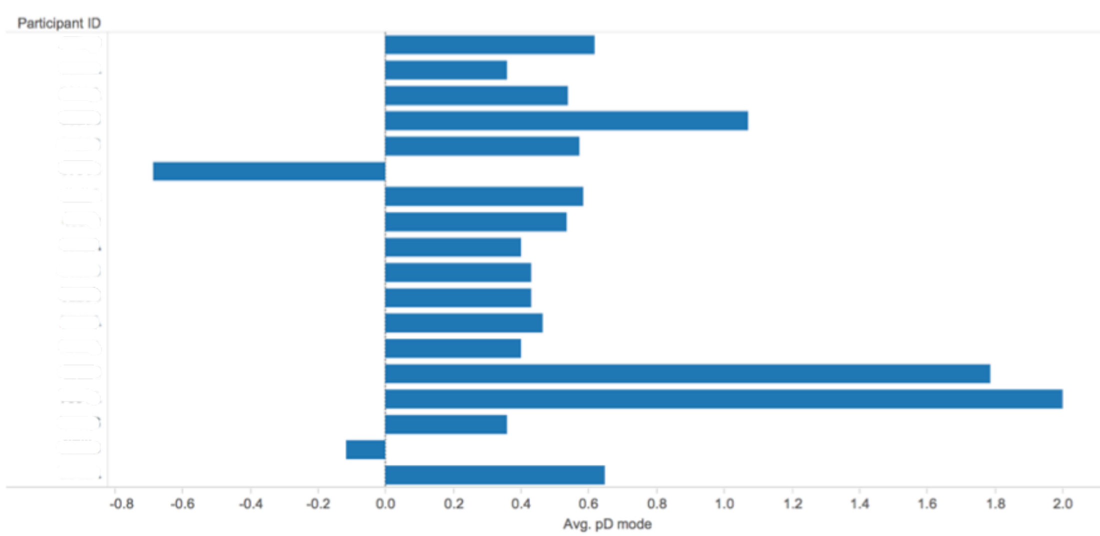


4.4.3 Approach

Data Collection and Cleaning

The infrastructure necessary for capturing the data during the study used Apache Cordova, Ubuntu, Node.js and Fitbit. The data came in as JSON objects, which April reformatted into CSV files. Yiwen used a Python script to match Fitbit activity data to the times when mood surveys were taken in order to create the intraday dataset. As necessary, each individual dealt with missing data. These methods will be expanded upon in our individual write-ups.

FIGURE 4.4.4 PARTICIPANTS' AVERAGE MOOD (-2 TO 2 SCALE)



Those who feel below the “0” threshold would likely be our target users for the application we envisioned.

Machine Learning

Visualizing correlation matrices of the data indicated from the outset that the mood and activity measurements were not very highly correlated. From this, we took that it might indeed be very difficult to predict mood from activity. As a result, we set out to run as many different models with variations of the master dataset as we could. We used the following tools and methods to create the best predictive model.

Sci-kit-learn: Decision Tree, Random Forest, SVM, Linear Regression, Logistic Regression. MATLAB: Neural Network, SVM, KNN, ensembling. NeuroLab: Neural Network

4.4.4 Results and Conclusions

FIGURE 4.4.5 DATA COLLECTION - SUMMARY OF RESULTS

		Data Sets**	
		Intraday Time series activity data Multiple mood ratings	Aggregate Total daily activity Average mood rating
Response Variables Valance and Activeness	Discrete (-2 to 2)	Logistic Regression 40%*	Logistic Regression 76.4%* Neural Network 80.5%* Random Forest 61.9%* Support Vector Machine 56.8%* K-Nearest Neighbor 55.5%*
	Continuous (0 to100)	Linear Regression R-squared: 0.0732	Linear Regression R-squared: 0.885

** Indicates accuracy/ prediction correct rate*
***Data subsets were divided into activity, sleep, and heart rate*

As a benchmark for comparison, the baseline accuracy was 45 percent — predicting the majority class, which was neutral. Many of the other classifiers we tried as group, such as SVM, linear regression, and decision trees, performed worse than the baseline accuracy. Our top results came from using the aggregated dataset (aggregate activity and mood by day) as well as discrete mood data. Our best results came from the random forest, neural network and logistic regression algorithms, with the highest accuracy at 80 percent.

Conclusions

While the results from our initial data analysis on the correlation of our mood and activity measurements were a bit dismal, finding meaning in our data was not impossible. It just took a lot of trial and error. Not only were we able to tell that some models can use activity data to predict mood better than others, we were able to determine the best kind data to do this with. For the most part, continuous data (on a scale of 0 to 100) delivered inaccurate results. However, two models that used discrete data (-2 to 2) resulted in 80 percent accuracy. This fact is useful for future research that will further require manual user entry. Perhaps the inaccuracy of our continuous data experiments was due to our models, but perhaps, in part, it was also due to the fact that users are better at self-reporting data on a 5-point scale. More investigation is required to know this for sure.

We found that the best neural network results came from the subset of data that only included users whose Fitbits reported heart rate data. In future work, we'd like to be able to create a larger dataset with this information. We don't necessarily believe that heart rate (especially the way it is measured by Fitbit's likely imperfect heart rate sensors) is largely responsible for predicting mood. However, a fuller dataset in general might be responsible for the boost in accuracy.

That said, we'd also like to be able to add other features to our dataset. Rather than only looking at activity as a possible way to predict mood, we could factor in other participant details such as demographic information, stress data, and survey data to the qualitative biometric data to see if features like gender or age also contribute to different individuals' mood states. In addition, beyond predicting mood, we could combine this measurement with other dimensions, such as arousal, in order to create different classes of emotional states such as excitement, depression, contentment, or distress. Using a recurrent neural network to predict on multiple classes was one method we were not able to complete during this round of analysis.

