BeWell+: Multi-dimensional Wellbeing Monitoring with Community-guided User Feedback and Energy Optimization

Mu Lin[†], Nicholas D. Lane[††], Mashfiqui Mohammod[‡‡], Xiaochao Yang[†]
Hong Lu[†], Giuseppe Cardone[†††], Shahid Ali[*],Afsaneh Doryab[**],
Ethan Berke[‡], Andrew T. Campbell[†], Tanzeem Choudhury[‡‡]
[†]Computer Science Department Dartmouth College, [*]Dartmouth Medical School
[‡‡]Information Science, Cornell University, [††]Microsoft Research Asia, [†††]University of Bologna
[‡]Dartmouth Institute for Health Policy and Clinical Practice, [**]IT University of Copenhagen

ABSTRACT

Smartphone sensing and persuasive feedback design is enabling a new generation of wellbeing applications capable of automatically monitoring multiple aspects of physical and mental health. In this paper, we present *BeWell+* the next generation of the BeWell smartphone health app, which continuously monitors user behavior along three distinct health dimensions, namely sleep, physical activity, and social interaction. BeWell promotes improved behavioral patterns via feedback rendered as an ambient display on the smartphone's wallpaper. With BeWell+, we introduce new wellbeing mechanisms to address challenges identified during the initial deployment of the BeWell app; specifically, (i) *community adaptive wellbeing feedback*, which automatically generalize to diverse user communities (e.g., elderly, young adults, children) by balancing the need to promote better behavior yet remains realistic to the user's goals; and, (ii) *wellbeing adaptive energy allocation*, which prioritizes monitoring fidelity and feedback responsiveness on specific health dimensions of wellbeing (e.g., social interaction) where the user needs most help. We evaluate the performance of these mechanisms as part of an initial deployment and user study that includes 27 people using BeWell+ over a 19 day field trial. Our findings show that not only can BeWell+ operate successfully on consumer-grade smartphones, but users understand feedback and respond by taking positive steps towards leading healthier lifestyles.

1. INTRODUCTION

Many diseases prevalent in society today are often the result of routine decisions people make on a daily basis; for example, diabetes [32], obesity, stress [28], anxiety [18] are influenced by the choices people make and how they live their lives. We believe once people are equipped with tools to actively monitor and manage the seemingly simple parts of everyday life then they will be able to better assume greater control and responsibility over their health.

In [23] we introduced BeWell, a *wellbeing app* that runs on off-the-shelf sensor-enabled smartphones. BeWell coarsely tracks the physical, social and sleep dimensions of wellbeing by monitoring several key behavioral patterns and providing feedback to the user. Ideally feedback would allow users to easily understand the consequences of their actions, enabling them to make appropriate changes in their behavior and more informed choices going forward. We evaluated BeWell through lab-based single phone experiments that measured the resource requirements of our design. In addition a small five person experiment was conducted to investigate the robustness of the activity inferences BeWell performs [23]. These benchmark experiments, although small-scale, highlighted key barriers to wider-scale deployments. First, we found despite careful engineering and even when using a high-capacity battery, BeWell exhausted the battery life of the smartphone after only 8-12 hours - forcing users to recharge multiple times per day. Second, even with a small number of users we encountered significant diversity, with users exhibiting a wide range of behavioral patterns. As a result of this initial finding, it became clear using a single static expectation of ideal "healthy" behavioral patterns would lead to feedback that was not consistently realistic for everyone; limiting the ability of the system to scale to a larger population of users with diverse health needs.

Guided by these insights gained in developing the original BeWell app we propose *BeWell+*[1], which incorporates a set of new wellbeing scaling techniques for generating community-guided wellbeing feedback and overcoming energy constraints evident in resource limited smartphones:

Community Adaptive Wellbeing Feedback. We design a wellbeing feedback mechanism in which the expectations of healthy behavioral patterns are adjusted to remain realistic for what is possible in the near-term for certain user communities. Instead of only relying on generally ideal (i.e., one size fits all) behavioral feedback (e.g., suggesting 8 hours of sleep) our new BeWell+ design is based on the community the user is associated with – their peer group. For example, it is unrealistic to expect an elderly person to meet the same goals for physical activity as a young adult; or for that matter a doctor that is on call having the same goal for hours of sleep as a high schooler. For each user a "wellbeing network" is identified in the user population, based on shared behavioral traits. Within the network positive and negative "role-models" are identified, and their behavior – along with established ideal behavioral goals – determine user wellbeing feedback. As a result, users are not provided with unrealistic expectations of behavior change, since they are compared to role-models/groups of peers. Here health goals are

Wireless Health' 12, October 23-25, 2012, San Diego, USA

[1]BeWell+ is available for download and use with any off-the-shelf Android Smartphone. Please download BeWell+ from: `http://www.bewellapp.org`

tailored to peer group norms as opposed to a generalized population wide norm. As the health of a user improves they join progressively more healthy communities of users with more challenging wellbeing feedback. We describe this new scalable mechanism as community adaptive wellbeing feedback.

Wellbeing Adaptive Energy Allocation. We design an energy allocation scheme that prioritizes resources so those dimensions of behavior that the individual is currently struggling with (e.g., physical activity) are: (i) more accurately assessed and (ii) provided with immediate feedback, helping to create awareness and promote change in individuals. User behaviors that consistently trend close to healthy norms are monitored less closely, with feedback provided on a slower time-scale – therefore, less system resources (e.g., energy) are required in this case. Using this approach, the more problematic user wellbeing behaviors still receive the attention they demand, while key elements of smartphone usability (e.g., stand-by time) also can remain within ranges acceptable to users. We describe this new scalable mechanism as wellbeing adaptive energy allocation.

Both of these techniques are implemented as part of the BeWell+ app deployment described in this paper. We present results from the first user study of any BeWell app in the wild, which includes 27 people using BeWell+ over a 19 day field trial. Findings from our study show that (1) community-adaptive wellbeing feedback can promote realistic personalized health goals for each user; (2) despite the complexity of multi-dimensional wellbeing feedback, users understand BeWell+ feedback and are able to identify appropriate corrective actions to take; (3) significant increases in energy efficiency result from wellbeing adaptive energy allocation; and, (4) users react positively to their overall experience and even show improvements in their ability to link everyday actions to wellbeing outcomes.

