

# **Designing for Healthy Lifestyles: Design Considerations for Mobile Technologies to Encourage Consumer Health and Wellness**

By Sunny Consolvo, Predrag Klasnja,  
David W. McDonald and James A. Landay

## **Contents**

---

<b>1</b>	<b>Introduction</b>	<b>169</b>
1.1	Our Mobile Technologies to Encourage Physical Activity	171
1.2	Roadmap	181
<b>2</b>	<b>Collecting Behavioral Data</b>	<b>183</b>
2.1	Tracking Food Intake	184
2.2	Tracking Physical Activity	205
2.3	Broader Considerations about Collecting Behavioral Data	217
2.4	Open Questions for Collecting Behavioral Data	222
2.5	Section 2 Wrap-Up	225
<b>3</b>	<b>Providing Self-Monitoring Feedback</b>	<b>226</b>
3.1	Forms of Feedback	227

3.2	Location of Feedback	241
3.3	Open Questions for Providing Self-Monitoring Feedback	250
3.4	Section 3 Wrap-Up	253
<b>4</b>	<b>Supporting Goal-Setting</b>	<b>254</b>
4.1	Goal-Setting in HCI Research	258
4.2	Our Experiences with Goal-Setting	260
4.3	Open Questions for Supporting Goal-Setting	273
4.4	Section 4 Wrap-Up	282
<b>5</b>	<b>Moving Forward</b>	<b>284</b>
5.1	Assessing Starting Level and Progress	284
5.2	Supporting the User When “Stuff” Happens	298
5.3	Supporting the User Over Her Lifespan	299
5.4	Section 5 Wrap-Up	304
5.5	Final Thoughts	305
	<b>Acknowledgments</b>	<b>306</b>
	<b>References</b>	<b>307</b>

# Designing for Healthy Lifestyles: Design Considerations for Mobile Technologies to Encourage Consumer Health and Wellness\*

Sunny Consolvo<sup>1</sup>, Predrag Klasnja<sup>2</sup>,  
David W. McDonald<sup>3</sup> and James A. Landay<sup>4</sup>

<sup>1</sup> *Google & University of Washington, USA, [sconsolvo@google.com](mailto:sconsolvo@google.com)*

<sup>2</sup> *University of Michigan, USA, [klasnja@umich.edu](mailto:klasnja@umich.edu)*

<sup>3</sup> *University of Washington, USA, [dwmc@uw.edu](mailto:dwmc@uw.edu)*

<sup>4</sup> *Cornell University, USA, [landay@cornell.edu](mailto:landay@cornell.edu)*

## Abstract

As the rates of lifestyle diseases such as obesity, diabetes, and heart disease continue to rise, the development of effective tools that can help people adopt and sustain healthier habits is becoming ever more important. Mobile computing holds great promise for providing effective support for helping people manage their health in everyday life. Yet, for this promise to be realized, mobile wellness systems need to be well designed, not only in terms of how they implement specific behavior-change techniques but also, among other factors, in terms of how much burden they put on the user, how well they integrate into

---

\* Dr. Consolvo did this work while at the University of Washington and Intel Labs Seattle.

the user's daily life, and how they address the user's privacy concerns. Designing for all of these constraints is difficult, and it is often not clear what tradeoffs particular design decisions have on how a wellness application is experienced and used. In this monograph, we provide an account of different design approaches to common features of mobile wellness applications and we discuss the tradeoffs inherent in those approaches. We also outline the key challenges that HCI researchers and designers will need to address to move the state of the art for mobile wellness technologies forward.

# 1

---

## Introduction

---

The world is facing a health crisis. Physical inactivity, poor diet, and other lifestyle behaviors (e.g., stress and insufficient sleep) are contributing to an epidemic of chronic conditions, including obesity, diabetes, and cardiovascular disease [48, 75]. These conditions now account for over two-thirds of U.S. healthcare expenditures [41], and their cost, in terms of economic impact and human suffering, is continuing to rise both in the United States and in other parts of the world. With an aging population further contributing to the rapidly rising health care costs, health leaders are encouraging people to take more responsibility for their own health behaviors. However, many of us are well aware of how difficult it can be to change our behaviors. As anyone who has ever made a New Year's resolution to get in shape or follow a healthy diet knows, changing one's habits is notoriously difficult. Too many of us end up making the same resolution year in and year out, only to fall back into our old habits after several weeks. The reasons may vary but the end results are often the same: little or no change for the better.

Mobile computing holds great promise for providing effective support for managing health in everyday life. Mobile devices include powerful processors, sensing capabilities, high-resolution display

screens, nearly pervasive connectivity, and they go with us everywhere we go. In June 2013, the Pew Research Center reported that more than 90% of American adults own a mobile phone, and more than 50% of American adults own a smartphone [78, 84]. Mobile computing represents a fundamental change in how wellness can be tracked and managed.

The promise of mobile computing has not gone unnoticed. Numerous commercial products have launched, and a growing number of research projects have been reported in the literature. Progress continues to be made in areas from innovations in sensing to new designs of mobile interfaces and techniques for engaging people in the process of managing their health. A number of recent survey articles have focused on a range of issues in mobile health and wellness. Tentori, Hayes, and Reddy [86] review mobile systems that address mobile clinical and end-user health and wellness applications. Tentori et al. focus on the diversity of systems and how each one addresses a specific health challenge. Klasnja and Pratt [53] categorize health interventions that have been developed for mobile phones, and discuss the features of modern smartphones that enable each type of intervention. Another review by Cowan et al. [24] focuses on the types of behavioral theories that are incorporated into mobile health applications that support behavior change. Cowan et al. found that mobile behavior-change applications use only subsets of a few well-established theories. Finally, in health sciences there have been a number of reviews of the use of text messaging (SMS) for supporting health behavior change (e.g., [31]).

While these recent surveys have addressed specific advantages of mobile computing, the types of health applications that can leverage mobile technology, and how those applications incorporate behavioral theory, there has not yet been a review of the design features of those applications and the design challenges and opportunities for mobile health and wellness technologies. Our focus here is to consider some of what has been learned about the design of such technologies and to articulate a set of design challenges that must be overcome for designing effective mobile health and wellness technologies.

The reasons for this focus are twofold. First, data suggest that design problems with current applications are adversely affecting

people's ability to use — and thus benefit from — mobile health and wellness applications. While the interest in mobile health applications is rising — for example, Pew recently found that nearly a fifth (19%) of smartphone owners have downloaded at least one health application [32] — continued active use of these applications is very low. A recent survey by the Consumer Health Information Corporation [23] found that 26% of downloaded health applications are used only once, and 74% are abandoned by the 10th use. Usability and design were found to be key considerations related to continued use. Improving the design of mobile health applications is thus critical if the potential of these technologies to help people reach their health goals is to be realized.

Second, while much of what we already know about the effective design of technologies in general will apply in this space, mobile health and wellness technologies have new or additional requirements that take center stage, such as the need to impact deeply-ingrained habits like daily food choices. In addition, these technologies raise a number of evaluation challenges, as those of us coming from an HCI and design background begin to develop systems that must satisfy not only end-users but also researchers and practitioners from the health sciences and related communities. For these reasons, a review of the design aspects of mobile health and wellness technologies seems to be in order. In this monograph, we attempt to provide such a review. Using our own research as examples throughout the review, along with other research and commercial health applications, we provide an account of different design approaches to common features of mobile health and wellness technologies and discuss the tradeoffs inherent in those approaches. We also outline the key challenges that HCI researchers will need to address to move the state of the art for mobile health and wellness technologies forward.

## **1.1 Our Mobile Technologies to Encourage Physical Activity**

Much of the discussion that follows uses our own mobile health and wellness applications as examples to illustrate the issues we discuss. We have been working in the space of mobile technology to encourage

health and wellness for many years. In general, our work has focused on people who are motivated to make healthy changes in their everyday lives (e.g., be more physically active and get better sleep), have the ability and desire to do so, but have not yet done so, or at least not done so consistently. That is, our work tends to target people who are in the *contemplation*, *preparation*, and *action* stages of change as defined by the *Transtheoretical Model of Behavior Change* [74]. Most of our work has focused on encouraging people to be physically active, though we have done some work on encouraging healthy eating [unpublished] and sleep habits as well (e.g., [7]).

In this section, we describe key aspects of three of our mobile health projects that attempt to encourage people to engage in physical activity. We cover *Houston* [19], a system to encourage people to take more steps, as well as *UbiFit* [20] and *GoalPost* [66], systems designed to help people incorporate regular and varied physical activity into their everyday lives. These technologies were pilot tested from weeks to months by members of the research team (and sometimes our colleagues and family members) prior to the field studies and deployments with target end-users that are mentioned.

### 1.1.1 Houston

In our first investigation, we were interested in encouraging *opportunistic physical activities*. That is, we were attempting to help people incorporate simple activities into their everyday lives such as taking the stairs instead of the elevator, or parking further away from their destinations. We were inspired by studies that found that people can achieve health benefits by merely increasing the number of steps they take each day and that social support from friends and family showed an increase in physical activity [15, 16, 89, 95]. With this in mind, we developed an application called *Houston* that encouraged small groups of friends to share their step counts and performance toward a daily step count goal via their mobile phones [19]. *Houston* was designed to promote self-reflection by providing personal awareness of daily step count through a mobile journal, goal-setting by providing progress toward and rewards for achieving a daily step count goal, and social influence by mediating physical activity-related social interaction among friends.



The *Houston* application was developed for the Nokia 6600 mobile phone, and the user's step count was detected by a commercially available pedometer (we used the Omron HJ-112 in our study). The user would read her step count from the pedometer, then enter it into the *Houston* application on the phone. She could enter her current count as often as she liked throughout the day, and she indicated when she was entering her final count for the day. She could enter her step count for today and yesterday, but no further back than that. If she had not reached her goal when entering her current step count, a pop-up message told her how many steps she still had to go (e.g., "<number of> steps to goal"). If she had not entered her final step count into *Houston* by the end of the day, she received a reminder on the phone to do so (and again the next morning if she hadn't entered yesterday's final count<sup>1</sup>). *Houston* provided positive messages when the user reached her daily step count goal (e.g., pop-up screens that read "Congratulations, you have reached your goal!" and "<number of> steps over your goal"), as well as a symbol next to her step count (i.e., an '\*') to indicate that her goal was met. Within the *Houston* application, the user could also choose to share her current step count with the members of her group, add notes to her step counts, send messages to the members of her group, and review trending information about her daily step counts and those of the members of her group, provided that they chose to share. She could also receive messages and step counts from the members of her group (see Figure 1.1).

We conducted a three-week field study of the *Houston* application ( $N = 13$ ) in Summer 2005 with three groups of women who were aged 28–42; each group's members were from pre-existing social networks. All participants regularly used mobile phones and wanted to increase their level of physical activity. During the study, participants carried a study-provided phone dedicated to *Houston's* use, in addition to their personal mobile phone.

We built three versions of *Houston* for the study: baseline, personal, and sharing. During the first week, all three groups of participants

---

<sup>1</sup>The Omron HJ-112 pedometer that we used in our study supported viewing the user's last seven days of step counts; the pedometer automatically reset itself to 0 steps at 12:00 am.

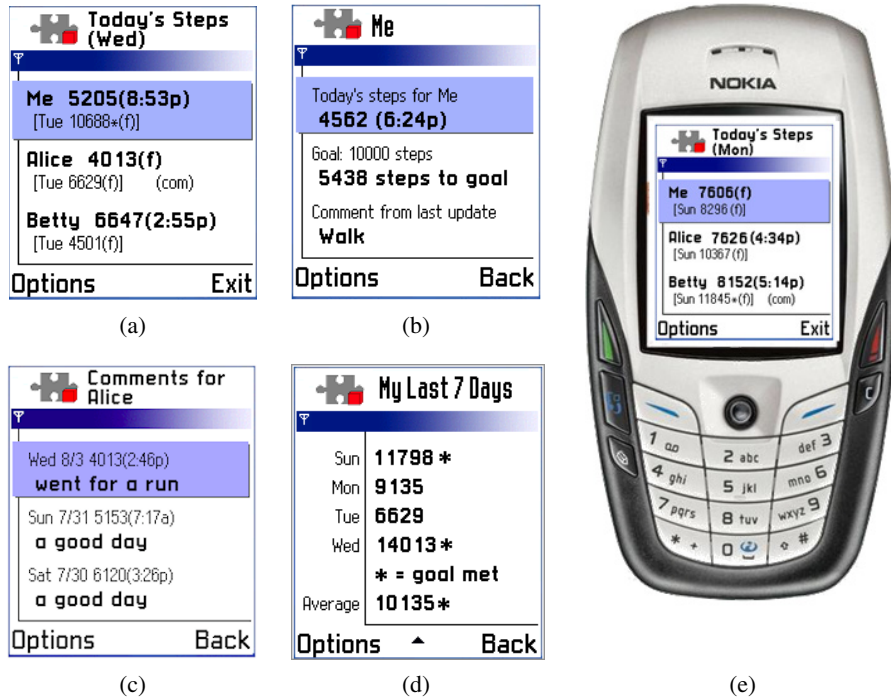


Fig. 1.1 **An overview of Houston.** (a) The main screen showing the user's daily step count for today and yesterday and the same information for members of her group; the "(f)" indicates the final count for the day, and the "(com)" indicates that the count includes a comment; (b) a daily detail screen showing progress toward goal; (c) comments that a member of the user's group, Alice, added to recent days; (d) step count totals for the user's last seven days, including a "\*" to indicate days when the daily goal was met, and (e) *Houston* running on a Nokia 6600 mobile phone.

used the *baseline* version, which was used to establish individual daily step count goals and familiarize participants with *Houston's* interaction model. With the *baseline* version of *Houston*, participants could: enter or edit a step count for today at any time during the day, as often as they wanted; enter or edit a final count for yesterday (e.g., if they did not enter a final count the previous day); and view final daily step counts for the last 7 days. For the remaining two weeks of the study, one of the groups used the *personal* version of *Houston*, while the other two used the *sharing* version. The *personal* version of *Houston* had all of the features of the baseline, and also provided a daily goal, progress

toward and recognition for meeting the goal, a daily step count average, and support for adding comments. The *sharing* version had all of the features of the *personal* version as well as additional features to support sharing of physical activity-related information with the other members of the user's group — that is, her “fitness buddies” — through the *Houston* application.

Additional details are described in [19]. Select findings from the three-week field study of *Houston* and how they relate to design are discussed throughout this monograph.

### 1.1.2 UbiFit

In our second investigation into developing technology to support health and wellness, we continued with the idea of using an application on a mobile phone accompanied by sensing and inference to detect activity. However, we changed our focus from encouraging an increase in daily step count to encouraging people to incorporate regular and varied physical activity into their everyday lives. We also took a step back from incorporating social influence into the system and decided instead to focus solely on the individual. *UbiFit* was designed to promote self-reflection by providing personal awareness of all of the physical activities that the user performs over the course of a week and goal-setting by providing progress toward and rewards for achieving a weekly physical activity goal.

The *UbiFit* application was developed for the Windows Mobile Smartphone, and the user's physical activities were automatically detected by the *Mobile Sensing Platform (MSP)* [17] and manually journaled by the user. The *UbiFit* system consisted of three main components: a glanceable display, an interactive application, and a fitness device (i.e., the MSP). The *glanceable display* used a non-literal but understandable and aesthetically pleasing image that represented key information about the user's physical activity behavior and goal attainment that was available essentially whenever and wherever she was because the display resided on the background screen (or “wallpaper”) of her cell phone. For the purposes of our study, we implemented the glanceable display as a garden that bloomed throughout the week as the

user performed physical activities. Different types of flowers represented different types of activities: cardiovascular activity, strength training, flexibility training, and walking. Upon meeting her weekly goal, a large butterfly appeared near the upper right corner of her display. Smaller butterflies represented goals attained in recent weeks, serving to reward and remind the user of recent successes. Yellow butterflies represented when the user met her *primary weekly goal*. White butterflies represented when the user met her *alternate weekly goal* — an optional goal that was intended to be less challenging to help the user through difficult periods (such as a busy period at work or a mild illness) in hopes that she would not give up for the week if her primary goal seemed out of reach. At the end of each week, the garden reset. It showed one calendar week's worth of activities (Sunday through Saturday) and four week's worth of goal attainments at a time.

The *interactive application* included detailed information about the user's physical activities and a journal where she could manually add, edit, and delete information about her activities. She could also see her weekly goal and the progress that she was making toward her goal. For the purposes of our field studies (described below), the user had to work with a study researcher to make any changes to her weekly goal; the application did not provide a way for the user to change the goal for herself. The *fitness device* automatically inferred and communicated information about certain types of physical activities (e.g., walking, running, cycling, using the elliptical trainer, and using the stair machine) to the *UbiFit* application on the phone. As with *Houston*, the user could add, edit, or delete activities for today and yesterday, and if nothing had been manually journaled for about two days, a reminder prompt asked if the user had anything to add (see Figure 1.2).

We used an iterative design process to develop *UbiFit*. This process included a paper-based survey, a 3-week field study, and a 3-month field study. The *survey* included a mix of multiple choice and open-ended questions about respondents' use of cell phones, their physical activity goals and practices, and two proposed designs — one of which was an early version of the garden design. Seventy-five people (46 female) who ranged from 18 to 63 years old and lived in 13 states across the United States responded. In the *three-week field study*, which was conducted in



Fig. 1.2 **An overview of *UbiFit*** (a)–(e) show the glanceable display’s garden. In (a), the user has not performed any activities yet this week, and she did not meet her goal in any of the prior three weeks. In (b), the user has not performed any activities yet this week, but the three small butterflies indicate that she met her goal in each of the three prior weeks (yellow = primary goal, white = backup goal). In (c), the user has performed one cardio activity so far this week and met her goal last week and three weeks ago. In (d), the user has had an active week, but only performed cardio and walking activities. In (e), the user has had an active week full of variety. In (f), the user is looking at a daily view within the interactive application where her activities are broken down by category. In (g), the fitness device — i.e., the MSP — is shown, and (h) shows the garden as seen on the background screen of a Cingular 2125 Windows Mobile Smartphone.