Real-World Implications

Finally, in future work, ideally we could make sense of the intra-day data that is available via Fitbit. If the end goal is to deliver timely behavioral interventions to users as their mood states change throughout the day, we need to be able to interpret real-time information. Perfecting these algorithms would make it highly possible that a behavioral intervention application like the one we envisioned could function effectively.

Research shows that when certain individuals experience sustained periods of time during which their mood is constantly in flux, it can lead to various health problems. This includes depression, insufficient sleep and in some cases mental instability. If we can identify the indicators that most strongly infer an individual's mood patterns, we could possibly catch these mood swings early enough and intervene in order to prevent these types of issues.

5 Looking Forward

5.1 Discussion

5.1.1 Privacy

Privacy is a key issue since with granular and sensitive biometric data is being collected.

Since great deal of effort to setup or change privacy preferences, we ensured that the platform was secure and protected their privacy by default from the outset. In collecting data for our study, made sure personally identifiable information was properly disassociated from their personal data in storage, so that the identifiable information for data analysis is separated from the user reporting and interaction. Each user had a unique participant ID that they used so that their real identities would not be associated with the data.

In the EudaeSense application, the users would need to enter other sensitive and private information such as their hopes, strengths, and gratitudes. We would also need to store this information in order to serve it back to the user at the appropriate times. It is important that we build trust with the user to let them know what data we are collecting, how we are storing it, and what we need their data for. We would also need to use their self-reported mood states to continue training and improving the predictive algorithm. This information could be linked back to the user. In order to protect the user's privacy, it would be important to encrypt their personally identifiable information and create unique user codes so that their data cannot be linked back to their identity.

5.1.2 Fairness

The current dataset used to train the predictive mood algorithm has less data about participants with diagnosed mood disorders, which may marginalize them and skew the predictive accuracy algorithm for such depressed people. If this were a full-fledged research study, recruiting for more people with varying mood disorders would help to make sure different personalities are represented in predictive model. In future work, sampling validity could be improved by making sure various minorities and cultures' moods are also included in the data.

5.1.3 Accountability

In the application of this mood prediction algorithm in an interactive affective computing model, there is a risk of false positives or false negatives influencing the user, which we would have to be held accountable as the creators of the algorithms.

False negatives would likely result in a null effect as the application or product would not launch or start. However, false positives could harm the user if the application launches and negatively affects the user's mood. In ensuring that the interactive information device does not make the user spiral downwards even more, the interactions should always be on a positive note, making them difficult to be negatively construed.

5.2 Future Work

5.2.1 Research Study

We hope to give our data and application to an EECS PhD student, who will be obtaining IRB approval to recruit participants with moderate depression or dysthymia to test the EudaeSense system of wearable and mobile device. In this study, the main research questions to address would be:

- How accurate was our predictive mood algorithm?
- How well did the application integrate into their daily lives?
- What were the drop off rates and percentages in comparison to the industry average?
- Did mood demonstrate significant change or improvement improve over the usage of the EudaeSense?

5.2.2 Extension of Interventions

We could continue to extend the number interventions to the application built on top of our mood prediction algorithm. Examples of these interventions include:

- A "smart" music playlist to lift or improve current mood
- Integrations of other phone applications to launch at times of critically low mood
- Automatically initiated suggestions of contact with friends upon low mood state
- A way of sending gratitude notes to close contacts or friends on the platform that display during algorithmically determined times of the day

Users could also be able to add and customize their own interventions if they have specific mood-elevating activities they enjoy doing that they would like to show up at particular times. This would take advantage of the behavioral economics principle of intertemporal choice, stating that people make different choices at different points in time. Someone may make a better choice for their future selves in order to prevent a negative event or mood that they expect their future selves to succumb to.

5.2.3 Extension of Features

Some of the usability testers suggested various features that would be able to help them combat their moods. One of those would be being able to see which interventions were most helpful for them in particular contexts and particular times. Thus, surveying the users about whether or not their mood improved after using an intervention would not only provide us information with whether the intervention was useful but also help the user be more aware of what activities help them to better their mood.

Another potential feature was mood sharing with family and friends, creating more of a support community around their different moods. EudaeSense could suggest people to contact or communicate with certain close contacts when feeling in particular moods. In addition, the EudaeSense app could also automatically ping friends or family to alert them when close contacts are feeling in particular mood and need a pick-me-up message, call, or hang out.

While we intentionally did not report the self-reported moods to back to the users in order to reduce their own over-analysis of their emotions, we also heard that it could help them be more aware of their feelings as well as find patterns in their data. Thus, we could also experiment with providing visualizations of their self-reported mood over time and by day of the week or time to see if that could help them find patterns and better self-regulate their moods throughout the week.

5.2.4 Improving Predictive Algorithms

Sustained periods of bad mood swings can lead to various problems for people like depression, insufficient sleep and to an extent even mental instability. The ability to predict what factors contribute the most to the mood of a person and identifying these mood swings at an early stage could possibly prevent or address various mood disorder problems.

Our result shows that with a logistic regression classifier, we were able to have 76% accuracy on the test set. We can reject the null hypothesis and accept the alternative hypothesis. In other words, we can predict the mood of a person by using a set of data collected from Fitbit.