The paper is structured as follows. We discuss related work and then provide an overview of the BeWell+ App in sections 2 and 3, respectively. We continue by detailing the design of community adaptive wellbeing feedback in Section 4. We describe wellbeing adaptive energy allocation in Section 5. We present the evaluation of the system and results from a user study in Section 6. Finally, we make some concluding remarks in Section 7.

2. RELATED WORK

Recently, encouraging progress has been made towards mobile systems that can monitor and improve specific health goals. Various research prototypes have been demonstrated to reliably track a wide variety of key heath factors (e.g., sleep [7], stress [16], diet [30], smoking [27], mood [26]). Similarly, a number of persuasive systems [17] have been designed to assist people in making desired behavior changes and to motivate them to become, for example, more physically active [10, 13, 25]. Commercial activity is also increasing, with products such as, Nike+ [5] and DirectLife [6] becoming more prevalent as mobile health gains mainstream consumer acceptance.

However, a person's wellbeing is shaped by a diverse combination of health and lifestyle factors. Effective personal management of wellbeing requires applications that address a large variety of daily behaviors which have broad health related consequences. As a result, there is a growing interest in building mobile systems that take a broader health perspective. Some approaches rely on developing a software suite of separate mobile applications that manage multiple aspects of wellbeing (e.g., [8, 4]). In contrast, [22] and [30] take a more integrated approach to wellbeing management but rely on manual data entry in the form of a diary to collect information. AndWellness [15] utilizes a mixture of sensor-based activity inferences and manual data entry to provide a general monitoring platform for a range of wellbeing concerns. However, AndWellness is designed to monitor the user rather than promote behavior change. BALANCE [13] also combines user and sensor input to closely monitor multiple wellbeing factors (diet and physical activity), but it neglects other important wellbeing dimensions including emotional and social wellbeing. Finally, purpose-built sensor systems (e.g., Fitbit [2]) can automatically monitor multiple wellbeing relevant behaviors, such as sleep and physical activity while also providing user feedback; but – unlike BeWell+ – these solutions require the user to carry an additional sensor at all times.

Wellbeing Feedback. Prior research has also investigated how ambient displays, different types of goal settings, classifier accuracy, and user interaction affect mobile system's ability to encourage positive behavior changes (e.g., [10, 25, 20]). Ubifit Garden [10], one of the first mobile persuasion system for improving physical wellbeing uses the wallpaper of mobile phones to dynamically provide feedback about the different types of physical exercise performed by the user. Although researchers have recognized certain groups within a user population will benefit from personalized persuasive feedback (e.g., [12, 11]), existing persuasive systems still typically provide the same type of feedback across all users. Under BeWell+, each user receives wellbeing feedback automatically tuned to match their particular lifestyle patterns.

Energy Allocation. One of the most significant practical challenges to the everyday usage of mobile health systems is the resource limitations of smartphones (e.g., battery lifetime). Continuously sensing wellbeing states and providing real-time feedback will consume a significant fraction of mobile device energy. Many proposed solutions consider the general form of this problem and apply resource optimization and/or adaptation techniques (e.g., [14]) to address smartphone energy constraints while executing resource-expensive tasks. Recent research (e.g. [31]) has focused on minimizing the energy cost directly related to mobile sensing applications. However, unlike BeWell+, none of these systems are designed to take the user's wellbeing into account while attempting to optimize resource usage on the phone.

3. BEWELL+ OVERVIEW

In this section, we describe the BeWell+ application and architecture. The BeWell+ application was developed for current smartphones as a proof-of-concept system for monitoring and promoting holistic wellbeing. In our prior work [23] we evaluated an earlier implementation of BeWell+, testing the accuracy of the human activity inferences that it builds upon and its ability to meet a series of system requirements (e.g., battery, computation). But we did not deploy the application or evaluate the system's ability to monitor and provide feedback along different health dimensions.

As shown in Figure 1, BeWell+ consists of two software components: (1) BeWell+ phone app and (2) Cloud infrastructure. The

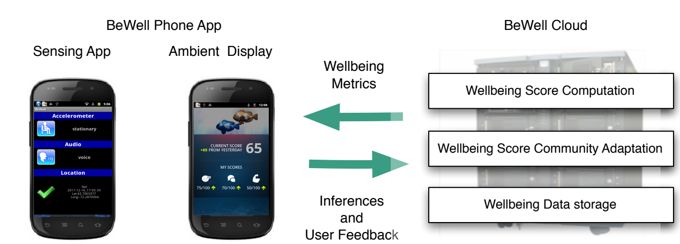

Figure 1: BeWell+ App implementation, including smartphone components supported by a scalable cloud system

Figure 2: Multiple wellbeing dimensions are displayed on the smartphone wallpaper. An animated aquatic ecosystem is shown with two different type of fish whose behavior are affected by changes in wellbeing (i.e., activity and social interaction); in addition the ocean ambient lighting conditions reflect the users sleep duration (shown on Nexus S.)

BeWell+ phone app automatically monitors user's everyday activities using the accelerometer and microphone sensors on phone. Inference results from the classifiers on the phone are then transmitted to the BeWell+ Cloud infrastructure. The Cloud infrastructure stores all the data and computes wellbeing scores. Wellbeing scores summarize the impact on overall health based of the inferred behavioral patterns. BeWell+ computes wellbeing scores for each health dimension it tracks. In the current prototype these are: physical activity, sleep patterns and social interaction as mentioned above. The BeWell+ phone app presents these scores back to users on the phone, using an ambient display rendered on the wallpaper of the device (see Figure 2).