Summer 2007, 12 participants (six female) who were recruited from the general public used the full *UbiFit* system for 21 to 25 days. Participants were from 25 to 35 years old, lived in the Seattle Metropolitan area, and were regular cell phone users who wanted to increase their physical activity. In the *three-month field study*, 28 participants (15 female) who were recruited from the general public used one of three versions of the *UbiFit* system for three months over the winter holiday season (from November 2007 to February 2008<sup>2</sup>). The three versions were: (a) *full system*, which included all three main components, (b) *no garden*, which included the interactive application and fitness device, but no glanceable display (i.e., there was nothing special about the phone's background screen, nor was there an aesthetic representation of activity), and (c) *no fitness device*, which included the glanceable display and interactive application, but no fitness device (i.e., all activities had to be manually journaled by the user). Participants were aged 25 to 54, lived in the Seattle Metropolitan area, and were regular cell phone users who wanted to increase their physical activity. During both field studies, participants carried a study-provided phone as their personal cell phone (i.e., their personal SIM card was put into a study phone, contacts were transferred over, and participants used the study phone as their personal phone for the duration of the study). Improvements were made to the system after each study. For example, the backup goal was added to the system for the three-month field study based on feedback we received in the three-week field study.

Additional details, including how theories from behavioral and social psychology influenced the design of *UbiFit*, are described in [20, 21, 22]. Select findings from the studies of *UbiFit* and how they relate to design are discussed throughout this monograph.

### 1.1.3 GoalPost

To further investigate some of the strategies that we used in our prior work to encourage regular and varied physical activity, we developed

---

<sup>2</sup>To put the timing of this work in the context of 3rd party development of smartphone applications, Apple released the original iPhone in June 2007 and launched the iPhone software developer's kit, which enabled 3rd party developers to develop applications for the iPhone, in March 2008. The first Android phone was sold in October 2008.

another mobile-phone application, *GoalPost*. Unlike *UbiFit* and *Houston*, with *GoalPost*, we focused solely on the mobile-phone application; we did not use any type of sensing or inference to detect the user's physical activities. *GoalPost* was designed to support goal-setting by encouraging users to set two goals per week — a primary goal and a secondary goal; rewards by giving users ribbons and trophies as they made progress toward and achieved their weekly goals; self-monitoring via an activity journal that used two styles of reminders to encourage users to record their activities and set their goals, and sharing via a feature that enabled users to easily share their goals, activities performed, and goals achieved with members of their Facebook network.

The *GoalPost* application was developed for the Apple iPhone. All physical activities were manually journaled by the user. As in *UbiFit*, *GoalPost* users set goals for a calendar week (Sunday through Saturday) that were broken down by category — cardio, strength, flexibility, walking, and other. Also as in *UbiFit*, goals could be specified at the category and/or specific activity level (i.e., 90 min of cardio OR 30 min of running and 60 min of elliptical) and could include any or all of the categories. Unlike *UbiFit*, *GoalPost* users could set and change their own goals whenever they wanted from within the application with no involvement from the researchers. When setting their goals, users could pick from a list of predefined activities or create their own. Similar to *UbiFit*, *GoalPost* users were encouraged to set two goals per week — one *Primary* and one *Secondary*. Users could choose whether or not they wanted to set both goals, and they chose how those goals were used (e.g., as a main and a backup in case the main became too challenging, or a main and a stretch to give them something extra for which to strive). Users were responsible for recording their physical activity in *GoalPost*, and they could record any physical activity, whether or not it counted toward a goal.

*GoalPost* provided users with pop-up reminders on their phone to journal physical activities and set goals, as well as a persistent reminder (in the form of a “notification badge”) on the application's icon of how many days since she performed a physical activity. Users earned trophies and ribbons as a reward for completing goals and activity categories within the goals. A ribbon was awarded for each category — cardio, strength, flexibility, walking, and other — within



the goal that they completed (blue for categories in their primary goal, red for secondary). A trophy (gold for primary, silver for secondary) was awarded when they completed all elements of their goal. Users were also able to post physical activity-related updates to their Facebook NewsFeed from within the *GoalPost* application. The user could choose to share her activity journal for a day or week, a single activity, a goal(s), progress toward the goal(s), her trophy case, or nothing. If she chose to share, she specified if the update should be shared with a subset of her Facebook network or her entire network (see Figure 1.3).

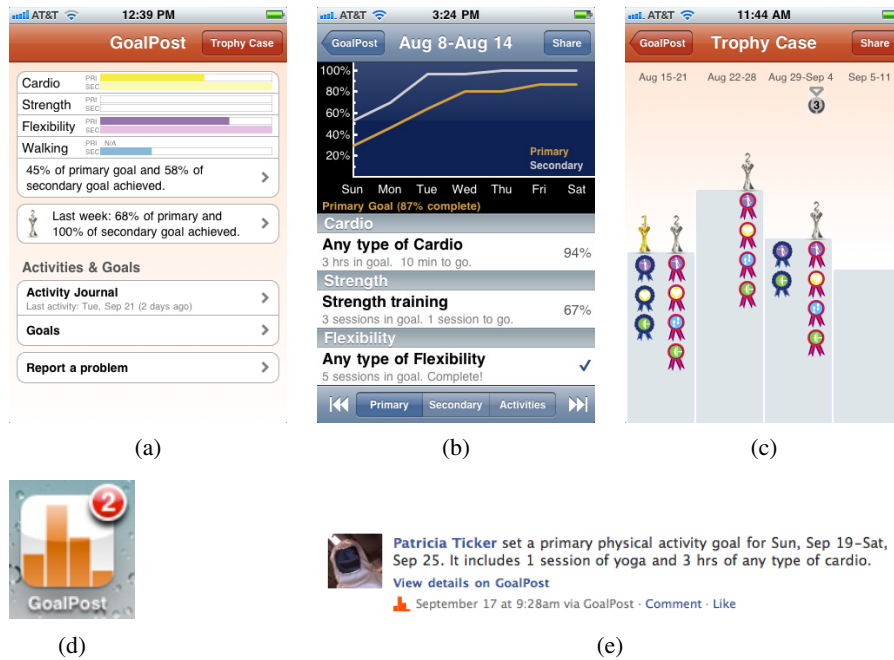


Fig. 1.3 **An overview of *GoalPost*.** (a) *GoalPost*'s main screen shows progress bars for each activity category of the user's goals as well as a percentage of how much of her goals have been achieved and how she did with respect to her goals last week. In (b), the *Goal* screen shows how the user is doing with respect to her goals this week, both in graph and text form; the user can navigate to the same view for prior weeks. In (c), the user's trophy case is shown; ribbons are for completed categories within a goal (e.g., cardio) and trophies are for achieving the entire goal. In the example, the "3" medal under the date range for Aug 29–Sep 4 shows that the user has met her secondary goal for three straight weeks. In (d), the reminder badge on *GoalPost*'s icon is shown; in the example, the user has not journaled any activities for two days. In (e), example user "Patricia Ticker" shares a goal with her Facebook network.



To help design the *GoalPost* application, we conducted a survey using a convenience sample ( $N = 55$ ) of our friends, family, and colleagues. In the survey, we solicited feedback on configuring goals, providing rewards, and default content for the Facebook NewsFeed updates that could be shared from *GoalPost*. Once the application was built, we conducted a four-week long field study of *GoalPost* in September and October 2010 with 23 participants in the Seattle Metropolitan area who were between the ages of 20 and 50. Participants were recruited from the general public and wanted to increase their physical activity. They also owned an iPhone 3G or more recent version and were willing to download the study application onto their personal phone for the duration of the study.

We built two versions of the *GoalPost* application for the study: *GoalPost* and a subset of *GoalPost* called *GoalLine*. *GoalPost* was the full application as described above. *GoalLine* was just like *GoalPost* except that it did not include the sharing features (i.e., if a participant wanted to post something about her goals or activities to her Facebook network, *GoalLine* did not include any features to facilitate that). Twelve participants used *GoalPost* for the duration of the study, while the other 11 used *GoalLine*.

Additional details are described in [66]. Select findings from the studies of *GoalPost* and how they relate to design of mobile health and wellness applications are discussed throughout this monograph.

## 1.2 Roadmap

In what follows, we discuss design aspects of the key features of mobile health and wellness technologies that people can use to adopt and sustain a healthier lifestyle. In our discussion, we use examples from our work as well as the work of other commercial products and research projects around mobile health and wellness tools. Our focus is on technologies intended for supporting people who want to change something about their health behaviors. In this monograph, we do not focus on medical or clinical work, nor do we focus on tools that encourage people to change behaviors they do not wish to change.

Most mobile wellness applications are built on top of three common functions: collecting data about health-related behaviors, providing users with feedback about the data they are tracking, and helping users to set and track progress toward goals. In this monograph, we focus on this common base. Of course, wellness applications may use other strategies in addition to these three (see [53] for a review). For instance, social influence — sharing of health-behavior information within the application and on social networks, competition, and provision of social support — is an increasingly common strategy used in wellness applications. Such social features are found both in commercial applications (e.g., Fitbit, Nike+, Jawbone UP) and in research projects (e.g., [8, 19, 35, 47]). Similarly, health is one of the key domains where gamification strategies have been used, and there is a growing number of mobile games designed to promote healthy behaviors (e.g., [38, 61, 72]). Such strategies are important and deserve careful consideration in their own right. Yet, these more advanced intervention strategies are often built on top of behavioral data tracking, self-monitoring and goal-setting, and those foundational features need to be designed well for the more advanced features to be effective. For this reason, we focus on the design of that common foundational base in this review.

One other note on scoping: as we mentioned, there are already a number of reviews that examine the use of SMS to encourage health behavior change. As our interest is in the opportunities that new mobile technologies are creating for supporting health and wellness, our focus in this monograph is on native applications and sensing systems that these new developments have made possible.

The remainder of this review is organized as follows. In Sections 2 through 4, we review the different ways in which behavioral data tracking, self-monitoring feedback, and goal-setting have been implemented in mobile health and wellness applications. For each of these three features, we consider the tradeoffs of different implementations and many outstanding design challenges. Finally, in Section 5, we discuss other areas that we believe need to be further investigated by HCI researchers and designers to truly make these types of mobile health and wellness technologies effective for helping people live healthier lives.

# 2

---

## Collecting Behavioral Data

---

Many mobile wellness applications provide a mechanism to let users record and track metrics related to the health activities that the application is trying to support. Tracking behavioral data is important for two reasons. First, the very act of tracking health behaviors can help people to change those behaviors in the desired direction. This phenomenon, typically referred to as *reactivity of self-monitoring*, has been extensively studied in psychology and health sciences since the early 1970s. Research has shown that tracking can help people modify a broad range of behaviors, from eating and exercise to hair pulling and obsessive ruminations [54, 68]. Simply recording the number of steps that one takes can help a person become more physically active, and tracking what one eats can help a person lose weight [76]. Many wellness applications take advantage of reactivity of self-monitoring to support healthy lifestyles by enabling users to track activities they wish to change.

Second, the data obtained through tracking can provide other important types of support, including graphs that help people reflect on patterns in their activities, goal-setting, different types of social influence (e.g., sharing step counts with fitness buddies), and gamification

(e.g., getting badges in a game based on how much exercise a person gets). Behavioral data is an essential ingredient of such interventions, making collection of the data a core function of many wellness applications.

How exactly behavioral data is collected varies from application to application and across different types of health behaviors. Though food intake may be tracked differently than physical activity, some themes cut across the different types of data. These include questions about *what data to collect* (e.g., is it sufficient for a physical-activity application to only track steps or should it support tracking different types of activities?), *at what granularity* (e.g., steps vs. minutes of walking or running), and *how data should be captured* (e.g., are activities automatically tracked or does the user need to journal them manually?). The ways in which a particular application answers these questions has a lot to do with how laborious data entry is for the user; what sensors are available, how accurate they are, and how usable they are; what can be done with the data; how and where the data will be stored; and, ultimately, how useful the data — and the application as a whole — will turn out to be.

In what follows, we examine these issues with respect to tracking food intake and physical activity — two behaviors commonly targeted by mobile wellness applications today. In addition, we address a number of general design issues related to the collection of behavioral data, such as the accuracy of the data, the user’s ability to edit and modify data that is automatically collected by a technology (e.g., by sensors), and who has access to and control over the data. Like the design of data entry, these issues can meaningfully affect users’ perceptions of an application and their willingness to use it longterm. We conclude this section by proposing open questions for HCI researchers and designers in the area of collecting behavioral data.

## **2.1 Tracking Food Intake**

Food intake (or diet) tracking is a common feature of mobile wellness applications. Some applications only focus on tracking food (e.g.,

*GoMeals*,<sup>1</sup> LiveStrong.com’s *MyPlate*,<sup>2</sup> *POND* [3]), while others include food tracking as a component of a more comprehensive wellness system that also includes tracking of physical activity (e.g., [14, 26, 90]; *Jawbone UP*,<sup>3</sup> *MyFitnessPal*<sup>4</sup>) and other health-related metrics (e.g., [1, 87]; *Calorie Counter PRO*<sup>5</sup>).

Whether food tracking is an application’s only function or a part of a more complex intervention, people who design food tracking functionality in mobile wellness applications are faced with the same problem: how exactly should the user experience of the tracking be designed? Unlike tracking physical activity, which can be partly or sometimes fully automated through the use of sensors, tracking of food is still predominantly a manual activity, making the burden of data entry a concern.<sup>6</sup> In addition, given the diversity of foods people eat, as well as the diversity of diet-related goals (e.g., losing weight, reducing intake of sugars, eliminating dairy, controlling blood glucose levels, or following the *Slow-Carb Diet*), the types of diet information that users might want to track can vary a great deal, increasing the potential complexity of the tracking functionality. For these reasons, achieving an effective food-tracking interface is a challenging design problem.

In this section, we review the main approaches that have been taken in tackling this problem and the tradeoffs of these approaches. These approaches include the tracking of individual food items with the help of a dietary database, tracking categories of food, tracking food using photos or audio, and automatically tracking food.

### 2.1.1 Tracking Individual Food Items Via a Database

One of the most common approaches to tracking food is to support logging *all* of the food and beverages that a person consumes — that is, her complete caloric intake — with the help of a dietary database.

<sup>1</sup><http://www.gomeals.com/> {Link verified 25 Aug 2013}

<sup>2</sup><http://www.livestrong.com/myplate/> {Link verified 25 Aug 2013}

<sup>3</sup><https://jawbone.com/up> {Link verified 25 Aug 2013}

<sup>4</sup><http://www.myfitnesspal.com/> {Link verified 25 Aug 2013}

<sup>5</sup><http://www.mynetdiary.com/> {Link verified 1 Sep 2013}

<sup>6</sup>This might finally be changing, however; see, for example, Noronha et al. [69], which we discuss below.

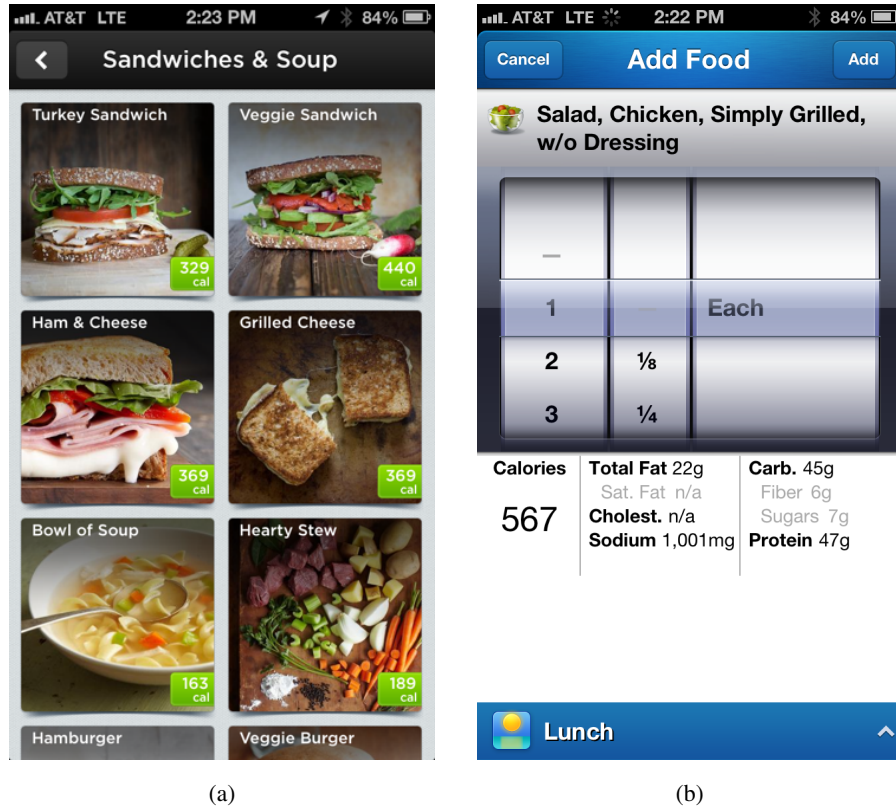


Fig. 2.1 **Food selection in a dietary database.** In many wellness applications, users can track the food they eat and drink by selecting their foods and beverages from a dietary database. While the look of the database can vary — (a) *Jawbone UP* uses photos, while (b) *LoseIt* does not — the basic interaction is the same: the user selects a food or beverage from a category or searches for it using a free-text search.

The basic interaction in this approach is selection of a food from a dietary database built into the application (Figure 2.1). Whenever the user eats something, she is supposed to use the application to find the foods or beverages she just consumed and enter them into her food log. If done regularly, this type of tracking results in a complete and detailed log of everything the user has consumed, making it easier to understand one's eating patterns and weight changes. This is a common approach with commercial food tracking applications, such as *LoseIt*,<sup>7</sup>

<sup>7</sup><http://www.loseit.com/> {Link verified 25 Aug 2013}

*Jawbone UP* and *MyPlate*, as well as several research applications, such as *PmEB* [90] and *BALANCE* [26].

#### 2.1.1.1 Main advantages

There are two main advantages of this approach. First, the dietary databases used for data entry contain calorie information for all foods they contain. By using an application with a database to log all of her food intake, the user should be able to have a complete picture of her caloric intake. Knowing how many calories one consumes makes it easier to keep the food intake within the limits needed to reach one's weight goals. In addition, when paired with detailed logging of physical activity — as many of the applications that take this approach support — tracking of all food enables the user to see her complete daily caloric balance (Figure 2.2), so she knows how much more she can eat or how much more physical activity she needs to do to stay on track with her diet goals.