There are a few ways we can extend our analysis as well as apply our findings from this study. First we could incorporate other features such as demographic information, stress data, and survey data to the qualitative biometric data to see if features like gender or age also contribute to different individuals' mood states. We would also like to experiment with predicting intraday mood using intraday Fitbit activity data and deep learning techniques such as LSTM neural networks. We could also combine mood with other dimensions such as arousal, creating additional classes of emotional states such as excitement, depression, contentment, or distress. We can also apply these predictions to create other products in affective computing -- either reporting the information back to a user or allowing a user to interact with the information in a novel representation as we had shown earlier displaying cat pics, attempting to change or respond to those mood states.

6 Our Team

Richard Chen is deeply passionate about finding a truly effective way for people to fight back and overcome their struggles with depression. His two biggest sources of motivation stem from a background in cognitive science as an undergraduate and his personal experiences with mental health therapy. As a masters student, Richard brings an HCI and UX/UI design skill set, and aims to be a generalist with a specific focus on helping the team scope out the landscape on the existing mood tracking mobile applications, creating prototypes, and conducting user research and expert interviews.

April Dawn Kester has 10+ years of management experience, leading cross-functional teams to meet business goals. With an undergraduate education in economics and computer engineering, she offers unique insight to improving human-technological interactions. April Dawn runs a successful website design company, managing the development of high performance sites, and also has experience developing iOS and Android applications. She has been focused on strengthening technical proficiencies in software development, systems architecture and machine learning.

Audrey Leung is passionate about understanding human behavior in all contexts order to create persuasive technologies that improve the well-being of others. She works at the intersection of business, psychology, design, and technology and has researched, designed, and prototyped several projects involving data analytics, information visualization, and web and mobile applications. Prior to graduate school, she co-founded two startups and managed teams in fraud investigation and auditing. As an undergraduate psychology major and prior volunteer at the Berkeley Greater Good Science Center, she is a strong advocate for emotional and mental well-being.

FIGURE 6.1.1 TEAM PHOTO

7 References

[1] Mental disorders affect one in four people. (2001, October 4). Retrieved October 27, 2015, from http://www.who.int/whr/2001/media_centre/press_release/en/

[2] A Call for Action by World Health Ministers. (2001). Retrieved October 27, 2015, from http://www.who.int/mental_health/advocacy/en/Call_for_Action_MoH_Intro.pdf

[3] Depression Fact Sheet. (2015). Retrieved November 11, 2015, from http://www.who.int/mediacentre/factsheets/fs369/en/

[4] Ben-Zeev, D., Scherer, E., Wang, R., Xie, H., & Campbell, A. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. Psychiatric Rehabilitation Journal, 218-226.

[5] Doryab, A., Min, J., Wiese, J., Zimmerman, J., Hong, J. (2014, June 18). Detection of behavior change in people with depression. Proceedings of AAAI Workshop on Modern Artificial Intelligence for Health Analytics (MAIHA). Retrieved October 27, 2015, from http://www.aaai.org/ocs/index.php/WS/AAAIW14/paper/view/8850

[6] Aver, D., Shah, S., Eder, D., & Wildschindtz, G. (1999). Nocturnal sweating and temperature in depression. Acta Psychiatrica Scandinavica Acta Psychiatr Scand, 295-301.

[7] Canzian, L., Musolesi, M. (2015). Trajectories of Depression: Unobtrusive Monitoring of Depressive States by means of Smartphone Mobility Traces Analysis. ACM

[8] Psychology Of Depression- Behavioral Theories. (n.d.). Retrieved May 07, 2016, from https://www.mentalhelp.net/articles/psychology-of-depression-behavioral-theories/

[9] American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5. Washington, D.C: American Psychiatric Association.

[10] Doherty, G., Coyle, D., & Sharry, J. (2012). Engagement with online mental health interventions. Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI '12. doi: 10.1145/2207676.2208602

[11] Cuijpers, P., & Riper, H. (2014). Internet Interventions For Depressive Disorders: An Overview. Revista De Psicopatología Y Psicología Clínica RPPC, 19(3), 209. doi: 10.5944/rppc.vol.19.num.3.2014.13902

[12] Emotion Sense. (n.d.). Retrieved November 3, 2015, from http://emotionsense.org/

[13] Paredes, Pablo, et al. "PopTherapy: Coping with stress through pop-culture." Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.

[14] Lin, Hsiao-Pei, et al. "Effects of stress, depression, and their interaction on heart rate, skin conductance, finger temperature, and respiratory rate: sympathetic-parasympathetic hypothesis of stress and depression." Journal of clinical psychology 67.10 (2011): 1080-1091.

[15] Sano, Akane, et al. "Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones." Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference on. IEEE, 2015.

[16] Liu, Guan-Zheng, Bang-Yu Huang, and Lei Wang. "A wearable respiratory biofeedback system based on generalized body sensor network." Telemedicine and e-Health 17.5 (2011): 348-357.

[17] Leahu, Lucian, and Phoebe Sengers. "Freaky." Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing - CSCW'15 Companion (2015). Web.

[18] Wang, Rui, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. "StudentLife."Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '14 Adjunct (2014). Web.

[19] Saeb, Sohrab, Mi Zhang, Christopher J. Karr, Stephen M. Schueller, Marya E. Corden, Konrad P. Kording, and David C. Mohr. "Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study." J Med Internet Res Journal of Medical Internet Research 17.7 (2015). Web.

[20] Osmani, Venet. "Smartphones in Mental Health: Detecting Depressive and Manic Episodes." IEEE Pervasive Comput. IEEE Pervasive Computing 14.3 (2015): 10-13. Web.