Wellbeing Scores. Wellbeing scores range between 0 and 100 and are calculated for each of the three dimensions (viz. physical activity, social interaction and sleep patterns). A score of 100 indicates the person is matching or exceeding recommended guidelines (e.g., averaging 8 hours sleep per day is represented by a score of 100). Our wellbeing scoring functions are a result of a careful design process which leveraged: the existing literature, guidelines from institutions (e.g., CDC), collaboration with medical researchers and short field experiments. In [23] we provide details of these functions, but briefly summarize them below:

Physical Activity Score: The physical activity of users are classified into walking, stationary, and running classes. Inferences are used to estimate a daily Metabolic Equivalent of Task (MET) value [9]. We rely on the Centers for Disease Control and Prevention's physical activity guidelines [1] to parameterize this function.

Sleep Score: Sleep monitoring is based on the total quantity of the sleep for an individual over a twenty-four hour period. We developed [23] a simple logistic model that estimates the amount of hours slept within ± 1.5 hours based on the phone usage pattern. Our scoring function uses the guidelines for sleep duration provided by the National Sleep Foundation [3].

Social Interaction Score: We detect significant changes in social isolation and social support based on the total duration of ambient speech during a day. This is estimated from the output of a speech/non-speech classifier. We rely on studies that connect social isolation and social support to psychological well-being, with low levels being linked with symptoms such as depression [19]. We experimentally develop a scoring function using small field trials to determine the typical daily quantities of speech encountered by people within our study.

Ambient Display. The ambient display is an animation that is rendered on the phone's lock-screen and wallpaper, making it visible to the user whenever the user glances or interacts with their smartphone. The display provides passive feedback to the user of their current wellbeing scores. Prior examples of successful persuasive systems [10] have found that wallpaper can effectively promote changes in user activity. These studies show phone wallpaper, when used as a glanceable display, can keep user goals "persistently activated" [21] in the mind of the user.

BeWell+ displays multiple wellbeing dimensions as an aquatic ecosystem, as illustrated in Figure 2. The animated activities of a clown fish (which mirrors the user's activity), the ocean ambient lighting conditions (which mirrors the user's sleep duration) and a school of small fish (which mirror the user's level of social interaction) provide a quick summary of the current wellbeing to the user. The relationship between the ambient display and the wellbeing scores is described below:

Clown Fish. The clown fish represents the physical activity of the user. The score modifies the speed at which the clown fish swims. At low levels of physical activity the fish moves slowly from left to right lethargically. As the user's physical activity increases, the fish swims more vigorously, even performing summersaults and backflips at high levels of activity.

School of Blue Fish. A school of fish swims with the clown fish and represents the users social activity. The closeness of the school of fish to the clown fish and its size grows proportionally to the amount of social interaction of the user.

Lighting of Ocean. Sleep patterns are captured by the light of the ocean. The ocean gets darker when the user lacks sleep and has a low sleep score. As the user sleep level increases, the ocean gradually become brighter.

The aquatic ecosystem represents a single point in the design space of the ambient display for BeWell+. Before selecting this visualization we performed small scale informal surveys of people from the target population. We found strong preference from people for the aquatic ecosystem, which is in agreement with examples from the literature where animated animals are effective at motivating behavior change (e.g., [25]).

4. COMMUNITY ADAPTIVE WELLBEING FEEDBACK

In what follows, we describe BeWell+'s data-driven community-adaptive approach to wellbeing feedback. The behavior goals that underpin wellbeing feedback are based on a combination of observations from the user population and ideal "healthy" behavioral patterns. This allows feedback to automatically tune itself to the population in which the system is deployed.

Implications of Community Diversity. Wellbeing problems and solutions can be highly personal. Each individual has their own challenges to wellbeing shaped by factors including, personal characteristics and behavioral tendencies. We first observed this problem even in our initial deployments of the original BeWell system, where we observe large differences even within small groups.

To help quantify this problem further we turn to data from our 27-person field trial (see §6.1 for further details). To measure differences in wellbeing we use the three dimensions of wellbeing scores previously defined in §3 (prior to any adaption). Figure 3 shows the distribution of all wellbeing scores for each user, irrespective of the particular health dimension. Surprisingly, even within a relatively small and homogenous group of people significant diversity is present. From this figure we see that the value and the variance of the wellbeing scores vary significantly across users.

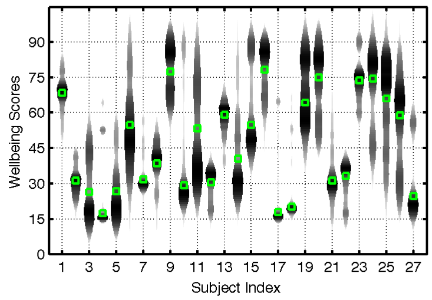

Figure 3: A high diversity of wellbeing behavioral patterns exists among our study population. A score of 100 refers to a "healthy" behavioral pattern.

Although not visible in the figure, we find the user behavior is also diverse within each separate dimension. For example, there are larges difference between the upper and lower quartile wellbeing scores of subjects for each dimension. Specifically, these differences are 61% for physical wellbeing, 83.1% for social wellbeing and 75.5% for sleep dimension.

Feedback from wellbeing apps (along with most mobile health systems) is typically goal based, implying an ideal behavioral pattern to promote. However, high levels of community diversity prevent a single goal behavioral pattern being applied to an entire user population. For example, it is unrealistic to expect an elderly person to meet the same goals for physical activity as a young adult. Similarly, doctors or students often can not conform to a "normal" sleep pattern – but would still benefit from appropriate feedback indicating how their sleeping habits could be made more healthy. Without adjusting the expectations that underpin this feedback it will be ineffective to many users while also damaging the confidence they have in the system. This is the main goal of our new community feedback paradigm – it presents a mechanism to potentially provide effective feedback for very large populations of diverse users in a scalable manner.

Adaptive Wellbeing Feedback. BeWell+ adapts generic wellbeing score functions based on the overall behavior similarity within the user population along with the similarity of users to ideal wellbeing behaviors. This novel process within BeWell+ allows feedback to adapt to the differences between user communities. Without adaptation improvements in behavioral patterns are not considered within the correct context. Another example is as follows: a shift worker who is able to increase her average quantity of sleep from 4 to 5 hours, but this improvement may still score poorly if compared to the general expectation applied to the general public. However, if compared to other shift workers this change could well place the individual in a high performing percentile of that community or peer group. Therefore, wellbeing feedback should recognize this as a substantial positive change, even if the change required to achieve "normal" sleep hygiene remains large.