A field study of *PmEB* [90], a mobile phone application for tracking daily caloric balance, found that engaging in this type of food tracking can increase users' awareness of caloric values of different foods and of their eating patterns. As one participant noted, using *PmEB* “*made me aware of the calories I was consuming and how high in calories some foods were*” [90]. Others commented that they started looking at food labels more closely and that tracking helped them find low-calorie foods that they could substitute for higher-calorie foods they previously ate. The awareness of how much one is eating and the increased knowledge about the caloric value of different foods are important benefits of keeping track of the individual foods one eats.

Another major advantage of this approach is that the accuracy of the data does not depend on the user's ability to estimate the caloric value of what she eats. Research has shown that people are not very good at estimating the caloric value of the food they eat, sometimes significantly underestimating its caloric content and sometimes slightly overestimating it [81]. Using a database decreases inaccuracies in the data and the subsequent confusion about why dietary goals are not being met. As long as the user enters everything she ate, the calorie

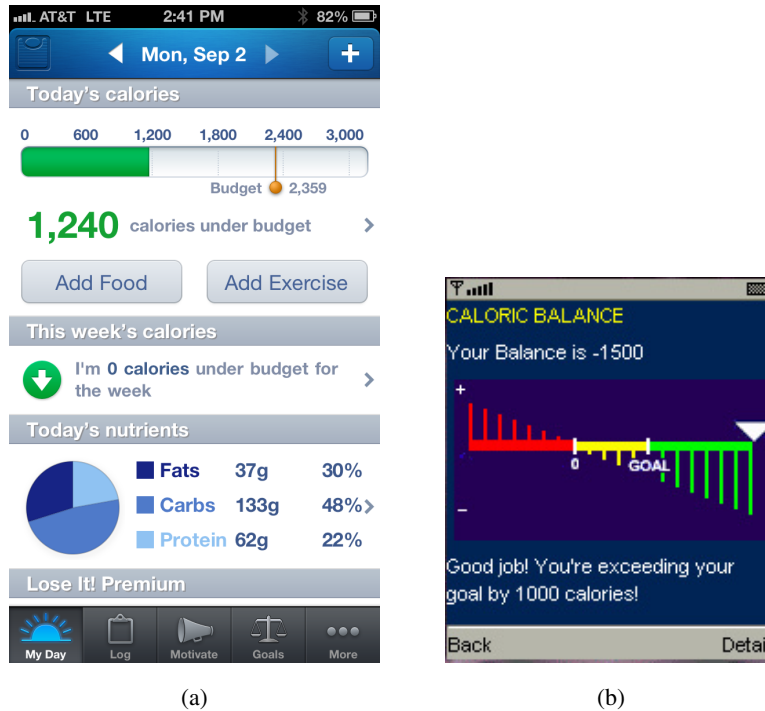


Fig. 2.2 **Feedback on daily caloric balance.** Applications that support tracking of complete caloric intake as well as of all of a user's physical activities can provide an accurate picture of the user's current caloric balance — that is, how many calories she is burning versus taking in. The images above show how caloric balance is visualized in (a) *LoseIt!*, a commercial application, and (b) *PmEB* [90], an early research application.

information in the application will be reasonably close to being correct. As we will shortly see, entering the food accurately can be problematic, but database-based food entry still makes the process less error-prone than if the entry weren't supported by a database.

### 2.1.1.2 Main disadvantages

In spite of its common usage in mobile wellness applications, tracking individual foods is not without problems. Chief among these is the laboriousness of data entry for the user, which can be particularly problematic if the intent is for the application to be used over the long term. For this approach to work as intended, it needs to be complete, which



means that the user needs to be able to enter into the application all the food that she eats. Even simple meals often contain multiple food items, however. For breakfast, a person might eat two scrambled eggs, a slice of toast, a piece of cheese, and a cup of coffee with milk. Just for this one seemingly simple meal, then, the person needs to enter four different food items into the application. And were those scrambled eggs made using Paula Deen’s recipe for *The Lady’s Perfect Scrambled Eggs*<sup>8</sup> (which includes sour cream, butter, and cheddar cheese), Alton Brown’s recipe for *Scrambled Eggs Unscrambled*<sup>9</sup> (which includes more modest amounts of milk and butter and doesn’t use sour cream or cheddar cheese), or another recipe? Was butter spread on the toast? What kind of milk was in the coffee? How much? Further, people commonly eat more complex meals than this example. Insofar as each food needs to be entered separately — and knowledge of what exactly was used to prepare the food may make a big difference in its caloric value — tracking can get burdensome very quickly. In fact, that is precisely what studies of applications that use tracking of individual foods show (e.g., [90]).

To deal with this challenge, wellness applications often include functionality to make data entry easier. One common strategy is to use the phone’s camera to scan bar codes on food packaging, so the user does not have to manually look up the food in the database. Although this can speed up data entry, bar code scanning only works for packaged foods purchased in supermarkets, leaving out a large number of common sources of food. Another way of making data entry easier is for the application to provide a list of favorite foods that contain food items that the user eats often. A related strategy is for the application to keep a separate list of all foods that the user has previously entered, so that those can be accessed again more quickly. *PmEB*, for example, keeps foods that were previously tracked in a separate personal database, so users don’t have to search through the much larger master database if they simply want to re-enter something they’ve tracked

<sup>8</sup><http://www.foodnetwork.com/recipes/paula-deen/the-ladys-perfect-scrambled-eggs-recipe/index.html> {Link verified 25 Aug 2013}

<sup>9</sup><http://www.foodnetwork.com/recipes/alton-brown/scrambled-eggs-unscrambled-recipe/index.html> {Link verified 25 Aug 2013}

before. Another strategy is to group food items by meal type, so that, for instance, all previously entered breakfast foods are grouped together in a *Breakfast* category and all foods that the user eats for afternoon snacks are stored in an *Afternoon Snack* category. Both *PmEB* and *LoseIt* take this approach. Insofar as certain foods are typically eaten only for certain meals (e.g., toast or cereal for breakfast), such grouping can speed up data entry. Finally, to deal with the issue of complex meals, applications sometimes allow users to save not only individual food items but also complete meals. Both *BALANCE* [26] and *LoseIt* support this type of functionality. So, if the user often eats a two-egg breakfast with toast and roasted tomatoes, she could save an entry for the whole meal and then enter it as a single unit whenever she eats that breakfast. For applications that support this functionality, once a meal is entered into the application it becomes much faster to enter on subsequent occasions. Although they do not completely eliminate the problem of laborious data entry, strategies such as these at least partly mitigate it, making the tracking of individual foods less frustrating, especially after the foods or meals are entered for the first time.

Another disadvantage of the individual-food tracking approach is that the usability of the application is highly dependent on the completeness of the underlying nutritional database. Unfortunately, even the best available databases, such as *CalorieKing*,<sup>10</sup> still have significant limitations. For example, although the best databases may have excellent coverage of common food items (e.g., eggs, milk, and pasta), packaged foods (e.g., Haagen-Dazs' "Dulce de Leche Ice Cream"), and foods sold by restaurant chains (e.g., Subway's "Tuna Sandwich"), users report that they are currently rather limited in their coverage of ethnic foods, and they don't contain information about foods from small restaurants or home-cooked meals. As a result, people using food-tracking applications often come across foods they want to log but which are not in the database. Each of the 15 participants in the four-week field study of *PmEB* experienced this problem at least once during the study. And the situation is even worse for users who typically eat ethnic or cultural cuisines. In Siek et al.'s [83] study of a

---

<sup>10</sup><http://www.calorieking.com> {Link verified 25 Aug 2013}

PDA-based application for nutrition tracking aimed at patients with chronic kidney disease, the participants had difficulty finding foods from discount stores — a common source of their food — in the application’s database. Although database food coverage continues to improve over time, cooking at home and eating in non-chain restaurants guarantees that some foods will not be in the database, which complicates tracking for the user. And, ironically, the less one tries to eat packaged and fast foods (which is a common suggestion to improve health and wellness), the more difficult food tracking using the database approach becomes.

### 2.1.1.3 Summary

To summarize, the main advantage of tracking individual foods with the help of a dietary database is the precision of the resulting data, which enables users to track their daily calorie intake relatively accurately and to set up precise weight management plans (e.g., *LoseIt* will calculate the exact daily caloric deficit the user needs to reach her desired weight by a certain date). The labor involved in doing this type of tracking, however, is substantial for the user, and the incompleteness of even the best nutritional databases means that the data entry effort cannot be completely eliminated even with the use of mitigating strategies such as separate lists of foods that the user commonly eats. For some purposes, this trade-off might be the right one to make; for others, though, simpler forms of tracking, like those we describe next, might be sufficient.

### 2.1.2 Tracking Food Categories

An alternative to detailed tracking of individual foods is to track food intake in terms of coarser categories, such as “fruit and vegetables” or “heavy meal.” While such tracking remains manual, not needing to find each food item in a database greatly simplifies and reduces the burden of data entry on the user. Depending on the goals of the food-tracking application, food-logging interfaces that follow this approach can often be reduced to a single screen with a handful of buttons that represent different food categories, making it possible to log a meal with a single tap. Of course, the data resulting from category-based tracking

is less granular and precise than what is obtained through tracking of individual foods with the help of a dietary database, but the gain in ease of use is substantial and, as we will see, the resulting data can still be valuable. For many applications, this trade-off could be the right one to make.

What exact categories are used to track food depends on the purpose of the application. Wellness applications that use category-based tracking have taken several different approaches. For example, *Wellness Diary* [64], an application for supporting weight management and healthy lifestyles developed by Nokia Research, focuses its categorization scheme on the amount of eaten food. The intent of the application was for it to be used long-term, so the food tracking was designed to be as lightweight as possible. Recording of food intake takes place through only four categories: “heavy meal,” “light meal,” “heavy snack,” and “light snack” The user roughly estimates how much she has eaten and records the meal or snack as being one of these categories. *Few Touch* [5], an application for encouraging healthy eating for patients with type 2 diabetes, takes a similar approach, but with an added focus on the foods’ carbohydrate content. *Few Touch* uses six categories: “high carb snack,” “low carb snack,” “high carb meal,” “low carb meal,” “high energy drink,” and “low energy drink” (Figure 2.3). The categories were chosen based on formative interviews and focus groups with diabetic patients who expressed that their main nutritional goals were to increase their number of daily meals, reduce eating of high-carbohydrate foods, and increase consumption of fruit and vegetables. Tracking of high versus low-carbohydrate foods and drinks, as well as the total number of times the user eats, supported those goals.

Depending on the application’s goal, category-based food tracking can be even simpler. Gasser et al.’s *Mobile Lifestyle Coach* [35] aims to support healthy lifestyles by encouraging physical activity and consumption of fruit and vegetables. As such, the only food tracking that the application uses is the number of servings of fruit or vegetables that the user consumed. Each serving of fruit or vegetables, and each ten minutes of cardiovascular exercise, counts as a single “lifestyle point” and the points are used to calculate users’ progress toward their daily goals and, in the social version of the application, toward the wellness



Fig. 2.3 **One-touch category-based diet logging in the *Few Touch* application.** The *Few Touch* application [5] tracks diet with single taps on one of six categories of food and drink.

goals of the user's team. Similarly, in *Health Mashup* [87], users track their food along only two dimensions: the amount they ate that day, and how healthy they think their diet was that day. Both ratings are made on a five-point Likert scale, making the food tracking a matter of a few seconds at the end of the day.

When more precision is needed, tracked categories can be made more fine-grained. *POND* [3], a mobile phone-based food diary, tracks foods based on nutritional components identified in the Healthy Eating Index (HEI) [46], an index developed based on the USDA's 2005 dietary guidelines [94]. The index assesses the healthiness of a person's diet based on the composition of the food consumed over the course of a day. HEI identifies twelve types of ingredients, including dark green and orange vegetables, eggs, dairy, meat, and oils, and provides daily goals for each ingredient. *POND* follows the HEI model (with slight modifications), and enables users to track what they have eaten by entering the number of servings they have consumed of each type of ingredient.

Although more complex than the category systems we reviewed so far, POND supports the tracking of more dietary information while still not requiring users to log each individual food separately.

### 2.1.2.1 Main advantages

Not surprisingly, the chief advantage of category-based food tracking is its simplicity and low burden for the user. Entering food by tapping a single button or by moving a couple of sliders is far faster than looking up in a database every individual food item that the user eats. As a result, people who use applications that employ category-based logging can often maintain high levels of use for extended periods of time. For example, Arsand et al. [5] report that in a six-month study of *Few Touch*, participants averaged 5.1 food and drink entries a day. Similarly, in a three-month study of *Wellness Diary* [64], participants averaged 3.15 food entries a day. Maintaining a food diary over such periods with traditional, per-food methods has been difficult [6], so these results are encouraging.

Low burden of category-based tracking would matter little if the resulting data were not useful. Luckily, category-based tracking can support important functions of wellness applications. First, qualitative data from studies that have used this type of tracking indicate that category-based tracking can be effective at increasing awareness of and reflection on eating behaviors. Participants in Mattila et al.'s [64] study of their *Wellness Diary* expressed that logging their food intake using the application made them more conscious of what they ate and as a result they began to choose less calorie-dense foods. Similar sentiments were expressed by participants in the studies of *Few Touch* [5] and *Mobile Lifestyle Coach* [35], and in both of those studies, participants reported an increase in their consumption of fruit and vegetables. These findings suggest that even stripped-down logging of food intake might result in reactivity of self-monitoring, a key feature of mobile wellness applications, although these results would need to be confirmed through more rigorous evaluations.

Category-based diet tracking can also provide useful quantitative data. A study of *Health Mashup* [87] found that the data recorded from

the two Likert-scale questions about diet was rich enough to enable the application's algorithms to find correlations between diet and participants' other activities (e.g., that a participant ate more on days when she got less sleep). One problem encountered during this study was the regularity with which users logged their food intake, not with the granularity of diet data. They addressed the problem of the regularity of logging by incorporating an unobtrusive notification in the application to remind users to log their food data at the end of the day [9]. These findings suggest that for some purposes, category-based food tracking might be good enough — that even apart from the reactivity of self-monitoring, the data collected via category-based tracking can be rich enough to support more advanced functionality of wellness applications.

#### 2.1.2.2 Main disadvantages

Of course, this data collection method is not without its downsides. Three are particularly relevant. First, category-based food tracking provides less information than tracking of individual foods. For some applications, such as those that support the tracking of a daily caloric balance, the data that category-based tracking can yield simply will not be precise enough. For applications targeting athletes who are trying to fine-tune their performance, or for weight-loss programs based on calorie counting, this method of diet tracking is probably not the solution.

Second, the usefulness of the data obtained through category-based tracking depends on the closeness of fit between the tracking categories, the application's goals and the users' ability to easily and accurately categorize the food they eat in terms of the categories used in the application. Regarding the first point, some of the participants in the study of *Few Touch* [5] complained that the application's categories were “*a bit rough.*” Although the review does not elaborate on this point, one potential issue might have been the lack of a separate category for fruit and vegetables — an explicit dietary goal that the application was trying to support. To be effective, category-based tracking needs to have sufficient coverage — it needs to include the types of things that the application's users will want to track or that the application

is trying to encourage. The granularity might not need to be high (e.g., *Health Mashup* only rates the overall amount and healthfulness of daily food intake), but if the user believes there are missing categories, the experience of using the application can be adversely affected. Regarding the second point, users might not be able to accurately categorize their food, especially if the categories used are somewhat vague, such as the “heavy meal” and “light meal” used in *Wellness Diary*. This situation can lead both to user frustration due to not knowing how to log a meal and to inaccurate data, such as if a high-calorie muffin is reported as a light snack. The lack of user knowledge can be particularly problematic for ingredient-based categories, such as those used by *POND*, as users might not understand the nutritional content of what they are eating. *POND* mitigates this problem by providing a fallback to a database lookup for individual foods (which also helps users learn about ingredients in different foods, though at the cost of increased effort), but without such a fallback, the problem could be significant.

Finally, category-based tracking makes it more difficult to discover why a wellness program is not working as expected. Without the food item-level data, it could be difficult to determine why a user is not losing weight or why a user’s blood sugar levels are higher than expected. In such cases, the user might need to switch to a more detailed data-collection method, at least temporarily.

### **2.1.2.3 Summary**

Although detailed food-item information can be very useful, for many wellness applications, the type of less detailed information that can be obtained through category-based logging might be sufficient. In those cases, category-based food tracking presents an alternative with many of the same, or at least similar, benefits (e.g., the ability to detect trends and reactivity of self-monitoring), but with much lower burden on the user, potentially enabling longer, more sustained data collection.

### **2.1.3 Tracking Food with Photos and Audio**

Arguably the easiest form of logging food consumption is multimedia, specifically, photos and audio. The use of photos has been the more



common strategy, employed both by research applications [8, 14, 63] and commercial ones (e.g., *The Eatery*<sup>11</sup> and *PhotoCalorie*<sup>12</sup>). Rather than enter textual and numeric data about food, this dataentry strategy enables users to use the camera built into their mobile phones to simply take a picture of the food they are about to eat. The application can then use the picture in different ways, such as displaying it on a timeline or sharing it with friends or healthcare providers. Audio recording works similarly: the user speaks into the phone what she has eaten, creating an audio record of the meal. In the HCI literature, audio recording has mostly been used in applications that target users from low-income environments where literacy might be an issue [83, 37].

### 2.1.3.1 Main advantages

A key advantage of the use of photos and audio is the ease of interaction. A photograph can capture a complex meal consisting of several foods and complicated ingredients (e.g., sauces) just as easily and quickly as it can capture a piece of toast, and a 10- or 20-second audio recording can capture a great deal of information about the food, including where it came from and what it consisted of, with little effort on the part of the user. This can be particularly helpful for people who eat a lot of foods that are not typically included in nutritional databases, such as ethnic foods or foods from non-chain restaurants. Finally, as we have alluded to, the use of multimedia can provide a way to capture diet information for people for whom traditional text-based data collection methods might be challenging due to low literacy or motor or cognitive impairments.