[21] Canzian, L., Musolesi, M. (2015). Trajectories of Depression: Unobtrusive Monitoring of Depressive States by means of Smartphone Mobility Traces Analysis. ACM

[23] Lee, R. M. (2000), Unobtrusive methods in social research, Open University Press, Philadelphia, PA. (pp. 116)

[24] Trull, Timothy J., and Ulrich W. Ebner-Priemer. "Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section." (2009): 457.

[25] LiKamWa, Robert, et al. "Moodscope: Building a mood sensor from smartphone usage patterns." Proceeding of the 11th annual international conference on Mobile systems, applications, and services. ACM, 2013.

8 Appendix

8.1 Prototypes - User Interface

FIGURE 8.1.1 PROTOTYPES - POSITIVE REMINDER

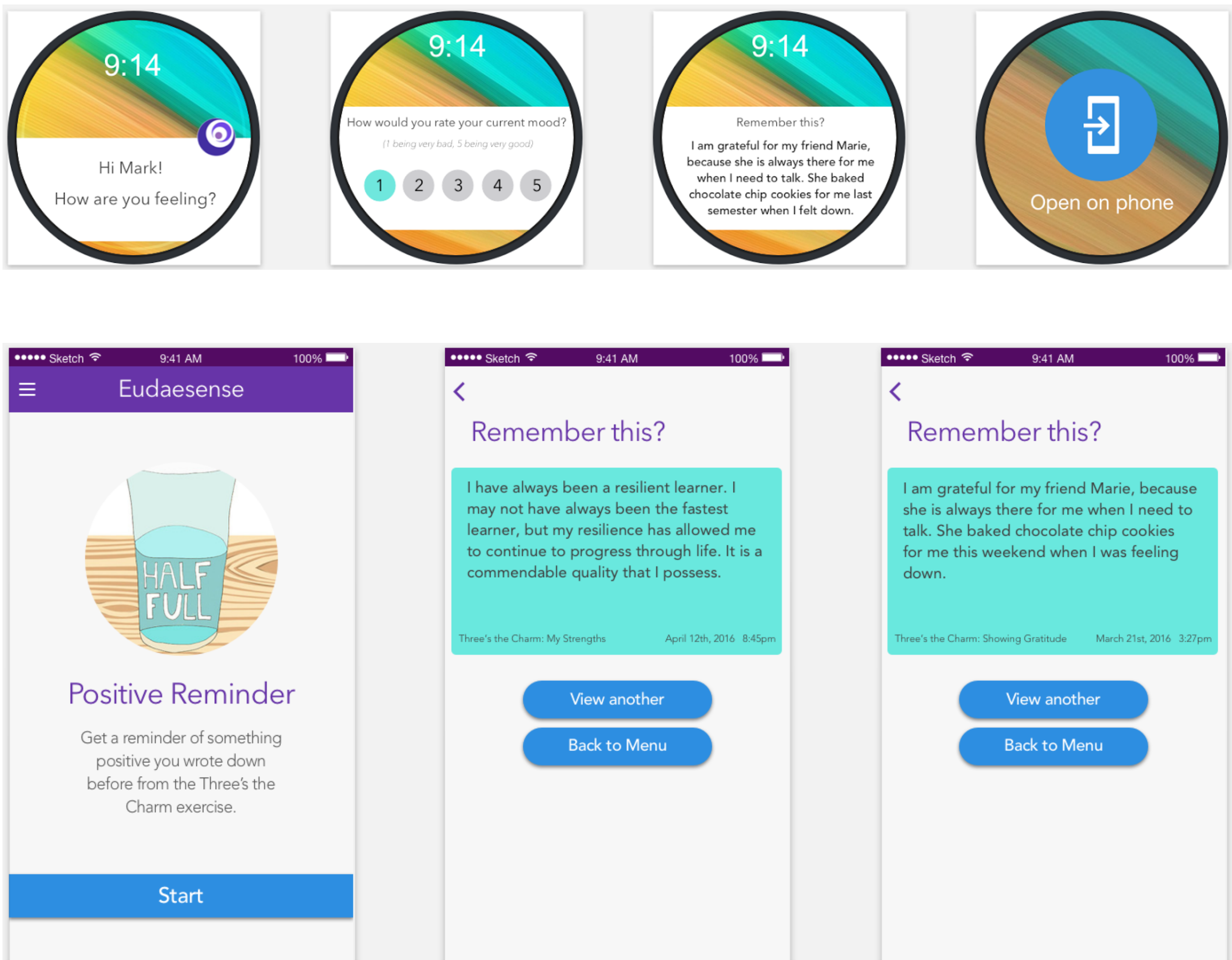


FIGURE 8.1.2 PROTOTYPES - THREE'S THE CHARM

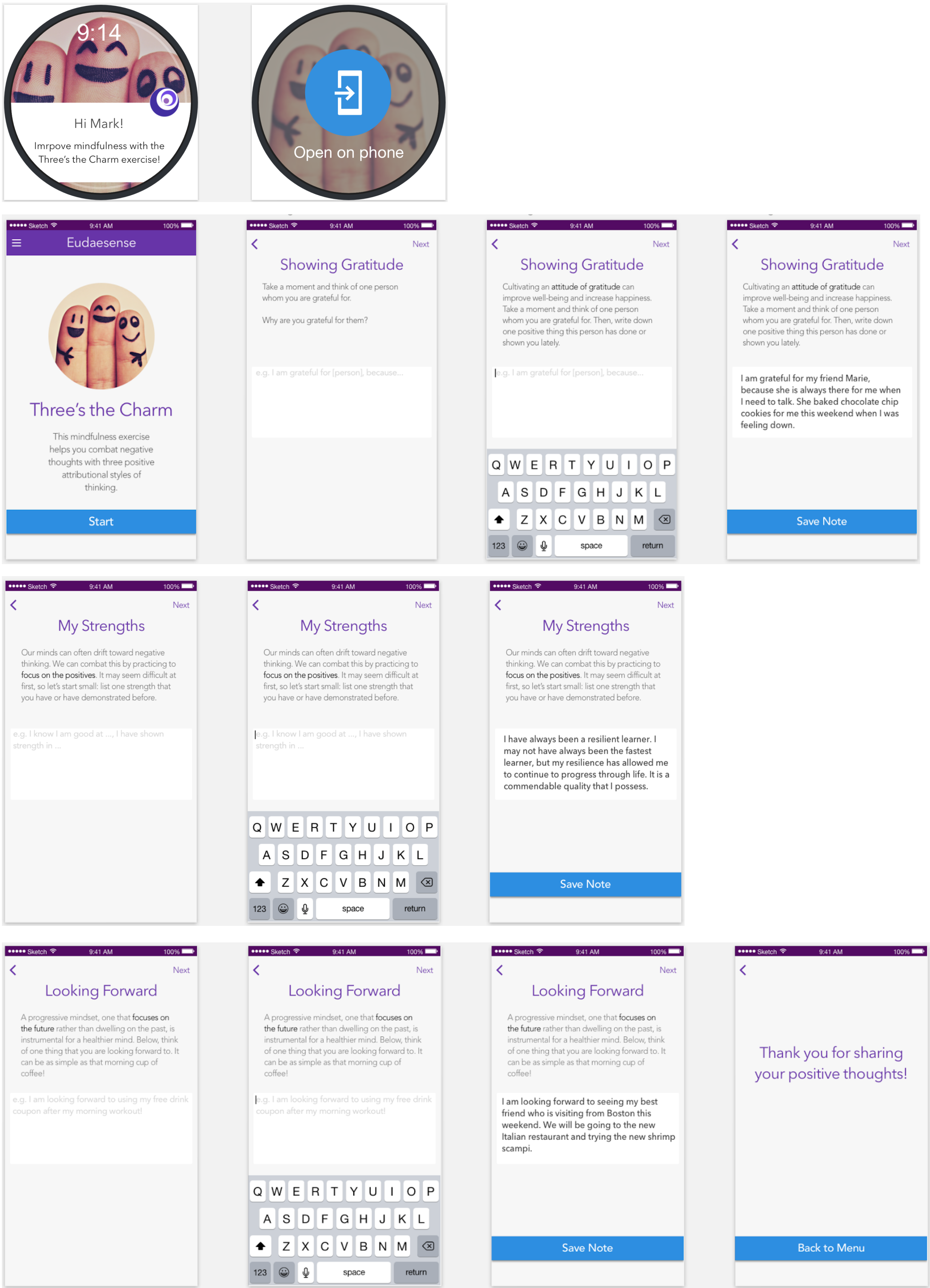


FIGURE 8.1.3 PROTOTYPES - CONTROLLING WORRY

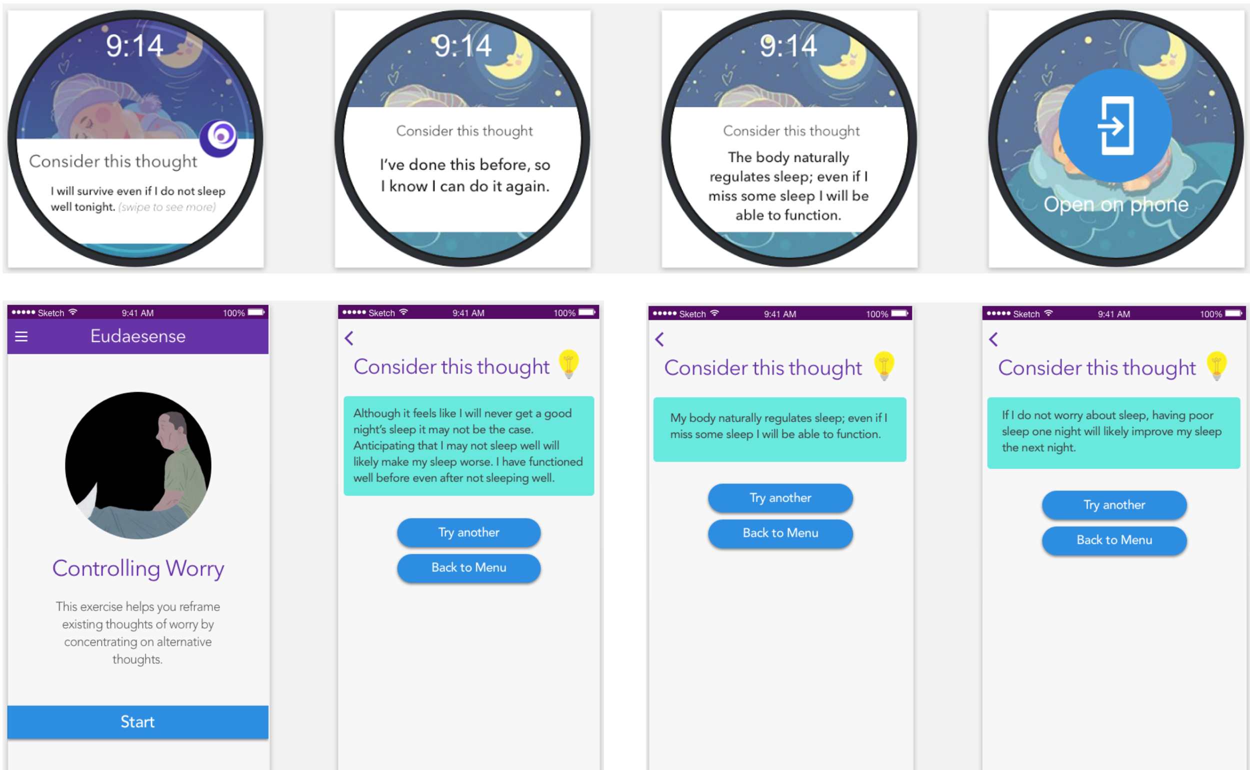


FIGURE 8.1.4 PROTOTYPES - BLOWING ON DANDELIONS

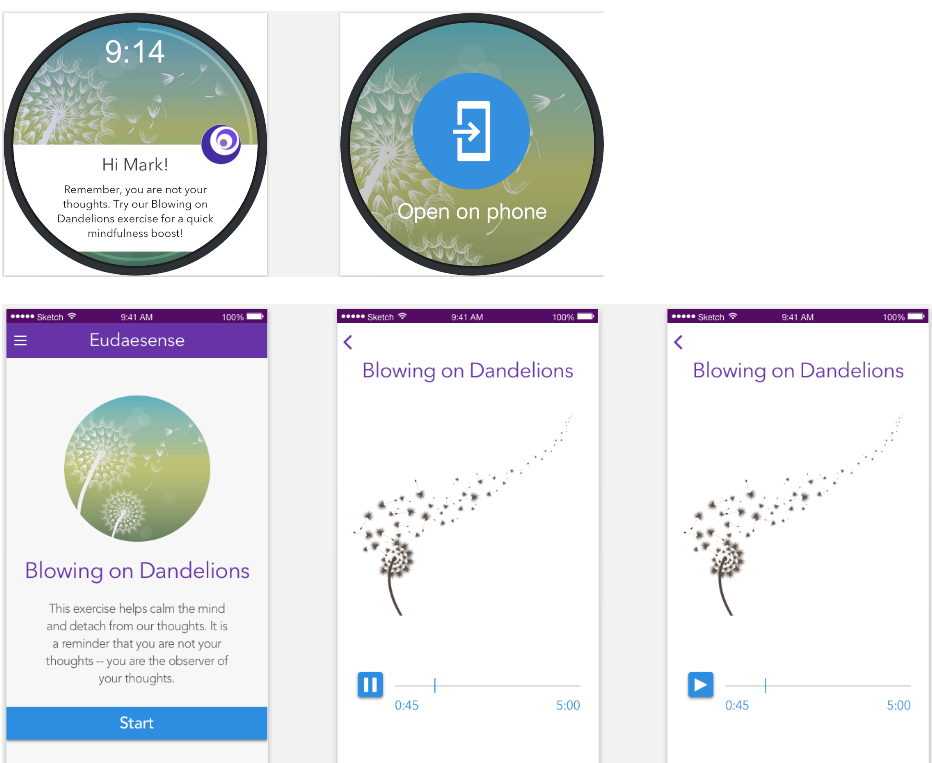


FIGURE 8.1.5 PROTOTYPES - POSITIVE NOTIFICATIONS