Adaptation is a data-driven process which relies on activity inferences, along with a trace of periodic GPS estimates, being transferred to the cloud from the BeWell+ app. Figure 4 illustrates each phase of the adaptation process, all computational stages of adaptation are performed by the cloud. Although only a single dimension is shown this process is repeated for all three dimensions. The detailed wellbeing score functions for these three dimensions can be found in [23]. Each function takes a specific statistic related to a user behavioral pattern. Adapted score functions maintain the same functional form, but with parameters being revised to accommodate user diversity. At the conclusion of this process a personalized set of wellbeing score functions are generated for all BeWell+ users.

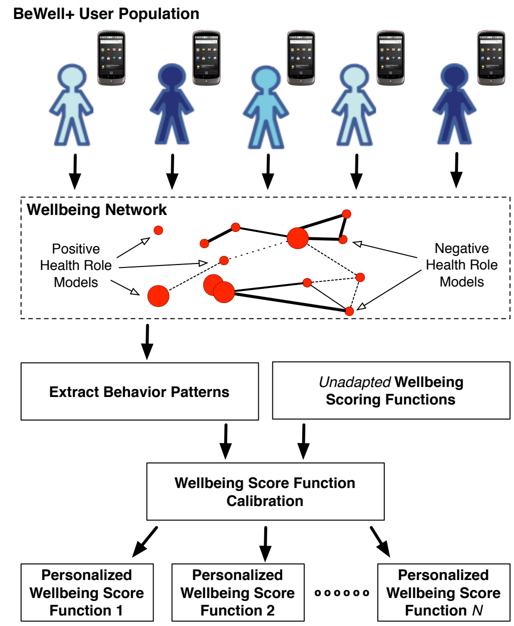

Figure 4: BeWell+ Community Adaptive Wellbeing Feedback

Guiding the adaptation process is a behavioral similarity network, a weighted graph in which nodes correspond to users and edge-weights quantify the level of similarity. This network attempts to identify people with related lifestyles and behavior constraints. BeWell+ computes similarity between two users based on mobility, temporal and activity patterns, adopting the lifestyle similarity definition proposed in [24].

It is critical to keep our personalized wellbeing score functions grounded with respect to "healthy" behavioral norms. Consequently, we use an *unadapted* wellbeing score function to balance the need to identify similar people with the need to recognize which of these people are either positive or negative wellbeing role-models. We begin by using the unadapted wellbeing score function once to score all previously collected user behavior. For every observation of user behavior (e.g., physical activity across a day) a data tuple is formed containing, (1) the wellbeing score, and (2) the relevant statistic concerning the wellbeing behavior (e.g., a daily MET value). Each tuple is weighted based on two factors: first, the similarity network edge-weight; and second, how "healthy" users are compared to an ideal behavioral pattern – determined by their average wellbeing score (using an unadapted scoring function). Tuple weight is a linear combination of these values with a parameter, `sim_strength` that determines how much influence user similarity has over the final weight used.

Finally, the adapted scoring function is generated by applying a weighted smoothing over the collection of tuples and fitting the wellbeing functions to the smoothed tuples. Adapted scoring functions set the underlying goals associated with high scores as realistic near-term objectives, rewarding improvement relative to people within their own community or peer group. Of course over the longer term, the ultimate goal of reaching a more ideal pattern remains important. Our adaptive scoring strategy incorporates this requirement by the process repeating as new data accumulates. Each time the adaptation process repeats it incrementally selects higher performing people as a frame of reference for the user while still emphasizing the need for these people to be relatively similar to the target user, with `sim_strength` controlling this trade-off.

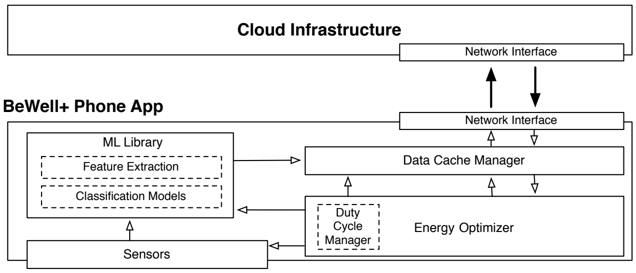

Figure 5: BeWell+ Energy Management Subsystem

5. WELLBEING ADAPTIVE ENERGY ALLOCATION

In this section, we discuss the design of our wellbeing adaptive energy allocation strategy. The novelty of this approach is to prioritize the resource allocation based on how well the user is coping with each individual health dimension. BeWell+ dynamically shifts resources between wellbeing dimensions (viz. physical activity, social interaction and sleep patterns) as the behavior of the user changes – dimensions with low wellbeing scores receiving more resources than those with high scores. As a result, the accuracy and responsiveness of the BeWell+ app are optimized within resource constraints and with an awareness of the user's wellbeing needs.

Insufficient Energy Resources. Monitoring wellbeing requires multiple aspects of daily life to be constantly monitored. This puts undue load on the battery of smartphones as this requires sensing and inference to be performed continuously across a range of sensor modalities. Figure 6 shows the battery life of five subjects using the original BeWell system, as reported in [23]. Even though each Android smartphone is equipped with an large-capacity battery (3200 mAh) battery life varies between 12 and 21 hours. If we assume the use of a factory standard battery (1400 mAh) then these lifetimes will be reduced to between 7 and 10 hours. At this level users will have to recharge their phone multiple times per day, otherwise BeWell will only be able to monitor them for the fraction of the day when the phone is active. This problem is more broadly applicable to the growing number of mobile health applications that consider multiple dimensions of behavior; and even further, is known to impact a variety of mobile sensing applications [31] and smartphone platforms [29].