In addition to ease of capture, photos and audio data have unique affordances that can make them, for some purposes, more useful than traditional textual and numeric data. When viewed by people (as opposed to processed by computers), photos of food can convey a great deal of rich information and can be inspected very quickly. This expressiveness of photos enables a range of applications, from self-monitoring to health education. For example, Brown et al. [14] created an application that lets users use their mobile phones to take pictures of their

---

<sup>11</sup> <https://eatery.massivehealth.com/> {Link verified 25 Aug 2013}

<sup>12</sup> <http://photocalorie.com/> {Link verified 25 Aug 2013}

food and of their physical activities and then displays those images in a timeline view on the user's desktop. Users can inspect the timeline at different time scales (e.g., a single day, several days, or a week) and can quickly see how they have been eating and how active they have been over time. Even without any automatic processing of the data, this type of a visual journal can act as a reminder to eat healthier and exercise and can easily reveal that a person has been slipping on her wellness behaviors or goals. Similarly, users of the wellness application *VERA* [8] can use their phone's camera to take pictures of any activity that they think affects their health, including what they eat. Users can also annotate these images, creating a free-form journal of their engagement with their health and the way they understand their health goals. Their visual journal can also be shared with other *VERA* users, who can provide feedback, encouragement, or even interact with the user. Studies of *VERA* have shown that the interactions that are generated in this way are quite rich and diverse, which is partly enabled by the richness of information contained in the shared photographs.

In addition, photos can be an excellent medium for supporting education. *MAHI* [63], a mobile phone-based application for people with diabetes, lets users take pictures of their food and upload the pictures to a Web site where the photos can be shared with a diabetes educator. By seeing the pictures, the educator has more concrete information about each user's self-management that the educator can use to help provide coaching for the user, and can, for example, explain to the user why her blood sugar reacted in a certain way by correcting the user's misconceptions about the carbohydrate content of a large muffin. Similarly, *The Eatery*, a commercial food journal for the iPhone, lets the user take photos of her food and then rate how healthy the user thinks the recorded meal was. The image of the meal is then anonymously shared with other users of *The Eatery* so they can also rate its healthfulness. The user's food journal displays not only her own rating of the meal but also the average of other users' ratings, which can help the user correct her understanding of the healthfulness of the food she eats.

Finally, a unique advantage of using photos to record food intake is that the recording has to happen before — rather than after — the person begins to eat. Photo recording thus injects itself into the eating

process, providing an opportunity for reflection before the food is eaten. This, in turn, can lead to a change in the decision on whether or not to eat the food. In a study that compared paper and photo diet journals, Zepeda and Deal [100] found precisely this. They explain [p. 696]:

*With the written method, participants may or may not have evaluated their food choices at the end of the day, but when they were forced to photograph something before they could eat it, they had to decide if it was really worthwhile to consume. The extra effort involved in the consumption made them evaluate the decision more carefully and led many to recognize the contexts in which they were engaging in poor dietary habits.*

Audio provides a different set of affordances. In particular, audio recording makes it easy to annotate the food entry with additional information, such as where the food was purchased or what changes the person made to the food (e.g., if the person took off the bacon from the salad). Grimes et al. [37] used this feature of audio recording as a centerpiece of their *EatWell* application, which encouraged healthier eating in an urban African American community. In *EatWell*, users use their phones to call into a server where they can record their experiences with trying to eat healthy in their community and to listen to the recordings from other users. Hearing others' voices made the system feel more personal, and the spoken stories were perceived as providing more authentic, "real," and situated (to their community) information than what the users received from healthcare providers.

### 2.1.3.2 Main disadvantages

Of course, the use of multimedia for recording food intake also has downsides. A major downside is the difficulty of computationally processing the resulting information. It is challenging to take photos or audio recordings of food and convert them into information that can be graphed, compared with historical data, combined with other kinds of quantitative data (e.g., from sensors), or automatically assessed for healthfulness. For applications where such automatic processing of data

is a key part of the application's functionality, the use of photos or audio recording is probably not a good approach to capturing food intake (or it should at least not be the only approach used).

Another downside has to do with privacy. Entering what one ate into a mobile application using standard on-screen forms is an activity that can be performed discretely — it looks to nearby people as any number of other activities that people commonly do on their phones. Taking a picture of a meal or describing it into a voice memo is much more conspicuous when done in the presence of others, especially when one is eating in public (e.g., at work or a restaurant). Social norms around such self-tracking activities might be shifting, but for many people, the attention that photo or audio capture might draw could very well be unwanted and experienced as embarrassing. This is particularly the case for photo capture which, by virtue of needing to be done before one eats, cannot be delayed until the person is in a more private environment.

The use of photos for food logging can also be complicated by the fact that a meal that is only partially eaten requires two images — before and after — so that the amount of food that has been eaten can be accurately represented. Applications that use photo logging thus need to provide a way to store more than one image to represent a single food entry.

Finally, photo and audio capture can be made more difficult by sub-optimal environmental conditions. Dark environments and low-contrast foods can make it difficult to capture good images that are easy to interpret, and noisy environments can interfere with the person's ability to create easily understandable audio recordings.

### **2.1.3.3 Summary**

The food tracking data obtained through the use of photographs and audio recording is very different than the data obtained through traditional food tracking techniques. For some purposes, such as conveying a lot of rich information for direct human inspection, the multimedia data are often well suited; for others — particularly those that require a lot of computational data processing — photo and audio data are usually a

poor fit. The ease with which such data can be obtained, though, makes the use of photos and audio an attractive data collection strategy and one that a growing number of wellness applications are turning toward. This trend is likely to continue, especially since some of the downsides of this approach are beginning to be addressed by crowdsourcing and computer vision research. We turn to these developments next.

#### 2.1.4 Tracking Food Automatically

As we mentioned above, a major downside of using photos and audio for food tracking is that the resulting data is difficult to convert to standard food intake metrics used by many applications, such as calories, portions, and amounts of different types of ingredients. Research has begun to address this problem. Two approaches have been proposed to automatically extract nutritional information from images of food: computer vision and crowdsourcing. We briefly review related work in these areas and the tradeoffs that the two approaches currently present.

Kitamura et al. [49] developed a Web site that uses computer vision algorithms to analyze the food groups and serving sizes in images of food that users upload into an online photo journal. The processing is done in two steps. First, the system analyzes the uploaded images and identifies those that contain food. For these images, the system does additional processing to identify the types of pictured food and their amounts. Kitamura et al.'s system is based on the Japanese version of the food pyramid concept, and it classifies food into grains, vegetables, meat and beans, milk, and fruit.

An alternative approach to the automatic analysis of food images is the use of crowdsourcing. For example, Noronha et al.'s *PlateMate* [69] estimates caloric value and the nutritional content from images of food using workers from Amazon's crowdsourcing platform, *Mechanical Turk*.<sup>13</sup> *PlateMate* analyzes images in stages: the crowd is first asked to mark each separate food in an image, then to label those foods, and finally to estimate the serving size for each food and based on that to look up its caloric content and nutritional components. There is error

<sup>13</sup><https://www.mturk.com/mturk/> {Link verified 25 Aug 2013}

checking and redundancy at each stage of the process to ensure that the quality of the estimates is held relatively consistent. To evaluate the experience of using *PlateMate*, Noronha et al. recruited 10 participants who tracked their food intake for four days — two days with *PlateMate* and two days by manually estimating the calories in what they ate.

#### 2.1.4.1 Main advantages

Some of the main advantages of the computer vision approach are that it is cheap, fast, and current algorithms are already good at detecting whether or not an image contains food. As to the cost, once the algorithms are developed and trained, the marginal cost of analyzing each new image of food is zero. In addition, image recognition using computer vision is fast, so the information extracted from the image can be displayed for user review and correction as soon as the image is taken. And finally, Kitamura et al.'s [49] algorithms that detect if food is present in an image work well; their system is able to identify which images contain food with over 90% accuracy.

Some of the main advantages of the crowdsourcing approach are that it is already accurate and users may prefer it to manually tracking their food. In an accuracy study of *PlateMate*, Noronha et al. [69] found that the system's calorie estimates correlated with the ground truth data (obtained through manual calorie calculations and precise weighing of food) at 0.86, very close to the accuracy of estimates made by expert dietitians. Just as importantly, *PlateMate* did not routinely underestimate the caloric value of food — a common problem with self-report — but tended to slightly overestimate it. Similarly, in their field study where participants used *PlateMate* to track their food intake, Noronha et al. found that *PlateMate* continued to slightly overestimate calories while the participants underestimated them. Further, in the field study's exit interviews, participants reported that *PlateMate* was less “*annoying*” and “*tedious*” than manual journaling, though it's important to keep in mind that the field study was on the shorter side (i.e., two days with *PlateMate* and two days with manual tracking).

### 2.1.4.2 Main disadvantages

Unfortunately, the main downside of the computer vision approach, at least at present, is poor accuracy. The classification of food based on the food pyramid food groups that Kitamura et al. [49] use is not nearly as good as their algorithms to detect whether or not food is present in an image. Even after tuning their algorithms and using users' correction data to personalize the classification to each user, the classifier identifies food groups correctly only 43% of the time. Vegetables are detected most accurately at 50%, while meats and beans are detected correctly least frequently, at only 31% of the time. Such low accuracy rates are not just a property of Kitamura et al.'s system. Other research projects fair similarly, with accuracy rates in the 30–40% range when used with images of food that have not been staged. For example, Joutou and Yanai's [44] computer vision system for food categorization had 61.34% accuracy in cross-validation tests with the 50 food images on which the system was trained, but its accuracy dropped to 37.35% when it was applied to images of food that users freely uploaded into the system. The 30–40% accuracy range appears to be the current state of the art in machine vision classification of food images.

Not surprisingly, crowdsourcing also has downsides. Perhaps the biggest current downside of crowdsourcing is cost. Processing each image in *PlateMate* cost Noronha et al. [69] \$1.40, and this cost would be incurred for *each* new image that is analyzed. This means that for logging even 3 meals a day (no snacks or beverages), the cost of using an application that uses an approach like *PlateMate's* would be \$125 a month. Given the very low cost of most wellness applications and Web sites (many are even free), it would appear that to make an application that uses this approach to food logging feasible, the cost would need to be reduced by at least an order of magnitude. Whether the cost could be substantially reduced (or perhaps minimized with the help of something like advertising) while maintaining accuracy is an open question.

Another potential downside with using crowdsourcing is response time. *PlateMate* takes around two hours to return an estimate for an image. There are crowdsourcing-based systems that have been much

faster (e.g., [12]), but it is not currently clear how quickly nutritional estimates need to be returned to make a food tracking feature based on crowdsourcing suitable for everyday use while maintaining high levels of accuracy. The level of needed responsiveness will likely depend on the nature of the application. For an application that tracks the user's daily caloric balance — for example, LoseIt — the delay from a crowdsourced nutritional analysis would likely be acceptable. On the other hand, an application that aims to provide real-time feedback on how the user should change her meal (e.g., by eating a smaller portion or not eating certain parts of the meal) would need a much faster response time than a crowdsourcing solution can currently provide.

#### **2.1.4.3 Summary**

The application of computer vision or crowdsourcing to food tracking is an exciting area of research that promises to remove the main downside of using photos to document food intake. However, at this point, the available systems are either too inaccurate or too expensive to use in applications aimed at everyday, long-term use. As the research in this area progresses, that could very well change.

#### **2.1.5 Summary of Tracking Food Intake**

Recording data about food intake is still fundamentally a manual activity, one that can come with a significant burden on the user. Category-based tracking and use of photos are currently the most promising approaches for tracking food intake over the long-term. While tracking of individual foods has its advantages, its high user burden suggests that this approach might be best treated as a fallback for recording data when simpler approaches fail or as a function to be used short-term — for instance, when the user is starting a new diet and is trying to get an accurate picture of her food consumption, or when she is unsure why she is not seeing the results she expected to see. After the short-term period ends, the application could switch to a lighter-weight form of food tracking to help sustain use, while keeping the detailed tracking available as an option when the user feels she needs to get a more detailed assessment of her diet again. However it's



done, balancing the application's benefits with the burden placed on the user is key for sustained use of food intake-tracking applications.

## 2.2 Tracking Physical Activity

Along with food intake, physical activity is the most common behavior currently targeted by mobile wellness applications. Unlike for food intake, however, automatic recording of physical activity data has been going on for many years, from simple sensors like pedometers to a range of sensing devices and sensors embedded in mobile phones themselves. Many of these activity trackers have become mature systems that have firmly transitioned from research prototypes to commodity consumer devices, bringing the automatic detection of physical activity into the mainstream. A result of this transition is that while manual tracking still often plays a role in collecting physical activity data, automatic sensing is increasingly becoming a key feature of mobile wellness applications that focus on physical activity. We review these approaches to collecting physical activity data — manual and automatic — next.

### 2.2.1 Tracking Physical Activity Manually

Manual tracking of physical activity has a number of similarities to tracking of food intake. As with food, physical activity can be tracked at different levels of detail. At the fine-grained end of the spectrum, data provided through tracking can be used to calculate a person's caloric burn, enabling calorie-balance features in applications like *LoseIt* or *PmEB* [90]. For example, *LoseIt* contains a database with energy expenditure data for over 110 activities, from racquetball to dancing to sex. The user specifies the duration of the activity and its intensity, and the application estimates, based on the user's height and weight, how many calories she burned doing that activity. As with its food journal, *LoseIt* lets users save their favorite activities into a separate list so they don't have to navigate a long alphabetical list every time they go for another run or play another tennis match. Applications that focus on tracking structured exercise may use even more detailed journaling. Users of *UbiFit* (one of our research projects), for instance, had the option to record the type of strength training exercises they did as well as

how many repetitions and sets of each exercise they performed. Commercial applications like *iFitness*<sup>14</sup> and *FitnessBuilder*<sup>15</sup> take a similar approach, recording the weight, number of repetitions, and the number of sets for each strength training exercise the user performs. On the coarser-grained end of the spectrum, some applications have offered tracking of only certain types of activities. For instance, Gasser et al.'s *Mobile Lifestyle Coach* [35] only lets the user journal the number of minutes of cardiovascular exercise, without distinguishing whether the user ran, biked, or performed some other form of cardio.

### 2.2.1.1 Main advantages

This flexibility of determining the appropriate level of detail for physical activity data is one of the main strengths of manual tracking. For users who are interested in recording every tennis match, every strength training session, and every Bar Method class, manual tracking can handle the task (a task that can be problematic for automatic sensing, which we discuss shortly). In fact, some participants in our studies of *UbiFit* were even interested in tracking their strenuous household chores, such as scrubbing the floor or cleaning their pets' cages — which *UbiFit's* journal was able to support. And when less detail is needed, as in the *Mobile Lifestyle Coach* application, tracking can be designed to be rather lightweight.

Another advantage is that physical activity is — with the exception of walking — typically both less frequent and less variable than food intake. Even rather detailed journaling of physical activity, such as the optional detail supported by *UbiFit's* journal, requires far less effort from the user than the tracking of individual food items that she consumes. In fact, data from our three-month field study of *UbiFit* [20] indicate that participants in the condition without the sensing device were able to maintain manual tracking for 3 months and that, judging by the sentiments expressed in the exit interviews, they did not find tracking to be overly burdensome (though few participants included the optional detail in their journals). In many cases, manual

---

<sup>14</sup><http://www.ifitnessapp.com> {Link verified 25 Aug 2013}

<sup>15</sup><http://www.fitnessbuilder.com/> {Link verified 25 Aug 2013}

tracking of physical activity is likely to be a practical option, especially for applications that aim to encourage varied physical activity and not just taking more steps.

#### 2.2.1.2 Main disadvantages

As with food intake, however, there is a tradeoff between the level of detail that can be captured through manual tracking and the burden on the user. Entering details for every set of strength training exercises is substantially more work for the user than tracking the same workout as 45 minutes of strength training or even 45 minutes of strength training that focused on biceps, quads, and glutes. Similarly, manually journaling *all* of the physical activity one does throughout the day — every step the user takes, every set of stairs she climbs — is a lot more work than limiting her manual tracking to just the structured exercise that she performs.

Another challenge to manual tracking of physical activity is a version of a problem we encountered in the discussion of manual tracking of food, namely, user knowledge needed to generate accurate data. In the context of physical activity, this issue often manifests in relation to recording the intensity of the activity. Applications that use manual tracking, especially those that aim to help users manage their caloric balance (e.g., *LoseIt*), often require that the user journals not only the duration and type of her activities but also their intensity. The same physical activity, when performed at different levels of intensity, can burn very different amounts of energy. In terms of caloric burn, leisurely cruising around town on a bicycle for an hour is very different than riding at 15 mph for an hour, which is very different than riding up a mountain road for an hour. For this reason, applications need intensity information to be able to do a reasonable job of estimating the user's caloric burn. We have found, both in our own studies and in an analysis of posts from a health and wellness message board, that people can experience difficulty estimating the intensity of their activities and are prone to overestimate how intensely they exercised. This, in turn, can lead to inaccurate expenditure calculations by the application, which can then lead to users' confusion about why they

are not reaching their health goals even when they are eating the number of calories that the application calculated they could eat. Providing examples for different intensity levels within the application can help mitigate this problem. Such examples would be particularly useful if they can be embedded into the interface that users use to record their activities, rather than just putting them into a first use tutorial, which many users skim or skip altogether.

### **2.2.1.3 Summary**

As with the manual tracking of food intake, physical activity can be manually tracked by the user at varying levels of detail from very coarse to very detailed; the more detail she must provide, the more burden on the user. Manual tracking of physical activity is better suited for keeping track of structured exercise rather than for having the user track every step she takes throughout the day. One particular challenge — especially when detail is required — is that many people have difficulty accurately estimating the intensity of their physical activities, which can impact the effectiveness of features that try to help users manage their caloric balance.

## **2.2.2 Tracking Walking Automatically**

As mentioned above, for many people the main exception to low frequency of physical activity is walking. While people can typically remember their longer walks — taking the dog for a 30-minute walk after dinner, for example — the various shorter episodes of walking that one does throughout the day — going to the ATM, walking around a grocery store, getting up from one’s desk to get some water or use the restroom, etc. — are difficult to remember, partly because they are embedded in other, more salient activities, as well as the fact that they are likely to be relatively short episodes of walking. Manual tracking is poorly suited for recording such episodes of walking. Reasonably accurate self-report of walking is possible, but it comes with substantial user effort, making it infeasible as a long-term data-collection strategy [51]. Yet, for many people, walking is the only kind of physical activity

they do regularly, and increasing walking has been a target of many health interventions and public-health initiatives [92, 95].