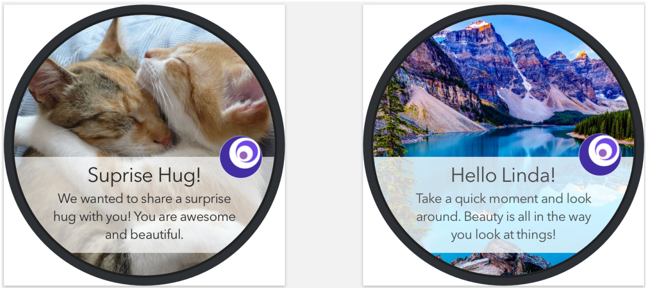


FIGURE 8.1.6 PROTOTYPES - NIGHTLY ROUTINE

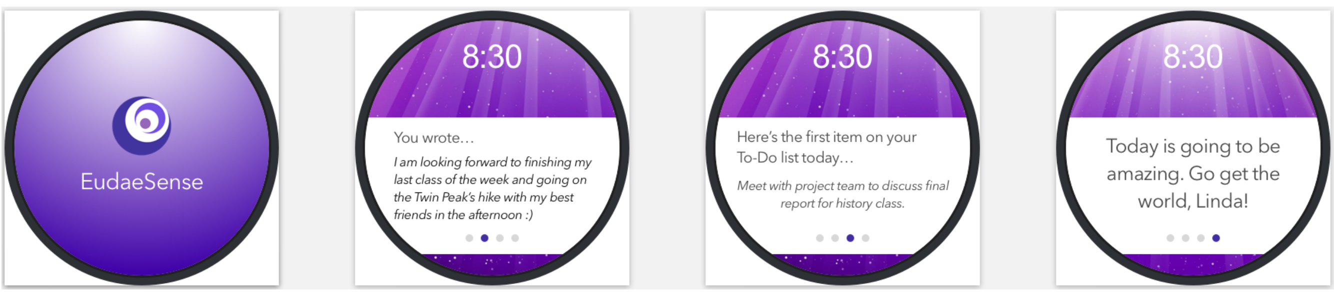


FIGURE 8.1.7 PROTOTYPES - BEDTIME NOTIFICATION

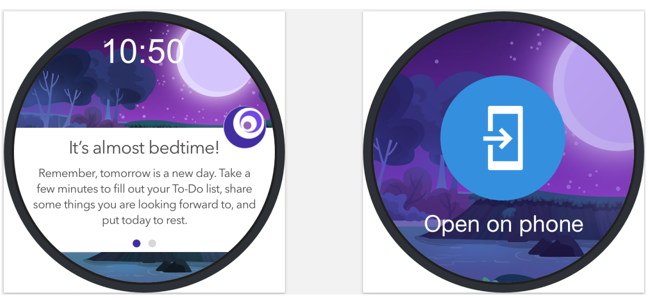
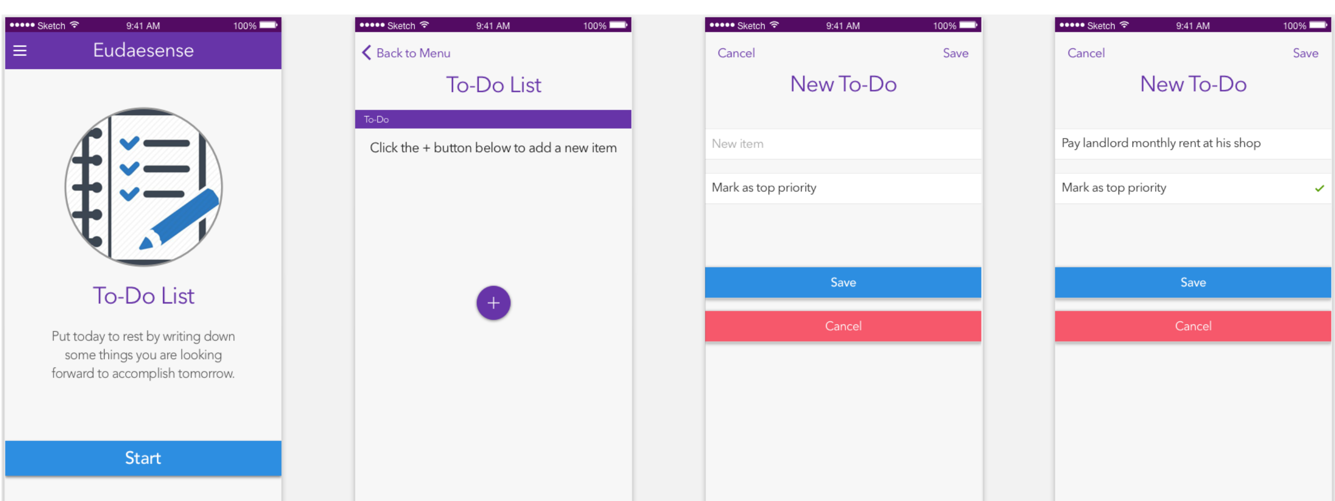


FIGURE 8.1.8 PROTOTYPES - TO-DO LIST



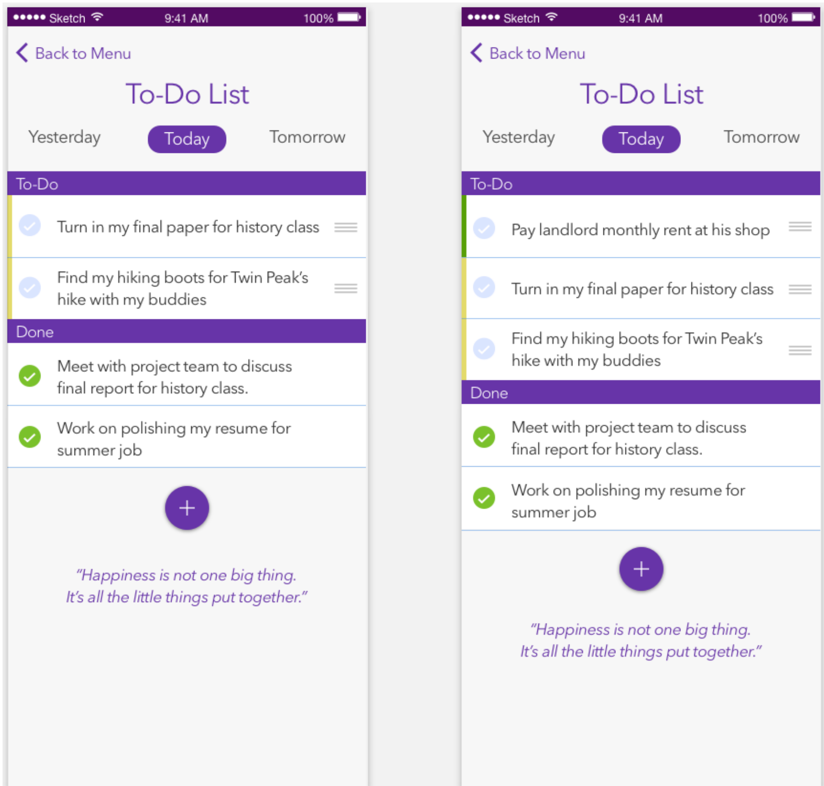
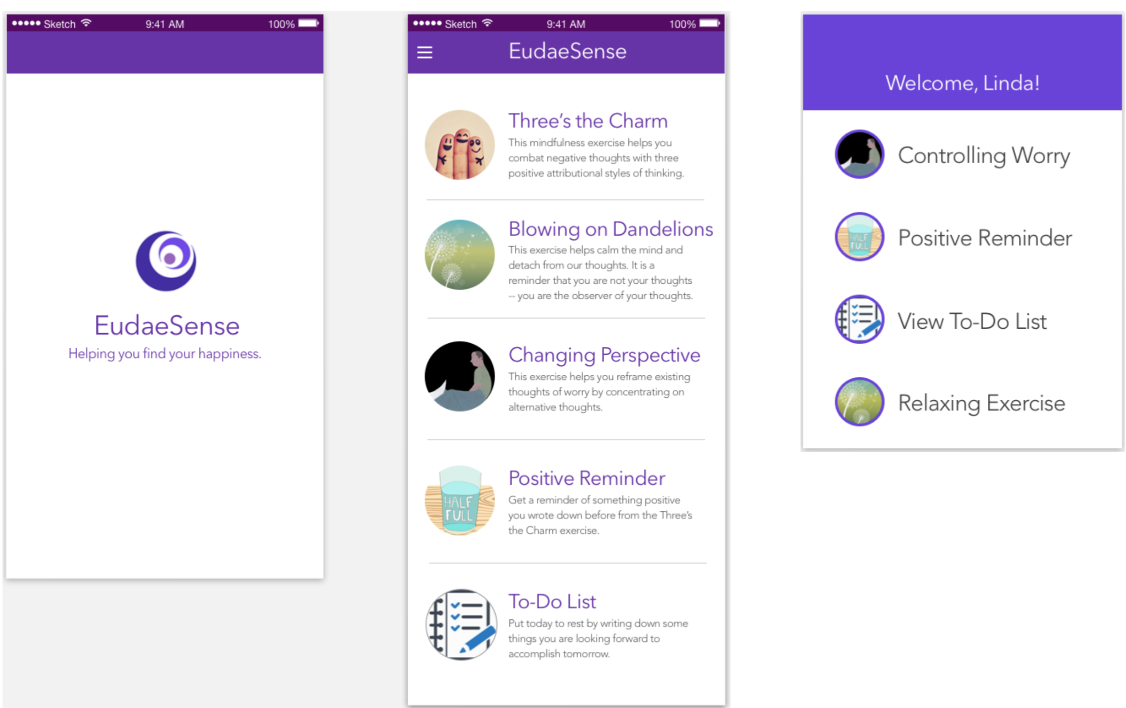