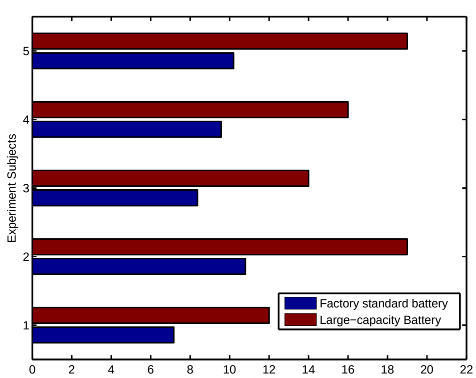

Figure 6: The spread of battery lifetime for the original BeWell system. For each person we show battery life using a large capacity battery (3200 mAh) and an estimate of battery life with a standard factory battery (1400 mAh).

Adaptive Energy Allocation. BeWell+ conserves smartphone energy usage by dynamically tuning the duty cycle of core system components based on the wellbeing score of the user. Figure 5 illustrates the control-loop used by BeWell+ to intelligently allocate the energy consumption, and highlights which component duty cycle parameters are tuned. Specifically, these parameters are: the rate at which sampling, feature extraction and activity inference routines are performed; along with how often BeWell+ interacts with the cloud to either upload user-specific statistics or collect revised wellbeing scores – both of which require community interaction and so necessitate the cloud to be involved.

Our energy management strategy is based on a simple yet effective optimization, which we will now describe:

Let $\texttt{duty}_\texttt{i}$ denote the ith duty cycle parameter in $\texttt{duty.all}$, the set of all duty cycle parameters in the BeWell+ app. Let function $\texttt{acc}_\texttt{j}(\texttt{duty.all})$ estimate the increase in error for specific dimensions of wellbeing scores (indicated by $\texttt{j}$) due to increasing levels of duty cycling. Further, let $\texttt{bat}(\texttt{duty.all})$ estimate the per day smartphone energy consumption due to BeWell+ operation, relative to potential $\texttt{duty}_\texttt{i}$ values. The functions of $\texttt{bat}$ and $\texttt{acc}_\texttt{j}$ use a polynomial regression fitted with data by profiling the BeWell+ app running with different $\texttt{duty}_\texttt{i}$ values, in addition to data from user experiments, which enables accuracy to be assessed. The values for each $\texttt{duty}_\texttt{i}$ parameter is found by optimizing the following objective function:

$$\arg\min_{\texttt{duty.all}} bat(\texttt{duty.all}) + \sum \alpha_j \cdot acc_j(\texttt{duty.all}) \qquad (1)$$

where, α_j is a weighting term allowing the accuracy of certain dimensions of wellbeing scores to be emphasized over others. Specifically, α_j is simply:

$$\frac{1}{\texttt{z}} \cdot (\texttt{score}_\texttt{max} - \texttt{score}_\texttt{actual})/\texttt{score}_\texttt{max} \qquad (2)$$

where $\texttt{score}_\texttt{max}$ is the maximum wellbeing score, $\texttt{score}_\texttt{actual}$ is the present value for the $\texttt{j}$th dimension of wellbeing and $\texttt{z}$ is the term used to normalize weights across all wellbeing dimensions.

The adaptive energy allocation component, shown in Figure 5, performs this optimization each time there is a change in the wellbeing scores. As the wellbeing of the user shifts (e.g., an unhealthy behavior improves significantly), BeWell+ can automatically reallocate energy to provide more accurate monitoring and more responsive feedback for the new wellbeing dimension of highest concern.

6. EVALUATION

In this section, we study the performance of BeWell+ with a 27 person field trial conducted over 19 days. We find that: (1) our community adaptive wellbeing feedback mechanism can reconcile health norms with the practical restrictions that limit near-term user lifestyle changes; (2) users can digest multi-dimensional BeWell+ feedback and are seen to make positive changes in their behavior; (3) wellbeing adaptive energy allocation is able to intelligently allocate resources to underperforming aspects of user wellbeing, while also adjusting to lifestyle changes; and, (4) users report an overall positive experience from the BeWell+ field trial.

6.1 Study Methodology

Our study population contains 16 men and 11 women aged between 21 and 37. Of these subjects, 9% are faculty or graduate students in a computer science department, 34% are doctors or medical researchers and the remaining 57% are students in the arts and life sciences graduate program. Each volunteer agrees to carry a

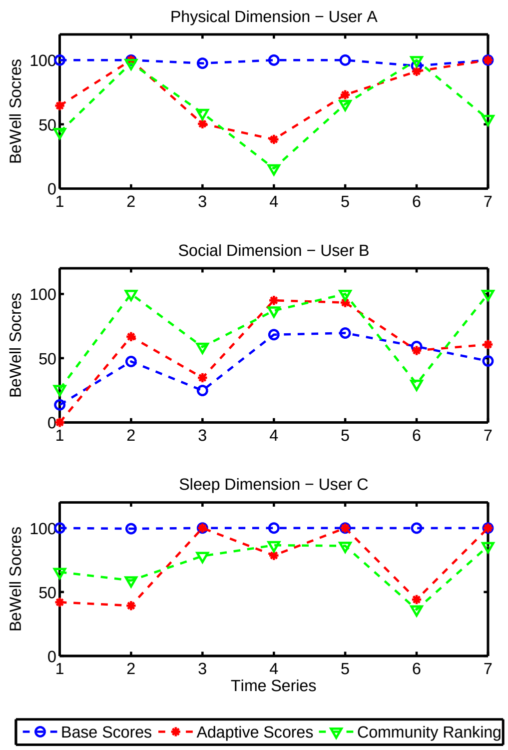

Figure 7: By adapting wellbeing score functions users receive feedback that considers their relative position with peers who have similar lifestyles.

High Performance Group (Sleep)			
	User A	User B	User C
Behavior Change	-4%	5%	10%
Baseline Score	88	90	100
Adaptive Score	36	67	100
Low Performance Group (Social)			
	User A	User B	User C
Behavior Change	-2%	5%	10%
Baseline Score	10	55	58
Adaptive Score	10	77	83

Table 1: Under adaptive wellbeing feedback users continue to receive feedback even if they are well above (or well below) the ideal expectations of healthy behavior.