For such reasons, automatic logging of physical activity in wellness applications was first used for tracking users' step count. Some of the early mobile wellness applications from the HCI literature, such as *Houston* (our first wellness application) [19], *Chick Clique* [88], and *Shakra* [2] all used automatic tracking of step count. Pedometers were often used to track step count, although some applications, like *Shakra*, used other forms of sensing.<sup>16</sup> One reason pedometers were (and still are) compelling is that they tend to do a pretty good job of capturing the full range of steps a person takes throughout the day, many of which would otherwise remain unaccounted for if walking were to be tracked manually. Pedometers enable users to see precisely how much — or little — they walk and how their step count varies from day to day. *Houston*, *Chick Clique*, and other early mobile wellness applications leveraged pedometers to encourage users to find even small ways to be more physically active in their daily lives, thus helping people become more active even if they could not or were not willing to adopt a structured exercise routine. This approach of encouraging opportunistic physical activity was to a large extent enabled by automatic tracking of step count, which made it possible to quantify and make visible how even small changes in users' daily routines (e.g., where they parked, whether they took a lunchtime walk) impacted their daily activity levels.

Since the early mobile wellness applications that used pedometers, automatic tracking of physical activity has evolved. One way it has done so (we discuss other ways below) is that with the recent availability of cheap 3D accelerometers, a range of small, reasonably fashionable commercial activity trackers are moving automatic physical activity tracking into the mainstream. *Fitbit*, *Nike+FuelBand*, and *Jawbone UP* are popular devices in this rapidly growing category. Like pedometers, these devices are fundamentally step counters, but many of them also track other metrics, such as total distance traveled, total active and inactive time, number of calories burned, and in *Fitbit's* case, the number of

---

<sup>16</sup> *Shakra* analyzed differences in cell phone signal strength to track activity.

flights of stairs climbed. In addition, some of the wrist-worn devices, such as *Jawbone UP* and *Fitbit Force* (or *Fitbit One*, when worn in a wrist strap), can also track data about the user's sleep.

### 2.2.2.1 Main advantages

The new commercial activity trackers have a number of attractive features. First, they have been designed to fit into the user's daily life. *Fitbit One* is small enough to be carried in a pocket, and bracelet-style trackers like *Nike+Fuelband*, *Fitbit Force*, and *Jawbone UP* are designed to be reasonably unobtrusive, which makes it possible for some people to wear them with a wide range of attire. The sensors also have a relatively good battery life, typically needing to be charged every several days or so — less often than most current smartphones. Pedometers have similar advantages. Their batteries can last for days, weeks, or even months, they have become smaller and less obtrusive, and their displays provide a way to get quick feedback about how much one has walked throughout the day and over the previous several days.

Modern activity trackers also usually come with custom mobile applications where users can see their activity history and progress, set activity goals, and, increasingly, share their data with others. Bluetooth-equipped sensors, like some models of the *Fitbit* and *Nike+*, can synchronize with the user's mobile phone wirelessly without any explicit user interaction, and other sensors like *Jawbone UP* can be connected directly to the phone, rather than needing a computer for data synchronization. Finally, some companies that make activity trackers have established relationships with other companies that support health and wellness activities, enabling their users to combine data from multiple sources. For example, a *Fitbit* user who doesn't like *Fitbit's* food tracking interface can connect her *Fitbit* account with her *LoseIt* account, allowing her to use *LoseIt's* food logging while still having the benefit of *Fitbit's* automatic tracking of step count.

### 2.2.2.2 Main disadvantages

From the user's perspective, perhaps the biggest downside of these devices, including standard pedometers, is the narrow range of physical

activities that they detect. Even the popular commercial activity trackers are little more than smart pedometers, and they do not explicitly recognize any physical activity other than walking and running (and even running is often treated as just more steps — rather than the more vigorous cardiovascular activity that it is). Other activities, such as tennis, biking, or using the elliptical trainer might generate some additional steps and are usually included in an “active time” metric, but those activities often have to be tracked manually to get proper credit for them. In our own work with physical activity sensors [20, 22], we discovered that people are often disappointed by the failure of sensors to detect other activities, even if they understand that the sensor is not designed to do that. They are particularly disappointed when they do something intense (such as vigorous gardening — a real example from our work), and the sensor doesn’t detect anything at all. Wellness applications that use sensors to track physical activities should find a way to handle such expectations from users. We will come back to this point.

From the perspective of developers — research or commercial — of mobile wellness solutions, a downside of using these devices is that it is often difficult to get activity data out of the systems. As of this writing, *Fitbit* is the only popular activity tracker provider that has a robust application programming interface (API), and even *Fitbit’s* API has serious limitations.<sup>17</sup> Most importantly, *Fitbit* provides API access only to the daily activity data (total number of steps for the day, total active time, etc.), and not to the intraday data (i.e., data about each walk the person took during a day or when active or inactive periods were during the day).<sup>18</sup> In addition, to our knowledge, neither *Fitbit* nor other companies with trackers currently on the market provide access to the raw accelerometer data that could be reprocessed and reanalyzed by a third-party application. Such limitations constrain the kinds of

---

<sup>17</sup> *Nike+* recently released an API as well, but it is not nearly as robust as *Fitbit’s*, at least not yet.

<sup>18</sup> *Fitbit* does have a way for developers to get access to intraday data, but this requires developers to be granted special-level API access. We could not find any explicit information about what kinds of developers or applications are eligible for such access.

applications that can be developed in conjunction with commercial tracker devices.

Finally, none of the companies that make popular tracker devices allow users to correct their data when the user discovers an error in the data (e.g., that the device didn't detect the duration of a walk correctly). From the standpoint of user experience and system credibility, this is a serious downside to using tracker companies' own applications. In principle, this downside could be mitigated by using a tracker in conjunction with a third-party application, but to do so, the third-party application would need to be able to get fine-grained, per-activity data, which is, as we mentioned, difficult to get even from *Fitbit* and impossible from most other trackers to date.

### 2.2.2.3 Summary

Pedometers and commercial activity trackers such as *Fitbit*, *Nike+Fuelband*, and *Jawbone UP* are pretty good at capturing the range of steps that a person takes throughout the day — something that is notoriously difficult for users to track manually. The devices are relatively small, unobtrusive, and often have good battery life. However, they do not capture the range of healthy physical activities that people do, which means that they might be best used by mobile wellness applications when they are accompanied by some form of manual tracking so that the user can record activities that the devices do not detect.

### 2.2.3 Tracking Other Physical Activities Automatically

Another line of development in automatic tracking of physical activity has been around systems that sense a wider range of physical activities. For sensing, *UbiFit* used the *Mobile Sensing Platform (MSP)* [17], a pager-sized, wearable research prototype co-developed by Intel Labs and the University of Washington. We trained the *MSP* to detect five types of physical activities: walking, running, cycling, using a stair machine, and using an elliptical trainer. These activities represent some of the most common forms of cardiovascular exercise, which means that for many people, the *MSP* could minimize the need for manual tracking.



However, the *MSP* was a research prototype that was used in several studies, but was never released for general use. Recognition of a range of physical activities by a single device is just now making its way into the commercial domain. A new commercial tracker, *Misfit Shine*,<sup>19</sup> claims to detect running, walking, biking, and swimming. *Shine* was recently released, and other devices with similar functionality are likely to follow. As with the step-based sensors mentioned earlier, *Shine* and the *MSP* are separate, special purpose devices.

Yet another line of development in automatic tracking has moved the detection of physical activities into mobile phones themselves (e.g., [80]). Over the last few years, sensors have increasingly been integrated into mobile phones, and even some of the cheapest smartphones today come equipped with an accelerometer, GPS, and gyroscope. The inclusion of these sensors has enabled the detection of physical activities directly on the phone, obviating the need for a separate device. A number of wellness applications have taken this approach to tracking activity. For example, *Moves*<sup>20</sup> is an iPhone application that runs in the background and continuously monitors the user's movements and physical activity. Based on data from the phone's GPS and accelerometer, *Moves* displays a timeline of the user's physical activities and location changes over the course of the day. Similarly, *BeWell* [56] uses phone-based sensing to monitor physical activity (walking, running, and stationary activity), along with sleep and social interactions, and provides ambient feedback on the phone's wallpaper about how the user is doing on these well-being dimensions.

### 2.2.3.1 Main advantages

As the range of activities that these sensors can detect improves, the burden on the user of having to manually track their activities goes down. Regarding the special purpose sensors, because they do little else, their design and battery life are getting better and better every day. Regarding sensing done via sensors on the user's mobile phone,

<sup>19</sup><http://www.misfitwearables.com/> {Link verified 25 Aug 2013}

<sup>20</sup><http://www.moves-app.com/> {Link verified 25 Aug 2013}

such an approach removes the need for an extra device — and the charging and other care that goes along with it — potentially increasing the consistency of activity tracking (as people are rarely without their phones) and lowering the cost of mobile wellness interventions.

### **2.2.3.2 Main disadvantages**

Complex, multi-movement activities such as biking or basketball are difficult to accurately detect — especially if the same sensor worn in the same place is being used to detect various types of activities. For this reason, automatic detection of physical activity will inevitably make errors. Sometimes, a person will do an activity and it will not be detected; other times, something that’s not a physical activity (e.g., riding a bus) will be detected as a physical activity; and sometimes even if the activity type is correctly detected, some of its properties (e.g., duration) will be incorrect. And this is saying nothing of activities that the sensing system is not trained to detect but which users might expect to be detected (see Consolvo et al. [22] for a detailed discussion of various types of sensing errors and user perceptions of errors). Yet, in spite of the fallibility of activity detection, none of the commercial trackers allow users to correct sensed data. A user might go for a 6 mile run, and her sensor might detect it as a 5.2-mile run or a 7.8-mile run. The user cannot correct the error.

We find this problematic for two reasons. First, in our work, we have repeatedly found that people want their activity logs to be accurate. Not being able to correct an incorrect activity record, even if the error is in the user’s favor (e.g., the device detected more activity than the person did), can be frustrating and negatively affect the system’s credibility and usefulness. Second, sensed activities are often tied to other components of the wellness application, such as goals. Users can find it infuriating when the system does not recognize that they met their goal because their activity was not detected correctly. If such errors occur repeatedly, a person might abandon the system altogether. For these reasons, we think that it is important to enable users to correct automatically detected data. Without this ability, an application can create

a great deal of needless user frustration, increasing the probability of system abandonment.<sup>21</sup>

A related issue has to do with the completeness or adequacy of the collected physical activity data. A strong temptation when developing wellness applications is to focus on the activities that the application's sensing can detect. The applications that accompany *Fitbit* and *Jawbone UP* focus on steps, and the one that comes with *Shine* focuses on running, biking, and swimming. Goal-setting, historical trends, and daily summary data are all limited to the activities detected by the sensors. What we found in our work on *Houston* — which also focused on steps — is that it is important for people to have the option to record (and receive credit for) the full range of physical activities they do, and that those activities typically go well beyond what sensors can detect today. Not being able to supplement automatic detection with manual tracking can make people feel like they are not getting credit for their efforts and can even discourage them from doing activities that cannot be recorded. In the case of *Houston*, some of the participants decided not to run, bike, walk uphill, or do other higher-intensity activities because the application only allowed them to record steps; they could make a note that they did these other things, but it didn't count toward their goal. In these cases, *Houston* actually had an opposite effect from what it was intended to do — encourage people to be physically active. Such side effects and associated user frustration can be avoided by supplementing automatic detection with robust manual tracking that can be used to record other activities that users do.

As with the step-based sensors, another disadvantage of the special purpose devices such as *Shine* and the *MSP* is that they are separate devices that need to be synced with the user's mobile phone (or in some cases, her computer), that need to be regularly charged, and that run the risk of being lost or damaged. Sensing that is integrated with the user's mobile phone, however, negatively affects the phone's battery

---

<sup>21</sup> We understand that the correction of sensed data could be problematic for some medical applications where healthcare providers need to be able to rely on the sensed data to get an accurate picture of the patient's functional status. Our focus here is on wellness applications, however, which typically do not have such requirements but for which minimizing user frustration is a major design goal.

life. In our personal experiences with *Moves*, for instance, with the application enabled, the iPhone routinely runs out of battery by 5 or 6 pm; without *Moves* running, the phone's battery would not only last all day, but would typically still have about 30% battery life remaining at bedtime. The battery hit can make phone-based sensing problematic for heavy phone users or for users who do not charge their phones frequently. In addition, for some forms of physical activity — running or biking, for instance — the small size of a dedicated tracker can be preferable to needing to carry the phone. Finally, though many people often have their phones nearby, not everyone carries their phone on their body in such a way that the sensors on the phone can accurately detect their physical activity. For example, for people who tend to keep their mobile phones in their bag or on a table, phone-based tracking can be far less accurate than using a dedicated, wearable device.<sup>22</sup>

### 2.2.3.3 Summary

Two different approaches are being investigated to help people automatically track more activities than just step count. One approach is to use more robust special-purpose sensing devices that can track a broader range of activities, such as the *MSP* research prototype that we used in our work with *UbiFit*, and the recently launched commercial device, *Misfit Shine*. Another approach is to use sensors already embedded in the user's mobile phone to detect activities without her having to carry or wear a special purpose device. We expect to see more developments in this space soon.

### 2.2.4 Summary of Tracking physical Activity

To briefly summarize, then, the most common forms of tracking physical activity used today are manual journaling and sensor-based automatic tracking. Given the comparatively low frequency of physical activity, manual journaling in this domain is less laborious than for food intake and it can be successfully maintained over long periods of

---

<sup>22</sup>In a 4-week field study of 28 smartphone users, Dey et al. [27] found that though participants were usually within the same room as their phone, it was only within arm's reach about half of the time, and "arm's reach" isn't necessarily on body.

time, depending on the level of detail required. The main exception to this rule is walking, which is difficult for users to capture accurately with manual tracking, especially if *all* walking episodes throughout the day — including short, non-structured walks — are of interest. A range of sensors fills this gap, enabling automatic detection of walking and, increasingly, other forms of cardiovascular exercise. While these sensors are relatively accurate and integrate well into people’s lives, they are still error-prone and the range of activities they can detect is relatively narrow when considered with the range of activities that many people actually perform or would like to perform. For these reasons, robust wellness applications should consider supplementing automatic detection with manual tracking and provide a way to correct errors in sensed data. Doing so enables users to maintain an accurate record of their activity, increases their trust in the system, and reduces frustration and associated risk of system abandonment.

## 2.3 Broader Considerations about Collecting Behavioral Data

While different forms of behavioral data have unique characteristics, a number of common considerations cut across different behaviors and states that are commonly tracked by wellness applications. Next, we briefly review those more general issues.

### 2.3.1 Adequacy of Data Coverage

As we mentioned, an important finding from our study of *Houston* was that our application had an unintended effect: some participants decided to forgo more intense physical activities because they would not receive the proper credit for those activities in the application. In the interviews at the end of the study, participants discussed how frustrated they were that they could not log other activities and worse, when they performed those other activities, the application sometimes made them look like they were being inactive. As a result of this experience, we designed *UbiFit* to support a much broader set of activities, including strength training and various forms of flexibility exercises, such as yoga

and Pilates. In our first field study of *UbiFit*, we found that even this was not enough, however. Our participants did physical activities that were not strictly exercise — chopping wood, scrubbing the floor — but which they still wanted to be able to track in the application (even though those activities did not count toward their physical activity goal). While *UbiFit* provided a way for users to add any activity they wanted to track (if it didn't fit into one of the main physical activity categories, they could add it to an “Other” category in the interactive application), flowers were only provided for cardio, walking, strength training, and flexibility training activities. In response to participants' reactions in the 3-week field study, we revised *UbiFit* to include a small, lavender colored flower for “Other” activities that were 10 or more minutes in duration. In the 3-month study of *UbiFit*, this “Other” category turned out to be popular: 17 of the 28 participants logged at least one “Other” activity, and a total of 61 such activities were logged during the study, including housework, shampooing the rugs, and gardening [20].

The issue of adequate coverage goes beyond physical activity and applies to wellness applications that track behavior more broadly. We submit that as a general design principle, an application should support recording all data that the users would reasonably want to record, given the goals and the nature of the application. For applications that track physical activities, this means being able to track a full range of physical activities, for applications focused on food, being able to track all the various foods that a person might eat and beverages she might drink, and so on.

It is important to note that we are not arguing that each wellness application should include long, exhaustive lists of activities, foods, and other behaviors. While that is certainly one approach (one taken by *LoseIt*, for instance), one could ensure adequacy of coverage with much coarser categories as well. For instance, a tracking tool for physical activity could achieve completeness of coverage with only five categories — walking, cardiovascular exercise, strength training, flexibility training, and other. To log an activity, the user could select a category and specify duration and potentially intensity. The application could even support the user adding details to that activity

(e.g., labeling the type of “Other” activity as “scrubbing the floor” or “housework” — whichever label she chooses). The data would lack the precision needed to calculate the exact calorie burn or to see increases in bench-pressing capacity, but it may still enable the user to see patterns over time, or how the amount of physical activity relates to other behaviors such as sleep or eating, for example. Similar considerations apply to tracking diet, mood, and other types of behavior. The main point is that the design of the tracking feature should ensure that there are no large gaps in coverage that would frustrate the user, not that tracking has to be done at a very fine-grained level of detail.

### 2.3.2 Units

Health-related activities can often be tracked in a number of different units. Food intake might be tracked in terms of calories, grams of protein, fat, and carbohydrates; the number of servings of grains, meats, dairy, and other food groups; or even sophisticated ingredient-based schemes such as the one used by *POND* [3]. Physical activity can be tracked in terms of duration, intensity, metabolic equivalents (METs), calories burned, or, as *Nike+* does it, in terms of an application-specific unit called “Nike Fuel” that detected activities get translated into. Which units should an application use?

The decision is a matter of tradeoffs. Some units, such as calories, require much more granular data collection than others (e.g., the number of servings of different food groups), but can enable an application to guide behavior in a more detailed way than the use of a coarser unit would be able to do.

Units also vary in the ease of user comprehension. Tracking exercise in minutes may be an easier unit for most users to grasp than METs, and the number of grams of carbohydrates tends to be easier to grasp than glycemic index. For some applications, though, the harder-to-understand unit might be the right choice if its use, once it is learned, helps better guide user behavior. For people with pre-diabetes or type 2 diabetes, for example, sticking to foods with a low glycemic index can be more effective for managing their sugar than focusing on the total amount of carbohydrates [43]. For an application targeting

this population, glycemic index might be a good unit to use, even if its use involves a more difficult initial learning curve.