FIGURE 8.1.9 PROTOTYPES - SPLASH SCREEN AND MENU SCREEN



8.2 Tracked Fitbit Data

The following attributes were tracked by Fitbit by individual on daily basis:

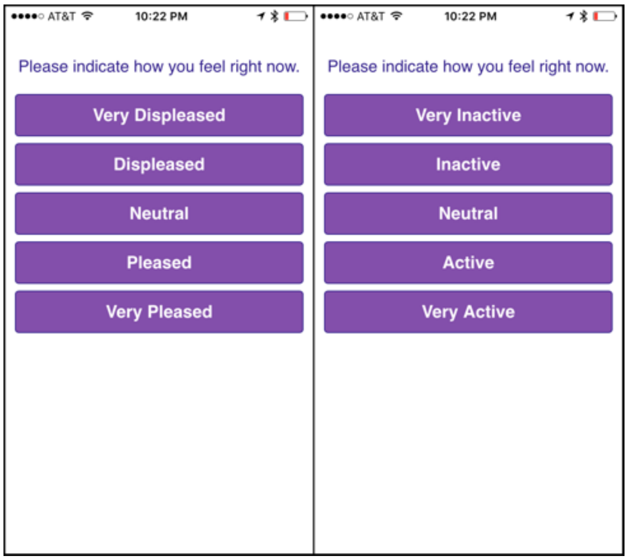
- activities-tracker-calories: calories burned inclusive of BMR according to movement captured by a Fitbit tracker.
- activities-tracker-steps: number of steps taken
- activities-tracker-distance: distance traveled
- activities-tracker-floors: floors climbed
- activities-tracker-elevation: elevation change
- activities-tracker-minutesSedentary: minutes sedentary with metabolic equivalents (MET) between 1-2
- activities-tracker-minutesLightlyActive: minutes lightly active; Fitbit trackers calculate active minutes using metabolic equivalents (METs), which help measure the energy expenditure of various activities.
- activities-tracker-minutesFairlyActive: minutes fairly active
- activities-tracker-minutesVeryActive: minutes very active
- activities-tracker-activityCalories: The number of calories burned during the day for periods of time when the user was active above sedentary level, but uses only tracker data. This value is calculated minute by minute for minutes that fall under this criteria. Manually logged activities are excluded.
- sleep-timeInBed: time spent in bed
- sleep-minutesAsleep: minutes asleep
- sleep-awakeningsCount: number of awakenings
- sleep-minutesAwake: count of minutes in awake state
- sleep-minutesToFallAsleep: minutes to fall asleep after getting into bed
- sleep-minutesAfterWakeup: minutes in bed after waking up
- sleep-efficiency: calculated by 100 * time asleep / (time asleep + time restless + time awoken during sleep) or minutesAsleep / (timeInBed - minutesToFallAsleep - minutesAfterWakeup). The time it takes to fall asleep is not factored into the calculation.
- heart: heart rate

8.3 Experience Sampling Screens

"pleasureDimension"
minResponse: -2,
maxResponse: 2

"arousalDimension"
minResponse: -2,
maxResponse: 2

FIGURE 8.3.1 ESM SCREEN - DISCRETE SCALE



"pleasureDimensionSlider"
minResponse: 0,
maxResponse: 100

"arousalDimensionSlider"
minResponse: 0,
maxResponse: 100

"stressSlider"
minResponse: 0,
maxResponse: 100

FIGURE 8.3.2 ESM SCREEN - CONTINUOUS SCALE