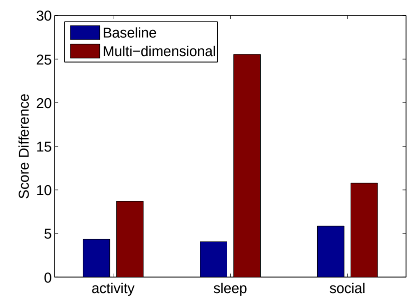

Figure 8: Increases in user wellbeing are largest for those subjects who receive multi-dimensional wellbeing feedback.

phone with the BeWell+ app installed. The subjects either move their mobile phone SIM card into the Nexus One or use call forwarding so they can use the study phone as their primary phone. We provide each user with a holster to clip the phone on to their belt or clothing. Users agree to keep the phone with them at all times.

To verify the effectiveness of presenting multi-dimensional feedback using the ambient display, the participants are randomly and uniformly split into two groups: *multi-dimensional group* and *baseline group*. All subjects have the core BeWell+ software installed that tracks sleep, physical activity and social interaction. However, the baseline group did not have the ambient display and could only view the collected information via a web portal that summarizes the time spent in each activity as a fraction of the day. The multi-dimensional group has the ambient display.

6.2 Community Adaptive Wellbeing Feedback

Our first series of experiments investigate two key aspects of wellbeing feedback: (1) the effectiveness of adapting feedback to keep implied healthy goals within realistic ranges for all users; and, (2) the benefit of multi-dimensional feedback, as observed in the behavioral decisions of our study population.

Adaptive Wellbeing Scoring in Action. To better understand how our adaptive wellbeing feedback can compensate for per user differences (e.g., lifestyles, occupation) we compare the use of adaptive and non-adaptive feedback on representative users from our study. Table 1 shows both forms of wellbeing scoring compared to different behavior changes which occur over the span of two days. In this table we examine two groups selected from the top and bottom 20% of study subjects in sleep and social dimensions respectively. We refer to users in the top 20% as *high performance group*, and those in the bottom 20% as *low performance group*. In Table 1 user A from high performance group in the sleep dimension declines in performance by 4% ($\approx$ 0.3 hours). However, when using an unadaptive wellbeing scoring scheme her score remains high as she continues to far outperform the expectation of this unadaptive scheme. User B has increased performance by 5% ($\approx$ 0.4 hours), so she gets a high score under the baseline scoring. But within high performance group, her performance is in the middle, higher than user A but lower than user C - this fact is only reflected in the adaptive version of the wellbeing feedback. These users only receive personalized feedback when using an adaptive scoring system that understands their performance relative to their peers. Low performance group illustrates an identical scenario. These users from the bottom 20% generally have low scores as they are far behind the performance of the overall user population. But if only compared with their low performance group counterparts, they will have significant changes in the scores, depending on their relative performance inside this group. Finally, Figure 7 presents a time-series view of wellbeing scores (12 days) for three different users from our field trial. For each user we show their performance within a single wellbeing dimension. From this figure one can see that these users hardly receive informative feedback (e.g., their scores remain at 100) without adaptive wellbeing scoring. This is again caused by their behavior exceeding (negatively or positively) the expected norms of the unadaptive wellbeing score system. In Figure 7, we also plot the relative percentile ranking of these users within their own group (the green curve). Clearly, the adaptive scores correspond much more closely to the users' actual peer performance compared to unadaptive scoring.

Multi-dimensional Wellbeing Feedback. We measure the quantitative benefit of providing feedback along multiple dimensions by comparing the changes in wellbeing scores between our two study populations (viz. *multi-dimensional group* and *baseline group*). To compensate for individual variation that could bias results (i.e., participants that have abnormally high or low wellbe-

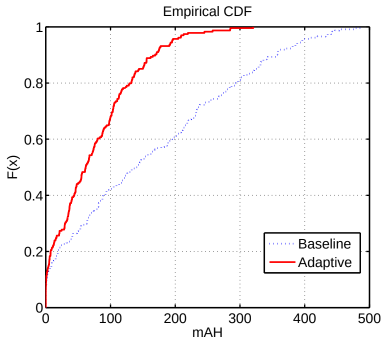

Figure 9: CDF of per user daily energy consumption under Wellbeing Adaptive Energy Allocation compared to a hand-tuned baseline

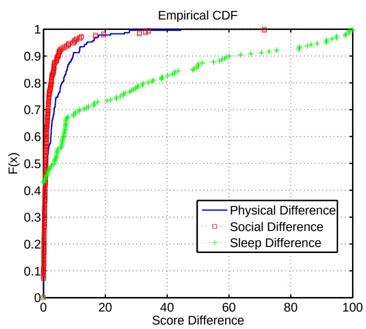

Figure 10: CDF of per wellbeing dimension score error (i.e., score difference) under Wellbeing Adaptive Energy Allocation

ing scores) we compare any changes during the study relative to a baseline average score for each person along each dimension. The baseline score is calculated from data collected during the calibration phase just before the start of the study – none of the subjects had feedback or ambient display during the calibration phase. Figure 8 shows the average difference in the daily score for each person during the study period, relative to their personal baseline. This figure shows a significantly greater increase in score for *multi-dimensional group* compared to *baseline group*. Specifically, this outperformance is 105% for physical activity, 88% for social interaction and 507% for sleep. Two-sample *t*-tests at the 95% significance level indicate that these differences between *multi-dimensional group* and *baseline group* are all statistically significant ($p = 0.049$, $p < 0.01$ and $p = 0.04$ for the physical, social and sleep dimensions respectively).

6.3 Wellbeing Adaptive Energy Allocation

In our next set of experiments, we investigate how efficiently BeWell+ manages smartphone energy, while still closely monitoring user wellbeing.