Finally, units vary in how easy they are to work with. For example, the number of servings, a common unit in food tracking, is easy for users to grasp conceptually but can be notoriously difficult to estimate. In a recent study, even patients with type 1 diabetes — a group that has to be extremely conscientious about tracking what they eat so they can effectively adjust their insulin dosing — expressed that they had difficulties estimating their food intake in terms of serving sizes [13].

The right decision regarding which units to use will vary from application to application. The choice of units is an important consideration, however, that should not be overlooked. What seems like an obvious choice — the use of calories for food tracking, for example — might not be an optimal one given the application's goals and its target users.

### **2.3.3 Ensuring Consistency of Manual Data Entry**

A key challenge with the use of manual tracking, whether of food, activity, or other health-related behaviors, is to ensure that users do it regularly enough to experience the benefits of the application. An obvious first step for ensuring the regularity of data entry — albeit one that is forgotten more often than one would expect — is to make sure that users actually have some benefit from the data they are entering. This issue has occurred repeatedly in medical applications that use mobile phones to collect data from patients but which don't provide any feedback (or provide only minimal feedback) to patients about the data they just entered (e.g., [4]). A more subtle form of the problem can also occur in wellness applications, though. For instance, in many self-monitoring applications, the bulk of the benefit of tracking occurs only when there is enough data in the system to be able to see trends and relationships among different forms of data. Yet, getting to that point can require a lot of data entry on the part of the user. For such applications, providing other forms of immediate benefits can be key to sustaining engagement with the application and the continuity of data collection needed for the application's larger benefits to come into view.



Another strategy that can help, if designed well, is to use reminders. Although they run the risk of being perceived as annoying [4], our research has shown that reminders can be an effective and unobtrusive strategy for increasing adherence to manual journaling. *UbiFit* used a reminder that (a) only appeared if the user had not logged any activities for about two days, (b) was unobtrusive, and (c) was gently positive. If the users logged their activities regularly, they would never see the reminder. It only appeared if no activities were added to the user's log for about two days. Even then, the reminder did not ring or vibrate; it just quietly appeared on the phone, waiting to be seen the next time the user looked at her phone. Finally, the text of the reminder was carefully designed. Rather than nag them to log more regularly, the reminder asked users whether they had done any activities they wanted to add to their journal — implicitly communicating that the system believes that the user is being active and just wants to make sure that the information is recorded and the user's log is kept accurate and up to date. Our participants were very positive about this reminder method and expressed that they often logged activities in response to seeing the reminder.

In a recent study, Bentley and Tollmar [9] used an even simpler, but effective, strategy. To encourage users of the application *Health Mashup* to journal their food intake, Bentley and Tollmar used a small icon that appeared in the phone's notification bar every evening. Like *UbiFit's* reminder, this reminder was also quiet. It did not vibrate or beep, but it just waited for the user to see it the next time she used her phone. By tapping on that small icon, the user could journal the day's diet directly from the notification, without needing to go to the trouble of opening the *Health Mashup* application itself. Bentley and Tollmar report that as a result of introducing this reminder system, the adherence to manual tracking in the application increased five-fold. This is particularly impressive given that even the initial version of the application contained a home screen widget that the user saw whenever she used her phone, which should have acted as a passive reminder in its own right. In spite of the widget, the adherence to tracking in the initial version was low, however.

These experiences suggest that non-intrusive reminders can support adherence to manual data entry provided that they are designed well. Making the reminder contingent on user behavior, so users are not reminded to do something they have already done, making it possible for journaling to be done directly from the reminder, and keeping the reminder positive, so that users don't feel like they are being nagged, can help create low-burden reminders that are both appreciated by users and are effective at promoting data collection.

## **2.4 Open Questions for Collecting Behavioral Data**

Given that this is such a challenging area to get right — that is, finding the right balance between collecting robust, accurate data and burden on the user — many open questions remain about how to best collect behavioral data in mobile wellness applications. In the following, we suggest at least four areas that would benefit from additional investigation: improving the experience and accuracy of manually tracked data, handling inaccuracy in sensed data, supporting reactivity of self-monitoring from sensed data, and controlling collected data.

### **2.4.1 Improving the Experience and Accuracy of Manually Tracked Data**

As we mentioned earlier, people often have difficulty understanding important aspects of the foods they consume (e.g., portion size, glycemic index, whether the foods contain certain ingredients) and the physical activities they perform (especially the intensity of activities), leading to potential inaccuracies in manually tracked data. While even inaccurate manual tracking can still support reactivity of self-monitoring [55], the inaccurate data that results from such tracking can interfere with other functions of an application, such as calculations of caloric intake and/or output. Further investigations are needed to understand how to best support users to develop skills for assessing their health-related activities when accurate and detailed manual tracking is required.

Another challenge for manual behavior tracking that would benefit from further investigation is in regard to *when* the user can enter

data about her food intake or physical activities. For example, should an application support the food or activity to be logged *before* it is consumed or performed or only *after*? If after, for how long after? For instance, research has shown that having to photograph the meal users are about to consume *before* they consume it can encourage reflection before eating, and in some cases, has actually helped users make better choices about what to eat [100]. At the same time, tracking before an activity is performed can lead to inaccurate data if the person ends up not performing the activity she intended to do (or she only performs partially). Future research should examine the tradeoffs of different forms of manual tracking and how to optimize the process to decrease user burden, and maximize data accuracy and reactivity of self-monitoring.

#### 2.4.2 Handling Inaccuracy in Sensed Data

As we mentioned earlier, using computer vision to automatically classify food from images has shown some promise, but is not ready yet for mainstream use. In their research prototype, Kitamura et al. [49] deal with the low accuracy rate by providing users with the ability to correct the estimates made by the system. We had good luck using this type of approach — that is, allowing the user to correct errors or omissions made by the sensing — in our work with *UbiFit*. However, with *UbiFit's* activity inference, errors were infrequent compared to the errors made in food detection by current computer vision algorithms. At the currently high error rate of automatic classification, it is unclear whether the user experience is acceptable for any sort of extended voluntary use. As such, until the algorithms improve, investigations could focus on how best to handle the inaccuracies in the sensed data.

#### 2.4.3 Supporting Reflection From Sensed Data

A more general open question remains for technologies that focus on automatic detection of the target behavior — that is, how does the automatic detection affect reactivity of self-monitoring and the ability of users to learn from their data? Recent work by Mamykina et al. [62] suggests that people have a more difficult time evaluating healthfulness

of images of foods that have already been tagged by other people (e.g., with the names of foods on the pictured plate) than when the images come with no attached description. Although such effects can probably be mitigated through careful design of user interactions in the application, the effects of increased user passivity from automatic estimates of food intake and physical activity need to be examined more closely. Put another way, if users aren't actively participating in the tracking of the target behavior, how can the technologies be designed to ensure that users will benefit from reactivity of self-monitoring that behavioral tracking has historically provided?

#### 2.4.4 Control of Collected Data

As an increasing number of wellness applications are social or include features to support sharing with others, a question facing designers and developers is if and how the collected data should be shared. In our work, we have found that the question of sharing behavioral data is a complex one. Not only do people think of different types of data as being more or less sensitive,<sup>23</sup> but sharing of the same kind of data can be more or less comfortable depending on the circumstances. For example, users of *Houston* were happy to share their step counts with their small group of fitness buddies on days when they were being active, but were more hesitant to share when they were being inactive [19]. Similarly, in our work on technology for patients with breast cancer, we found that cancer patients wanted a lot of control over how the metrics they were tracking (their symptoms and wellbeing parameters such as energy level) were shared even with their closest family members [unpublished results]. While they were okay with sharing this information most of the time, sometimes they wished to keep certain records private, often in order not to worry their partners and other loved ones. To further complicate matters, even seemingly innocuous data, such as breathing patterns, can be used to infer sensitive information (e.g., drug use) [77], making automatic sharing of sensed data

---

<sup>23</sup>For example, as part of our 3-month study of *UbiFit*, we found that people tend to be relatively unconcerned about their accelerometer data but perceive the GPS data captured by their phones as being far more sensitive [50].

problematic — especially because users may not be aware of or consider this type of risk. These and other issues are likely to be raised when applications support the sharing or storage of behavioral data with application developers, service providers, the user’s employer, the user’s health insurance provider, and so on.

For such reasons, the data sharing policy needs to be carefully considered. To minimize risks to the user’s privacy, for sensed data, it might be better to only keep high level inferences and not the raw data itself, to avoid unintended exposure of sensitive information. For high-level behavioral data, we believe that users need to maintain full control over their data and should be able to decide on a per-record basis if and how the data is shared. Although the sharing functionality will clearly depend on the nature of the application, when possible, we suggest that the data be kept private by default but that the users be given easily accessible and understandable ways of sharing the data when they so desire. In whichever way it’s implemented, however, making both the sharing controls and the current state of the data transparent to the user is crucial for the application’s credibility and for maintaining user trust.

## **2.5 Section 2 Wrap-Up**

In this section, we discussed current approaches to manually and automatically tracking two behaviors commonly targeted by mobile wellness applications today — food intake and physical activity. We also discussed general design issues related to the collection of behavioral data and proposed some open questions for HCI researchers in the area of collecting behavioral data including improving the experience and accuracy of manually tracked data, handling inaccuracy in sensed data, supporting reactivity of self-monitoring from sensed data, and controlling collected data.

# 3

---

## Providing Self-Monitoring Feedback

---

As we discussed in Section 2, many wellness applications help users record information about their health behaviors, such as physical activity and diet. This act of recording one’s own behavior is called *self-monitoring*, and it is an effective technique for supporting health behavior change both because it helps the user change her behavior directly and because it generates data needed for other techniques, such as *goal-setting* [54, 68], which we discuss in Section 4. While the use of technology to facilitate self-monitoring has a long history — for example, mechanical counters and paper diaries have been used for decades to help people track their behaviors — mobile applications and unobtrusive, wearable sensors have drastically reduced the burden of tracking, enabling people to record their behaviors more easily, more accurately, and over longer periods of time.

What makes mobile technology particularly well suited for supporting self-monitoring, however, is the variety of feedback that these tools can provide based on the data that users record. Even simple forms of feedback — such as seeing the number of steps that one has to take to meet one’s goal — can increase the efficacy of self-monitoring for changing behavior [45]. But mobile technologies can go well beyond such

simple feedback. Many wellness applications present their users with sophisticated graphs of the patterns in their behavior over time; some, like *UbiFit* and *BeWell* [56], use stylized representations that users see every time they use their phones; and some, like *Health Mashup* [87], even try to interpret the user's data and provide her with explicit insights about her behavior — for example, that she is more physically active on days when she has gotten more sleep the night before.

In this section, we provide a brief review of the ways that mobile wellness systems provide self-monitoring feedback based on users' behavioral data. We discuss the forms that such feedback takes and where it is presented, as well as what we see as the strengths and limitations of each approach. What will be clear from our discussion is that while wellness applications have used several different approaches to providing feedback, there has been little systematic investigation of the effectiveness of those feedback mechanisms especially compared to other feedback mechanisms. We know that having feedback helps; but we still know little about what kinds of feedback work best — not only for whom, but also for what forms of wellness behaviors and metrics different types of feedback work best and why. We close this section by suggesting open questions in designing self-monitoring feedback for mobile technologies that support health and wellness.

### 3.1 Forms of Feedback

Mobile wellness systems have used four main forms of feedback: counts, graphs, stylized representations, and narrative information. Counts and graphs are by far the most common approaches, although both stylized representations and narrative information can play important roles in wellness tools. We describe these forms of providing feedback in turn.

#### 3.1.1 Counts

The simplest form of self-monitoring feedback is to present the user with counts of the activities that she is tracking: the number of steps that the user has taken, the number of calories or the number of servings of fruit and vegetables that she ate, the number of hours she slept the night before, and so on. Among the wellness applications that we

have reviewed, every application that included a self-monitoring component also provided users with feedback about the tracked behaviors in the form of counts. For applications that include manual journaling of exercise or food categories, such as *Mobile Lifestyle Coach* [35] or *Few Touch* [5], such feedback amplifies the increased awareness from noticing and recording the behaviors in the first place. For applications that include automatic recording of behavior, such as *Fitbit* or *Jawbone UP*, such feedback takes the place of recording as a chief way in which the salience of the tracked behavior is increased. In addition, for frequent and low-salience activities, such as walking or sitting, without explicit feedback about how much the user walked or sat, she would have a very difficult time developing an accurate sense of the frequency and duration of such activities. The same reason these activities are so difficult to record manually also makes them difficult to estimate accurately. Even simple feedback in the form of the number of steps taken or the amount of time the user spent sitting, provides much more awareness of these activities than the user could have without such feedback.<sup>1</sup>

#### 3.1.1.1 Main advantages

Simple counts are the least abstract type of information about tracked activities. Their chief strengths are that they capture a very important dimension of the user's behavior — that is, the amount — and their basic interpretation is straightforward. There is little to misunderstand about “7236 steps” or “3 servings of fruits and/or vegetables.” Seeing counts of tracked activities increases users' awareness of their activities and it helps to “keep them honest” — it becomes more difficult to delude oneself that one is being active if *Fitbit* is consistently showing 2,000 to 3,000 steps per day. And, of course, the counts of tracked behaviors are the basic building blocks of other types of feedback about users' activities. Patterns over time, comparisons with other people, assessments of how close one is to one's goals, are usually based on the

---

<sup>1</sup>In fact, in our studies of both *Houston* and *UbiFit*, the participants were consistently surprised with how inaccurate their estimates were of how much activity they had in their day-to-day lives.



counts of user's behavior from day to day. This is why counts are such an essential part of self-monitoring applications.

### 3.1.1.2 Main disadvantages

The main limitation of counts is that their meaning is ambiguous without further information. Knowing that one has walked 5,000 steps means little without knowing how much one should be walking, if this is more or less than the person walked on previous days, or how much other people with similar health goals tend to walk. It is only in context that counts are meaningful for the user, and the more contextual information the user has, the better she is likely to be able to make sense of the counts.

Consider the simple step count again — one of the most common metrics offered today by applications that track physical activity. Today's popular commercial applications that track steps — *Fitbit*, *Jawbone UP*, *Nike+Fuelband*, *Misfit Shine*, *Basis*, and so on — give meaning to the user's daily step count by recommending a default goal of 10,000 steps per day that originated in Japanese walking clubs of the 1960s, but which has since been widely adopted as a goal by the public-health community [92]. It turns out that for many people with office jobs, this number of steps can be difficult to reach on a consistent basis. Given that, how should a user interpret her common daily counts of, say, 4,000 to 6,000 steps per day? What should the person make of this number? Other than knowing that she is doing about half of the recommended amount of walking, it can be difficult for the user to know what it means for her health. Is reaching 10,000 really necessary? What's the minimum number of steps per day the user needs to do to start losing weight, given her current diet? To see health benefits? Does it matter from what activities the steps come? For example, are steps walked up hill better than, worse than, or the same as steps walked on a level surface? On its own, the count doesn't answer questions like these. Yet, for the user to be able to effectively use the counts that, say, her *Jawbone UP* shows her, she needs such additional information. Applications that mostly limit themselves to counts and historical trends of counts provide users with less information than is needed to make

informed decisions about how to change behavior. The counts are not useless, but they are a lot less useful on their own than they are when they are contextualized within richer forms of feedback.

Another limitation of counts is that they can provide people with a false sense of knowledge and comfort. Many wellness applications use some basic information about the user — age, height, weight, and gender — to try to estimate the number of calories that the user is burning from day to day. Such calculations are notoriously inaccurate, although that is rarely made very visible within the application itself. Consider, for instance, the feedback on a recent day from the second author's (PK) *Misfit Shine* (Figure 3.1). According to the *Shine*, PK made 9,292 steps and burned 3,200 calories. While even the step count is probably too high for that particular day, the calorie expenditure count is almost certainly way too high. If PK consumed 3,000 calories a day, and was about as active as he was that day, he would inevitably gain weight (PK knows this from experience). Furthermore, looking through the data, it is clear that the application's calorie calculations are not even internally consistent over time. On the following day, the *Shine* reported that PK took 12,438 steps but only burned 3,066 calories. And on the day after that, PK apparently took 14,144 steps and burned 3,202 calories. So, within three days, the application reported roughly the same calorie expenditure for two different daily step counts, of which one was over 50% higher than the other. If PK were using this data to determine his caloric intake budget — which he could reasonably do, given how the information about his caloric burn is presented — he would quickly find himself in trouble on the scale without understanding why.

There is nothing in how this information is presented in the *Shine* application that suggests that the calories-burned counts could be inaccurate. The feedback presents with the kind of definitiveness that is just not warranted by the underlying algorithms and information on which the count is based. Nor is *Shine* the only culprit in this regard. In every application we've seen that reports energy expenditure based on similarly limited data (e.g., weight and height), calories burned are presented in equally absolute terms, without any information about the confidence for the calculations on which they are based.

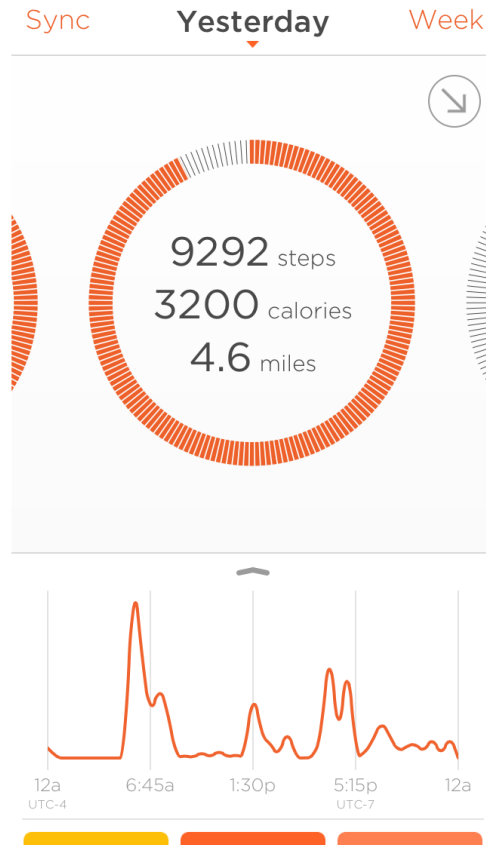


Fig. 3.1 **Calories burned view in *Misfit Shine*.** *Misfit Shine* provides no visual indication about the potential inaccuracy of its calorie-expenditure calculations.