Energy Efficiency. In this experiment we compare BeWell+'s adaptive resource management to a baseline in which BeWell+ performs no duty cycling. This baseline represents the upper bound accuracy of wellbeing scores with respect to errors that are caused by duty cycling. To compare these two schemes within identical experiment conditions we perform a trace based experiment. We begin by profiling the energy consumption of key energy consuming stages of our BeWell+ prototype when using both the adaptive and baseline approaches. We replay all 19 days of raw data sensor data for each participant, which we collect during our field trial. For each day of each participant we estimate the energy consumed, in addition to computing wellbeing scores.

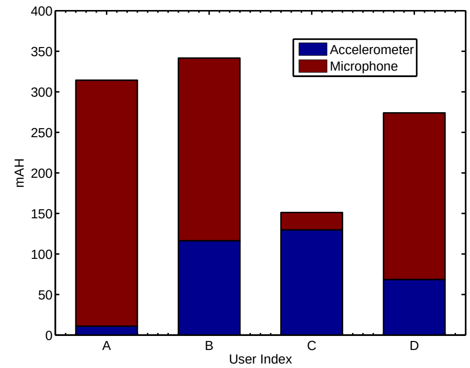

Figure 11: Breakdown of daily energy consumption (by sensor) for four different BeWell+ users

Survey Questions	Answers
1. User would prefer different wallpaper	-1.00
2. Multi-dimensional Display easy to interpret	1.50
3. Multi-dimensional Scores helped keep balance	1.56
*-2: Strongly disagree, -1: Disagree, 0: Neutral, 1: Agree, 2: Strongly agree	
4. I showed others my wallpaper	83.5%
5. Animation was annoying	0.00%
*Percentage of person choose	

Table 2: Ambient Display Results from Exit Interview

Figure 9 shows a CDF of the average energy consumption for each day in this experiment. This figure shows our adaptive scheme is able to reduce average energy consumption by more than 50% for 80% of the days, which is approximately a 3-hour increase in battery life. Reductions in energy consumption should be considered in comparison to Figure 10 which shows the impact to wellbeing score accuracy. For example, lowering energy consumption by 50% results in approximately 18 points of error in the wellbeing score across all three dimensions. We consider this score difference, which corresponds to 5% error in voice fraction measurement or a 1.4 hour error in sleep duration, tolerable given the large increases in energy efficiency that result.

Adaption to User Wellbeing Profile. Figure 11 provides some further insight into the findings of the prior experiment. This figure illustrates the energy consumed for four representative subjects, and the relative allocation of energy to each sensor (and associated computation). For example, user B consumes the most energy as this subject has uniformly poor wellbeing scores across all dimensions, making it difficult to conserve energy from any one dimension. As expected in this case the allocation of energy between dimensions is evenly split. In contrast, user C uses significantly less energy as she has comparatively high wellbeing scores, allowing the adaptive scheme to lower energy used for these dimensions. The reason why the accelerometer is allocated a larger proportion of the energy budget for user C is that it is still the weakest dimension (in comparison to other dimensions.)

6.4 Exit Interview

In the remainder of this section we explore user reactions to: (1) the multi-dimensional ambient display; and, (2) subject attitudes and preferences to general usage of the BeWell+ app.

Reactions to Ambient Display. Table 2 summarizes exit interview questions related to the ambient display. Participant responses indicate they have a positive reaction to the phone wallpaper as a means to visualize multi-dimensional wellbeing scores. A natural concern is that the use of multiple dimensions will overwhelm the user and they will not be able to easily digest the information. However, for example, question 2 in Table 2 shows that people overall had little difficulty in interpreting the ambient display.

During exit interviews we discover friends and co-workers often casually ask *how is your fish today?* Many of the participants mention that they compare scores with other participants; 83.5% of *multi-dimensional group* report that they show the display to their friends and colleagues. Exit surveys highlight an unexpected amount of social activity attributable to the ambient display in only a few weeks. Still, this enthusiasm may be due to a potentially short-lived novelty effect among subjects, this observation requires further testing as part of a long-term followup study.

From Table 2 we find very few subjects prefer an alternative wallpaper – we believe this number may rise when deployed in a broader population. During discussion we find that participants commonly turn off the phone screen when in more formal settings (e.g., meetings or while giving presentations) because of concerns it may be mistaken for a game or lead to them not being taken seriously by their peers. The ability to temporarily hide the display seems to be a necessary feature. Still, none of the subjects describe the visualization or the frequent animation as annoying (see question 4 in Table 2).

BeWell+ Application Experience. We find 70% of subjects believe that BeWell+ is a helpful and enjoyable application. A common theme with subjects is that they are surprised by what they learn from the study about their lifestyles. They report they find themselves motivated to actively change their daily behavior.

Encouraged by some early interview responses we decide to investigate some of the reasons for improved behavioral patterns during the study. We are curious if such increases are partially due to an improved ability within *multi-dimensional group* to connect everyday actions to wellbeing outcomes. To test this we perform a simple recall test. We show a timeline of participant wellbeing scores along different dimensions (viz. sleep, activity, social) and ask the participant to annotate and explain the variations seen in the timeline. Our findings show that the subjects that have access to multi-dimensional feedback on the phone are better able to connect life events to fluctuations in wellbeing. On average *multi-dimensional group* recalls 4.28 events per week compared to just 1.8 events for *baseline group*. Similarly, *multi-dimensional group* is able to recall a larger number of unique events as well. Common annotated events included: friends visiting for the weekend, change of (hospital) rotation, or pressure from work.

7. CONCLUSION

In this paper, we presented the next generation of the BeWell app – *BeWell+*, a smartphone application for monitoring and providing feedback across multiple dimensions of wellbeing. The primary goal of our field trial was to deploy BeWell+ to mainstream users in a real-world setting. Our deployment allowed us to both investigate fundamental issues that may influence the design of future generations of wellbeing apps and validate some of the assumptions that underpin BeWell+. Due to the relative short duration of this study it is not possible to make any claims of long-term behavioral change. The behavioral changes we do observe we believe are positive indications of the ability of BeWell+ to convey information and increase awareness. A longer-term field study and a more diverse population of users are both desired to further study BeWell+ and in particular the novel wellbeing mechanisms it introduces, namely, community adaptive feedback and wellbeing adaptive energy allocation.

8. REFERENCES

[1] Center for Disease Control and Prevention. `http://www.cdc.gov`.

[2] Fitbit. `http://www.fitbit.com`.

[3] National Sleep Foundation. `http://www.sleepfoundation.org`.

[4] bLife. `http://www.myblife.com`.

[5] Nike+. `http://nikerunning.nike.com/nikeos/p/nikeplus`.

[6] Philips Directlife. `http://www.directlife.philips.com`.

[7] Shuteye, `http://dub.washington.edu/projects/shuteye`.

[8] A. Ahtinen, E. Mattila, A. Vaatanen, L. Hynninen, J. Salminen, E. Koskinen, K. Laine. User Experiences of Mobile Wellness Applications in Health Promotion: User Study of Wellness Diary, Mobile Coach and SelfRelax. In *PervasiveHealth '09*.

[9] B. Ainsworth, W. Haskell, M. Whitt, M. Irwin, A. Swartz, S. Strath, W. O'Brien, D. Bassett, K. Schmitz, P. Emplaincourt, D. Jacobs, A. Leon. Compendium of Physical Activities: An Update of Activity Codes and MET Intensitiess. *Medicine and science in sports and exercise*, 32(9 Suppl), Sept 2000.

[10] S. Consolvo, P. Klasnja, D. W. McDonald, D. Avrahami, J. Froehlich, L. LeGrand, R. Libby, K. Mosher, J. A. Landay. Flowers or a Robot Army?: Encouraging Awareness and Activity with Personal, Mobile Displays. In *UbiComp '08*.

[11] S. Consolvo, P. Klasnja, D. McDonald, J. Landay Goal-setting considerations for persuasive technologies that encourage physical activity In *Persuasive '09*.

[12] S. Halko, J. Kientz Personality and persuasive technology: an exploratory study on health-promoting mobile applications In *Persuasive '10*.

[13] T. Denning, A. Andrew, R. Chaudhri, C. Hartung, J. Lester, G. Borriello, G. Duncan. BALANCE: Towards a Usable Pervasive Wellness Application with Accurate Activity Inference. In *HotMobile '09*.

[14] R. Balan, M. Satyanarayanan, S. Park, and T. Okoshi. Tactics-based remote execution for mobile computing. In *Mobisys '03*.

[15] J. Hicks, N. Ramanathan, D. Kim, M. Monibi, J. Selsky, M. Hansen, D. Estrin. Andwellness: an open mobile system for activity and experience sampling. In *Wireless Health '10*.

[16] P. Ferreira, P. Sanches, K. Höök, T. Jaensson. License to Chill!: How to Empower Users to Cope with Stress. In *NordiCHI '08*.

[17] B. J. Fogg. Persuasive Technology: Using Computers to Change What We Think and Do. *Ubiquity*, 2002(December):2, 2002.

[18] K. Fox. The Influence of Physical Activity on Mental Well-being. *Public Health Nutrition*, 2(3a):411–418, 1999.

[19] L. George, D. Blazer, D. Hughes, N. Fowler. Social Support and the Outcome of Major Depression. *The British Journal of Psychiatry*, 154(4):478, 1989.

[20] D. Hareva, H. Okada, T. Kitawaki, H. Oka. Supportive Intervention using a Mobile Phone in Behavior Modification. *Acta Med Okayama*, 63(2):113–20, 2009.

[21] P. Klasnja, S. Consolvo, D. McDonald, J. Landay, W. Pratt. Using Mobile and Personal Sensing Technologies to Support Health Behavior Change in Everyday Life: Lessons Learned. In *AMIA Annual Symposium Proceedings*, vol. 2009, pp. 338–342.

[22] E. Lamminmaki, J. Parkka, J. Hermersdorf, J. Kaasinen, K. Samposalo, J. Vainio, J. Kolari, M. Kulju, R. Lappalainen, I. Korhonen. Wellness Diary for Mobile Phones. In *EMBEC '05*.

[23] N. D. Lane, M. Mohammod, M. Lin, X. Yang, H. Lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, A. Campbell. BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing. In *PervasiveHealth '11*.

[24] N. D. Lane, Y. Xu, H. Lu, S. Hu, T. Choudhury, A. T. Campbell, F. Zhao. Enabling large-scale human activity inference on smartphones using community similarity networks (CSN). In *UbiComp '11*.

[25] J. Lin, L. Mamykina, S. Lindtner, G. Delajoux, H. Strub. Fish'n'steps: Encouraging Physical Activity with an Interactive Computer Game. In *UbiComp '06*.

[26] R. LiKamWa, Y. Liu, N. Lane, L. Zhong Can Your Smartphone Infer Your Mood? In *PhoneSense'11*.

[27] A. Ali, S. Hossain, K. Hovsepian, M. Rahman, K. Plarre, S. Kumar. mPuff: automated detection of cigarette smoking puffs from respiration measurements. In *IPSN'12*.

[28] R. Norris, D. Carroll, R. Cochrane. The Effects of Physical Activity and Exercise Training on Psychological Stress and Well-being in an Adolescent Population. *Journal of Psychosomatic Research*, 36(1):55–65, 1992.

[29] E. Miluzzo, J. Oakley, H. Lu, N. Lane, R. Peterson, A. Campbell Evaluating the iPhone as a mobile platform for people-centric sensing applications In *UrbanSense'08*.

[30] K. Patrick, F. Raab, M. Adams, L. Dillon, M. Zabinski, C. Rock, W. Griswold, G. Norman. A Text Message–based Intervention for Weight Loss: Randomized Controlled Trial. *Journal of Medical Internet Research*, 11(1), 2009.

[31] Y. Wang, J. Lin, M. Annavaram, Q. Jacobson, J. Hong, B. Krishnamachari, N. Sadeh. A framework of energy efficient mobile sensing for automatic user state recognition. In *Mobisys '09*.

[32] H. Yaggi, A. Araujo, J. McKinlay. Sleep Duration as a Risk Factor for the Development of Type 2 Diabetes. *Diabetes Care*, 29(3):657, 2006.