The concreteness of numbers has a potential to deceive. Careful design is needed to offset this risk.

### 3.1.1.3 Summary

Counts are one of the most commonly used and most valuable types of feedback that self-monitoring applications can provide. Yet, to be truly useful, counts need to be contextualized so that the user can use them effectively to make decisions about her behavior, and the counts need to be presented in a way that does not provide a false sense of precision where that precision is not warranted.

### 3.1.2 Graphs

Besides counts, the most common form of feedback that we have seen used to support self-monitoring in mobile wellness applications is graphs. Based on the data from tracked activities, graphs aim to help people understand how their activities change over time, how they are progressing toward their goals, and how different types of activities that the user is tracking might relate to one another.

The most common function of graphs in mobile wellness applications is to display users' tracked activities over time, typically using line or bar charts. The temporal resolution of such graphs varies, from short time frames such as a single day (e.g., see the lower portion of Figure 3.1) to high-level views of tracked activities over weeks or months. Such representations enable users to understand when and how much they are doing their health behaviors, and in what direction activity levels are shifting over time. For instance, the top portion of the graph in Figure 3.2 shows that over the course of this week, the user averaged around six hours of sleep a night and only had one night with eight hours of sleep. A longer-term view, such as a monthly or yearly view, could show how typical this pattern is and whether the user is succeeding in coming closer to being able to consistently get the recommended eight hours of sleep per night.

Another important function of graphs is to support reflection on the possible relationships among different types of activities and metrics. *Jawbone UP*, for instance, enables users to view graphs of multiple metrics that they are tracking aligned by time. Figure 3.2 shows the daily view of a graph that contains both step count and sleep data. Such graphs can help a user determine if the amount of sleep seems to be related to how much physical activity she does the next day, enabling her to create and test hypotheses about the relationship. In our work on a mobile application for patients with breast cancer, we found that such combined graphs can be very helpful for investigating what factors may be influencing one's behavior or states such as mood or fatigue [52].

Finally, graphs can provide easy-to-understand feedback about goal progress. While goal progress can be shown in purely numerical



Fig. 3.2 *Jawbone UP's data trends display.* *Jawbone UP* provides users with a way to examine trends in their data over time, including by examining potential relationships of multiple tracked behaviors — in this example, sleep and step count.

terms — for example, in terms of percentage of the goal reached — graphical representations can be easier to grasp at a glance, making them a good match for glanceable displays, device displays, or as always-present interface elements within an application that the user is unlikely to spend a lot of time examining. For instance, *GoalPost* provides bar graphs of goal progress for each component of the user's weekly activity goals on the application's main screen, providing an at-a-glance view of goal progress every time the user opens the application. Additional information about goal attainment is provided in the form of line charts on a separate *Goal* screen (see Figures 1.3(a) and 1.3(b)).

### 3.1.2.1 Main advantages

Graphs often do a good job of helping users understand patterns in their data. The diversity of available graph types makes it possible to communicate a great deal of information in a compact representation that fits the display constraints of a mobile device. In addition, graphs can be made interactive, allowing users to access additional information by touching points on the graph or by drilling down. Such interactivity enables graphs to scale elegantly as the amount of data grows and the user continues to use the application over an extended period of time. These characteristics make graphs one of the most flexible forms of self-monitoring feedback.

### 3.1.2.2 Main disadvantages

Although it might seem obvious, it's worth keeping in mind that graphs are only as good as the underlying data. As we discussed in Section 2, few applications that use sensing for tracking behavioral data allow users to correct data that has been incorrectly detected or to manually enter information for periods when the sensing device was not used. Activity recognition is not perfect, however, nor do people always have their sensing devices with them. Even if she tries to use it regularly, a person might forget to move her *Fitbit* from one pair of pants to another or the device might run out of battery. Similarly, people's estimates of their activities or food intake can be poor, so manually tracked data can be inaccurate as well. In such cases, the data in the application will be incorrect and this will be reflected in the resulting graphs (something the authors have personally experienced). Missing data is a particularly big issue for self-monitoring feedback when graphs are used. Since the application doesn't know whether the data is missing or the activities were not performed, it usually assigns values of zero to the periods with missing data. Once graphed, such gaps can obscure trends and potential relationships that the graphs are designed to reveal. While the same problem applies to simple counts, the graphs' function of displaying trends amplifies the effect of missing or incorrect data. Providing ways for users to tag or otherwise indicate incorrect or missing data

could help an application generate more accurate graphs, increasing the likelihood that they would help users develop useful insights into their behavior.

A second limitation relates to the use of graphs to support discovery of relationships among multiple behaviors or metrics. While graphs can be very useful for this purpose, graphs in typical wellness applications can only support relationship discovery by helping users generate hypotheses about possible relationships — hypotheses that users then need to test by making changes to their behavior and seeing what happens. Graphs themselves provide no information about the direction of a relationship (is more sleep contributing to more physical activity or vice versa) or even if the relationship is real at all or just apparent. Yet, the concreteness of the representation can obscure this fact, especially for users with lower levels of scientific training. For these reasons, applications that use graphical representations of multiple behaviors might try to couch such graphs within a module that guides users through formulating hypotheses and helping them come up with ways to test their hypotheses, rather than just providing graphs by themselves and leaving the users to make of them what they might.

Finally, numeracy levels in the general population are quite low — nearly half of the population has problems with understanding even basic numerical concepts, such as percentages and ratios [71] — as are levels of graphical literacy — the ability to understand the meaning of graphically presented information [34]. These findings raise the question of just how effective graphs are as the main form of self-monitoring feedback for broad populations. We are not aware of any systematic research that examines this issue in the context of wellness applications, but the literature on numeracy and graphical literacy at least suggests that more complex graphs should be tested for understandability before being incorporated into mobile wellness tools.

### **3.1.2.3 Summary**

Graphs are a powerful tool for exploring trends and relationships in tracked wellness data. They can represent a great deal of information

in a compact form, they scale well for large data sets, and they can be made interactive, further supporting users' ability to understand their behavior. The usefulness of graphs is likely to be constrained by users' numeracy and graphical literacy levels and their ability to accurately understand what graphs are able to tell them about the relationships that appear to be present in the data. The quality of the underlying data that the graphs use is also critical for their effectiveness at supporting wellness behaviors.

### 3.1.3 Stylized Representations

In addition to graphs and counts, wellness activities and metrics can also be presented in abstract, stylized representations that map tracked data to images or animations. This is precisely what we did with *UbiFit's* glanceable garden display. As we discussed in the *Introduction*, *UbiFit* maps different types of physical activity — walking, cardiovascular exercise, strength training, and flexibility training — to different types of flowers that appear in a garden over the course of a week. Additionally, butterflies represent weekly goal attainment for the current and the previous three weeks. Just by glancing at the garden display, the user can see how active she has been, how diverse her activity has been, and whether she reached her weekly goal for any week in the past month.

Lane et al.'s *BeWell* system [56] takes a similar approach to self-monitoring feedback. Unlike *UbiFit*, which only focused on physical activity, *BeWell* provides stylized feedback on three different wellness activities: sleep, physical activity, and social interactions. *BeWell* uses sensors on the mobile phone to detect users' levels of these activities and then represents that data in the form of an aquatic ecosystem on the phone's live wallpaper. *BeWell* maps the data about sleep to the movements of a turtle, which sleeps in the animation when the user is not getting sufficient sleep. Physical activity is mapped to the movements of a clown fish, which becomes more animated and playful as the user becomes more active. Finally, social interactions are mapped to the size of a school of small yellow fish that swim across the screen. As with *UbiFit*, a quick glance at the screen is enough to provide a



rough sense of how well the user has been doing on the three activities tracked by *BeWell*.

While both *UbiFit* and *BeWell* use stylized displays on the background screen of a mobile phone, this is certainly not the only place where such representations can be used. In fact, as we note shortly, one of the advantages of this form of feedback is that it can expand the range of locations where self-monitoring feedback is provided.

### 3.1.3.1 Main advantages

Stylized representations such as those used by *UbiFit* and *BeWell* have a number of strengths. First, by mapping health-related activities to an image from a completely different domain, stylized representations can support user privacy, enabling their use in places where people other than the user could potentially see them, without revealing that the user is tracking health behaviors. Phone wallpapers and lock screens, digital picture frames, widgets on large-screen displays are just some of the locations where stylized representations can be used to provide self-monitoring feedback but where a more literal representation of health behaviors might not be feasible due to privacy concerns. This could be particularly important for health behaviors and metrics that users can perceive as being sensitive, such as mood, weight, or medication adherence (e.g., imagine if the user is tracking adherence to her psychiatric or HIV medications).

Second, stylized representations can be made to be attractive, increasing the probability that a user would be willing to use them in a highly visible place, such as her phone's lock screen. In this way, stylized representations can help increase how often the person sees self-monitoring feedback, potentially increasing the effectiveness of self-monitoring itself.

Finally, stylized representations can be made to support themes, allowing users to personalize how they receive feedback and to change the theme when they tire of the previous one. This ability to personalize and change up the feedback might increase user engagement with a wellbeing application and help users maintain interest in using the system over longer periods of time.

### 3.1.3.2 Main disadvantages

Since the use of stylized representations to support self-monitoring is very recent, not much data is available about their various characteristics. One potential limitation has to do with their learnability. Since wellness behaviors are mapped to representations from a different domain, users need to learn this mapping to be able to interpret the feedback. Though *UbiFit*'s garden display was successful in this regard in all of our evaluations of it, how many different metrics can be mapped and still maintain easy learnability is an open question, as is whether certain types of representations allow for easier learning of mapping than others.

By their nature, stylized representations are also less precise than graphs, and they scale less well when there is a lot of data. A new flower in the *UbiFit* garden tells the user that an activity of a certain type has occurred, but the flower doesn't encode information about the duration of the activity or at what time it was performed. Similarly, while a new flower is easy to notice when the garden only has a few other flowers, when the garden is full, a new flower becomes much harder to detect and at some point the garden can run out of room for new flowers (though over the two or three years that *UbiFit* was in use, no participant's, research team member's, or pilot tester's garden ever came close to meeting the display's maximum number of flowers). Depending on the activities that are being tracked, such scaling and precision issues will need to be kept in mind when a stylized representation is being designed.

### 3.1.3.3 Summary

By providing feedback in a privacy-preserving and aesthetic way, stylized representations can extend where and how self-monitoring feedback can be provided. A lot still needs to be learned, however, about how to make such representations easily learnable and how they can scale to accommodate different types and amounts of wellness data.

### 3.1.4 Textual Feedback

Some wellness applications have begun to use textual messages to give users feedback about their data. *Health Mashup*, for instance, monitors

a broad range of wellness behaviors and then runs collected data every night through a set of algorithms that look for significant correlations among different types of data. If a correlation is found, the user is presented with messages — located in a widget on the phone’s home screen — that explain the found correlations in everyday language. For example, a user might be told that “*you lose weight on days you have more scheduled time*” or that “*you walk 80% further on weekends vs. weekdays (15000 vs. 8300 steps)*” [87]. *Health Mashup* also tries to characterize the significance of the found correlations by following their descriptions with a plain-text indication of confidence — marking the message with indicators such as “possibly” or “very likely.” This form of textual feedback serves a similar function as graphs, but it avoids the risk that the users will perceive a pattern in their data that is not actually there or that they will misunderstand the data on the graphs.

*Jawbone UP* uses a slightly different form of textual feedback with its “insight engine.” Based on tracked data, *UP* pops up textual messages that recommend small changes that the user can make in her routine to be healthier or provides descriptions that can help the user make more sense of the data. *UP* might say, for instance, “*you walked 8000 steps,*” followed by “*equivalent to walking across the Golden Gate Bridge and back.*” With such insights, *UP* provides additional context for the data and helps the user generate ideas for concrete actions that can lead to a healthier lifestyle that the user might not have come up with by herself.

An even simpler form of textual feedback is messages that some wellness applications use to notify users of substantial changes in their activities or to provide feedback on goal progress. *Misfit Shine*, for instance, creates “highlights” to explicitly acknowledge goal attainment or achievement of personal best activity levels, such as “*you outdid your previous record by 255 points.*” *Fitbit* and *Nike+* use a similar strategy as well, mostly as a way of reinforcing good performance. In *Houston*, we used textual messages to provide clear feedback about goal progress. Whenever a user entered a step count from her pedometer, *Houston* would display a message telling the user how many more steps she had to go to reach her goal or a congratulations message telling her how far over her goal she was. The message made goal progress more salient

and enabled the user to get this important information without having to search for it in the application or perform a calculation.

#### **3.1.4.1 Main advantages**

The chief strength of textual feedback is that it can provide people with information about their data that they might not have discovered themselves. Machine learning algorithms can detect patterns in the data that would be too difficult to spot unassisted. Textual feedback provides a way of presenting such patterns to users, as well as providing suggestions for concrete actions that the user might want to make given those patterns.

By providing feedback in everyday language, textual feedback can also overcome the problem of giving self-monitoring feedback to users with low numeracy. As such, this form of feedback might be particularly helpful for systems targeting populations where numeracy and graphical literacy are likely to be a widespread issue, such as low-education and low-socio-economicstatus groups [71].

Finally, textual feedback can make changes in the tracked activities more salient either by acknowledging achievements or by drawing the user's attention to the fact that she is slipping. *Shine*, for instance, creates textual messages not only for achievements but also for decreases in activity, with messages such as “34% less active than last week.” While graphs can certainly show that one's activity levels have gone down, such an unambiguous statement can make the point more strongly and with less room for self-delusion.

#### **3.1.4.2 Main disadvantages and summary**

While text messaging (SMS) has been used to promote healthy behaviors for a long time, the use of textual self-monitoring feedback in mobile wellness applications is relatively recent and its limitations are not yet apparent. Framing, frequency, length, how uncertainty is presented and other such properties likely affect the effectiveness and acceptability of textual feedback, as they affect other forms of health communication. We are not aware of any research that has investigated these issues in the context of in-application self-monitoring feedback for health behav-

iors, however. More research is needed to determine optimal forms of textual feedback for this domain and medium.

## 3.2 Location of Feedback

In addition to taking different forms, self-monitoring feedback can also be provided in different locations. In this section, we review four locations where such feedback is commonly provided by mobile wellness applications: within the application itself; on a glanceable display on the device, such as on a lock screen or as a home screen widget; on a sensing device used to detect health activities; and on a Web site (or another system) connected to the wellness application.

### 3.2.1 In-Application Feedback

The most common place for providing feedback about tracked data is within the wellness application itself. All wellness applications we have reviewed or developed ourselves include, at a minimum, in-application feedback. *Houston*, for instance, provides users with views of their daily and weekly steps counts, goal progress, and comments about tracked step counts that the user received from her fitness buddies. Commercial applications like *Fitbit* and *LoseIt* also offer rich in-application feedback, including information about the user's current activities, food intake and caloric balance, and progress toward activity and calorie-balance goals.

In-application feedback is highly flexible. An application can provide multiple kinds of feedback, each tailored to fulfill a different function. For example, *UbiFit* contains two key types of feedback representations: a journal with details of the user's activities, and a view of goal progress that indicates all activities that count toward the user's weekly activity goal. Similarly, *GoalPost* contains an activity journal, a goal-progress display, and a view of the rewards that the user has received for performing physical activities and reaching her goals (the trophy case screen, shown in Figure 1.3(c)). *Jawbone UP* uses an even larger number of feedback representations, including custom views that users can create to see daily, weekly, and monthly

views of multiple behaviors and states that they are tracking, and the “lifeline” — a summary view of all tracked behaviors.

### 3.2.1.1 Main advantages

A key advantage of providing feedback inside an application is that feedback representations can be highly sophisticated. In-application feedback can take advantage of all features of a mobile phone, enabling the use of multi-touch gestures to pan and zoom, the use of drilling down or fly-outs to get more information, or superimposing tracked data on other information. The iPhone application *Moves*,<sup>2</sup> for instance, provides feedback about physical activity based on the locations where the activity took place. The user is able to drill down into the location to see it on a map, and even to create custom names for common locations, such as *home* or *office*.

Another advantage is the ability to have multiple types of feedback. An application can be arbitrarily complex, accommodating any type of feedback that the designers determine is needed. In particular, views of historical data — which can be voluminous if the user has been using an application for a while — are often served well with in-application representations.

### 3.2.1.2 Main disadvantages

The key disadvantage of in-application feedback is that the user has to remember to go to the application to be able to see the feedback. While this is typically not a problem when the user first starts using an application, after the novelty wears off or when the user gets busy with other things, the frequency of application use can go down substantially. After a while, the user can forget about the application altogether, effectively abandoning it. In such cases, in-application feedback stops providing help for encouraging and supporting health behaviors.

The problem of forgetting can potentially be mitigated by the use of notifications that remind the user to keep using the application. For this to work long-term, though, notifications have to be extremely

---

<sup>2</sup><http://moves-app.com> {Link verified on Sept 2, 2013}

well designed. “Alert fatigue” [10] is a known problem across a range of systems. People can quickly habituate to notifications and begin to ignore them. When that happens, little is accomplished in terms of supporting engagement with the application, and the burden of exposing the user to a constant stream of notifications may lead her to abandon the application even more quickly.

### 3.2.1.3 Summary

In-application feedback is the most common type of feedback used by mobile wellness applications. Such feedback can be rich, flexible, and highly interactive. However, to be effective, users need to keep coming back to the application, which is an important challenge that designers have to address.

## 3.2.2 Glanceable Displays

In addition to notifications, another way to deal with the downside of in-application feedback is to provide additional feedback that is visible outside of the application, in a location that users will frequently see whether they go to the application or not. This is precisely what *UbiFit* did with its garden display. *UbiFit*'s garden display was implemented as the phone's wallpaper, which meant that users would see it whenever they used the phone, whether to check the time, send a text message, see upcoming appointments, or use the phone for any other purpose. As the garden display was updated any time a new activity was detected or journaled, users always saw up-to-date feedback about their activities that week.

*UbiFit*'s garden display is an example of a *glanceable display*, a variation on the concept of the ambient display. In *UbiFit*'s case, the glanceable display takes over a prominent part of the phone interface (like the wallpaper), so it can be seen every time the phone is used. On Android phones, a glanceable display can be implemented as a live wallpaper or a phone-screen widget, both of which are visible whenever the user returns to the phone's home screen. *BeWell* [56] was implemented as a live wallpaper as was the display for *ShutEye* — a mobile application that we designed to encourage good sleep hygiene [7], while some of

our current work on medication adherence uses the home-screen widget for this purpose. On iPhones, an application can provide feedback by changing the phone's lock screen, although currently this operation cannot happen automatically in the background (e.g., in response to sensor data), a limitation that iOS 7 is supposed to remove.

### **3.2.2.1 Main advantages**

The greatest strength of glanceable displays is precisely their high frequency of being seen. Although the mechanisms that mediate effects of glanceable displays have not been carefully studied, our hypothesis is that their frequent visibility primes the person's health goals [40], making the cognitive representations of these goals repeatedly active. Such goal activation has been shown, in turn, to trigger goal-pursuit behaviors, including noticing and taking advantage of opportunities for goal-directed actions [30]. If this is the case, it may partially explain why participants in the 3-month study of *UbiFit* who had the glanceable display were more active than those who used the same self-monitoring journal and fitness device but who did not have the glanceable display.

In addition to providing feedback about tracked data, glanceable displays can potentially also remind the user to use the wellness application itself. This might lead to higher levels of application use, which could help make the application more effective. However, this potential advantage has not yet been tested.

### **3.2.2.2 Main disadvantages**

One limitation of glanceable displays is that in comparison with in-application feedback, glanceable displays are very limited in terms of the user interactions that they support. The user cannot interact with a lock screen or wallpaper in any robust way, and interactions with an Android home-screen widget are limited to scrolling to see more content and tapping to enter the application. If an application wants to provide feedback that requires a higher level of interaction, a glanceable display is not a good option as the only feedback location.

Another limitation stems directly from the glanceable displays' greatest strength: their visibility. For example, when implemented on



the phone's lock screen or wallpaper, it is very likely that a glanceable display will sooner or later be seen by people other than the user. Glanceable displays thus present a privacy risk, especially if they are used to represent sensitive health behaviors, such as medication adherence. For this reason, we recommend that systems that employ glanceable displays use stylized representations that appear unconnected to health. While other forms of feedback can certainly be placed on a glanceable display, their acceptability by the target user group would need to be carefully researched before a more literal representation is seriously considered.

Finally, lock screens and home screens are very limited in terms of real estate. Only one application can control the lock screen or wallpaper at any given time. This means that the use of multiple applications that have implemented glanceable displays could be problematic. Widgets are less limited in this regard, but only slightly. Even with widgets, there is just so much room on the main home screen where a widget can be placed and still be visible. Relegating the widget to a second or third home screen may not provide much additional value over what is offered by in-application feedback. Given the limited real estate of always-visible parts of a phone's interface, users will need to make decisions about what information is valuable enough for them to want to always see on their devices. A glanceable display that is attractive, understandable, and provides timely feedback has a much better chance of swaying users to give up their photos of babies, kittens, or their beach vacation in order to use the display.

### 3.2.2.3 Summary

Glanceable displays are a promising new method for providing frequent self-monitoring feedback. When they are in a position to be seen often, glanceable displays can help users remain engaged with their health goals and with the wellness application. However, glanceable displays can also present privacy risks that should be addressed through careful design and by paying close attention to users' privacy concerns. Additionally, a user can have only so many glanceable displays at a given time.

### 3.2.3 Feedback on Sensing Devices

In addition to providing self-monitoring feedback on the phone, mobile wellness technologies that use a sensing device to detect health activities and metrics sometimes also provide feedback on the device itself. Digital scales, glucose meters, digital blood pressure cuffs all provide immediate readouts of the results. Increasingly, sensors for monitoring physical activity have taken this route as well. *Nike+ Fuel Band*, for instance, includes a simple display consisting of white LED lights that can be activated by pressing a button to let the user know how many steps she has taken, the number of “fuel points” (Nike’s activity metric) that she accumulated over the course of the day, and how close she is to her daily goal, as well as to provide an acknowledgement when the goal has been reached. The user switches among different pieces of information by repeatedly pressing the button on the band. Other sensors include similar, although less elaborate, feedback. *Fitbit One* and *Withings Pulse* provide the current daily step count, distance traveled, and the number of steps remaining to reach the daily goal. The *Pulse*, which can also measure heart rate, provides the heart rate information right after a measurement is taken as well. Both sensors provide feedback as numerical information on a small display built into the device. Finally, rather than providing the exact count of the user’s activity, *Misfit Shine* and *Fitbit Flex* use a set of white LEDs to indicate the proportion of the daily step goal that has been reached so far. *Flex* does this with only five LEDs — providing feedback in 20% increments — while *Shine* uses a set of 12 LEDs arranged in a circle to provide a little more granular feedback.

#### 3.2.3.1 Main advantages

The main advantage of on-device feedback is that the user can receive feedback without needing to check her phone. This is particularly useful for receiving feedback while a person is engaged in a high-intensity physical activity (e.g., running or biking), during which people often don’t have or can’t effectively manipulate their phones, or when a person wants to quickly see how much activity she has done without fumbling to unlock the phone and open the wellness application. When one

is first beginning to increase one's physical activity, such quick checks are very useful for developing a sense for distances and the number of steps that one gets from different parts of one's daily routine, for example. Users can miss the ability to get quick feedback when they use a device like *Jawbone UP* that does not include a way to provide feedback or when the available feedback is rather course-grained, such as the feedback on *Fitbit Flex*.

Another advantage of on-device feedback is that the feedback is always up to date. Many of the current sensing devices sync with the phone either manually (e.g., *Jawbone UP*, *Misfit Shine*), or periodically throughout the day (e.g., *Fitbit*, *Withings Pulse*). This means that the feedback on the phone can be out of date until the user forces a sync to refresh it, which adds time and complication to an operation that should be fast and painless. Viewing feedback on the device is often the fastest way to get up-to-date information about one's activities, allowing users to go to the phone only when they need additional information or insights.

### 3.2.3.2 Main disadvantages

The chief limitation of on-device feedback is that due to the size and the computational capabilities of sensing devices, the feedback they are able to provide is typically far simpler and more limited than the feedback that can be provided on a mobile device like the phone. Basic counts are the primary form of feedback supported by sensing devices, and feedback provided by some devices, like *Fitbit Flex*, is even more course-grained. That said, the main function of on-device feedback is to provide users with a quick way to check their current status and goal progress, and for this purpose the limited feedback that sensing devices can provide is often enough. We wonder, however, whether the feedback from devices like *Shine* and *Fitbit Flex* is too simple even for this purpose.<sup>3</sup> More research is needed to understand the minimum

---

<sup>3</sup>In our personal experiences, we have found that the very simple feedback provided by devices like the *Shine* and *Flex* isn't very useful, especially when compared to what devices like the *Fuelband* provide.

level of feedback that is needed to effectively support the quick, intraday checks for which users often look to the sensing device.

### **3.2.3.3 Summary**

While more limited than the feedback that can be provided on a phone, on-device feedback can be exceedingly useful for quick checks on activity and goal progress throughout the day. Devices that do not provide such feedback, like *Jawbone UP*, make this operation more complicated, which may increase user frustration and adversely affect their willingness to use the system long-term.

### **3.2.4 Feedback in Other Locations**

Finally, few of the modern wellness applications are standalone systems. Most, even if they initially started as purely mobile applications, such as *RunKeeper*, have evolved over time into cloud-based services that are capable of providing feedback to users in a number of locations. The most common location to provide such additional feedback is the application's Web site. The majority of popular commercial wellness applications have companion Web sites, and this is becoming more of a trend for research systems as well (e.g., Nokia's *Wellness Diary* syncs to a companion site *Wellness Diary Connected* [98]). Such Web sites can provide a variety of services, from just giving users a larger screen on which to view their graphs to incorporating additional forms of support for changing behavior. The *Nike+* Web site, for instance, has a well-developed gamification component that includes quests to conquer different parts of the world, all built on the users' fuel point scores. Social sharing of wellness tracking data and ability to create and join health-activity challenges are other common components of such Web sites.

#### **3.2.4.1 Main advantages**

The main advantage of using a Web site as an additional location for self-monitoring feedback is that a Web site makes it easier to examine

graphs, look for patterns in the data, and get a comprehensive picture of historical trends. While the computational capabilities of mobile phones have greatly increased, a 4" screen is still far more constrained than a 27" desktop monitor or even a 13" laptop display. A larger screen can be of great benefit for examining complex data. This is particularly obvious for applications that make heavy use of graphs as a form of feedback, such as *Fitbit*.

Cloud-based services are not limited to only providing feedback on Web sites, however. Although we are not aware of any current systems that do this, there is no reason why such services can't provide feedback in other places as well. Digital picture frames, wall-mounted displays, widgets on large-screen monitors are just some of the places where wellness applications could provide additional feedback to users. Taking advantage of some of these options would enable wellness applications to provide a broad range of customized feedback that is finely tailored to its mode of delivery, potentially substantially increasing the applications' effectiveness while keeping the user burden low. We hope that future research explores this possibility.

#### 3.2.4.2 Main disadvantages

The main drawback of self-monitoring feedback on Web sites or other devices is that providing such feedback requires additional development and design resources. For an organization or research team that is strapped for resources, it might make more sense to focus on providing a best-of-breed mobile experience than to dilute their efforts across multiple platforms and end up with several mediocre products. In addition, if a Web site does not provide any clear additional value to the user over what is provided by the mobile application, there is a high probability that the Web site will rarely be used. If the resources are going to be put into creating one, its value needs to be made very clear. It is possible to have a successful mobile wellness application even without a companion Web site. While it has a cloud-based back-end, the user-facing aspects of *Jawbone UP* are still purely mobile, for instance. Similarly, *RunKeeper* thrived as a purely mobile application

for a number of years before its companion Web site was launched. Using a Web site is additional work for the user. If the user is asked to put in this extra work, it needs to be worth it.

Another potential disadvantage of using additional locations for self-monitoring feedback is that it increases the potential for security and privacy problems. Multiple feedback locations necessitate that the user's data resides in a cloud service. Although such services have become common and many of us entrust sensitive data to cloud-based services, it is important to note that health data can be particularly sensitive [77] and that if it is compromised, users' confidentiality and privacy can be put at risk. A steady stream of reports about security problems at even the most reputable technology companies indicates that these risks are real.

A larger number of feedback locations also increases the number of people who will likely see the feedback, amplifying risks to the user's privacy. As with glanceable displays, the additional feedback representations need to be carefully designed to mitigate those risks.

#### **3.2.4.3 Summary**

Cloud-based mobile services enable provisioning of self-monitoring feedback in a number of locations. Web sites are currently the most common location for such additional feedback, but in the future additional, highly tailored feedback can be distributed across a range of devices to maximize the application's effectiveness. The use of additional feedback locations increases privacy and security risks, however, and those will need to be addressed during the design process.

### **3.3 Open Questions for Providing Self-Monitoring Feedback**

As we noted throughout this section, although wellness applications have used a variety of feedback methods, there has been little effort to systematically evaluate how well those different forms of feedback work. We know of no evaluations that compare effectiveness of different types of feedback for supporting self-monitoring (e.g., do counts or graphs work better? For whom? For what behaviors?), nor are we aware of any studies that investigate how to optimize feedback for a particular

type of wellness behavior (e.g., what kinds of graphs work best for providing feedback on food-intake data?). There is still plenty to learn about the properties of different types of self-monitoring feedback for wellness behaviors.

In addition, there are a number of open questions related to individual forms of feedback that we have been discussing. We briefly discuss three areas that we think could be particularly fruitful for future work: presentation of uncertainty, designs of stylized representations, and optimizing multi-location feedback.

### 3.3.1 Presenting Uncertainty

One issue we touched on in our discussion of counts is that they can make the self-monitoring data appear more accurate and precise than it really is. An important open question is how to effectively mitigate this risk. Inferences from sensor data, correlations among different types of data, and other types of statistical calculations all have confidence intervals — a measure of how likely it is that the calculation is correct. One way to mitigate risks in current ways data is presented is to provide users with some representation of the confidence of displayed information. How exactly that should be done is unclear, however. Low numeracy rates suggest that presentation of raw confidence intervals would probably not be effective. Future research on different ways to present uncertainty and on the effects that those presentations have on people's behavior could help designers create representations of self-monitoring data that would enable people to develop a more accurate understanding of their behavior patterns.<sup>4</sup>

### 3.3.2 Designing Stylized Feedback

So far, stylized representations have only been used by a handful of mobile wellness applications, including our own application *UbiFit*. As such, there has been no systematic exploration of the design space for

---

<sup>4</sup>A related challenge is to develop better ways for users to indicate uncertainty of their estimates of their health behaviors during manual tracking. This information could then be used, along with confidence of automatic calculations, as part of confidence displays in feedback representations.

this type of self-monitoring feedback. Consequently, in addition to questions about learnability discussed above, there are a number of other open questions about how best to design stylized feedback. Among those questions are the following:

- How many different health behaviors and other metrics can be presented in a stylized representation while still maintaining the user's ability to quickly assess their status?
- What are the various dimensions of data that need to be supported by stylized representations to be able to represent a broad range of wellness data? Do they need to be able to represent time? Intensity? Duration? Another dimension? What types of images or animations support those various dimensions well?
- Do certain types of images (or themes) work better for certain populations than for others or for certain types of health behaviors than others? In our evaluations of *UbiFit*, we found that men were often initially skeptical about the flower theme of *UbiFit*'s glanceable display, but that the display still worked to keep them aware of their activities and their goals [20]. Would an image that they liked better have worked better?
- Would changing the theme regularly increase the display's effectiveness by minimizing habituation? If so, how often would the theme need to be changed? How often is too often? How many options need to be provided? What type of variety needs to be offered?

Research on questions such as these would contribute to the development of more robust — and more effective — stylized feedback displays.

### **3.3.3 Optimizing Multi-Location Feedback**

Finally, we have suggested that glanceable displays on the phone and in other locations (digital picture frames, large-screen displays, etc.) could help people reflect on feedback about their health behaviors more often, which, in turn, might help them to manage those behaviors more



effectively. Does having feedback in additional locations, over just using a phone-based glanceable display, help? If so, when and where should the additional feedback be displayed? At what point do additional locations of feedback stop producing any benefits?

A related question is what feedback locations are best suited for what types of information, for what types of feedback representations, and for what populations. A glanceable display on a phone is more private than a digital picture frame on a cubicle desk or a wall-mounted display. What types of information would people be willing to have displayed in each location? Would a set of suggestions for healthy activities based on the user's context and her wellness data be acceptable even on the phone? Would a *UbiFit*-style garden display work as feedback in a digital picture frame on a desk or is its cartoony nature too conspicuous for a public place such as the office? More research is needed to answer such questions.

Many other questions remain about how to provide effective self-monitoring feedback. We hope that HCI researchers and designers will begin to fill these gaps in knowledge in the coming years.

### 3.4 Section 3 Wrap-Up

In this section, we reviewed different ways that mobile wellness applications provide feedback about health behaviors and metrics that users track. Such feedback varies both in the type and location of where it is presented. Increasingly, wellness applications are combining multiple forms of feedback to provide users with ways to stay aware of their health activities and goals throughout the day. How to optimize such multi-location feedback to maximize reactivity of self-monitoring while minimizing user annoyance and privacy risks is still an open question.

# 4

---

## Supporting Goal-Setting

---

Attempts to encourage health and wellness behaviors frequently draw from behavioral psychology, incorporating, for example, Locke and Latham’s Goal-Setting Theory [58]. *Goal-Setting Theory* was based on nearly four decades of empirical research and Ryan’s [79] work claiming that a person’s actions are affected by her conscious goals. It describes how people respond to different types of goals, and thus which types of goals are more effective at motivating behavior. As industrial-organizational psychologists, Locke and Latham focus on the relationship between conscious performance goals and level of task performance — particularly in the workplace — though goal-setting theory is not limited to the workplace.

In their work, Locke and Latham found that people give the highest levels of effort and performance to the highest or most difficult goals. They also found that “*Specific, difficult goals consistently led to higher performance than urging people to do their best*” [p. 706]. The belief is that the more general “do your best” type of goal has too wide a range of acceptable levels of performance and is therefore difficult to judge if the goal has been attained. Though not all specific goals lead to high performance (as difficulty varies), they do reduce variation

in performance, as there is less ambiguity about what is expected. In accordance with this principle, physical activity recommendations, such as those provided by the American College of Sports Medicine (ACSM), follow the model of proposing specific, unambiguous goals.<sup>1</sup>

The relationship between goal and performance is strongest when people are committed to their goals [58]. The two factors that most contribute to *goal commitment* are:

1. the importance of goal attainment to the person, including the importance of the outcomes she expects to result from attainment, and
2. self-efficacy, that is, belief that she can achieve the goal.

Therefore if the person does not consider the goal to be important or does not believe she can achieve it, then she is unlikely to make the goal. Two ways to increase the *importance* of a goal for the person are to [58]:

1. get the person to make a public commitment to the goal, or
2. provide a monetary incentive.

If a monetary incentive is used, the amount of the incentive is important (more money = more commitment), as is the rate at which the person is paid. If she is paid only for *achieving* the goal and the goal is difficult, performance may drop significantly (e.g., if she realizes that she is not going to meet the goal and thus not receive the reward, her performance and self-efficacy will drop). The drop does not tend to occur if the goal is only moderately difficult or if the person is paid for performance (e.g., piece rate) rather than only for goal attainment.

Three ways to improve *self-efficacy* are to [58]:

1. ensure adequate training that leads to successful experiences,
  2. provide role-modeling with whom the person can identify,
- and

---

<sup>1</sup> A general exercise program recommended by the ACSM is to do at least 3–5 sessions per week of cardio training for 20–60 minutes per session, 2–3 sessions per week of resistance training involving 8–10 muscle groups per session, and 5–7 flexibility training sessions per week involving the static stretching of all major muscle groups [99].

3. provide persuasive communication that expresses confidence in the person's ability to achieve the goal.

Locke and Latham caution that conflicting goals may undermine performance if the conflicting goals motivate incompatible behaviors. They also stress the importance of *providing feedback* regarding how the person is progressing toward her goal [p. 708]:

*For goals to be effective, people need summary feedback that reveals progress in relation to their goals. If they do not know how they are doing, it is difficult or impossible for them to adjust the level or direction of their effort or to adjust their performance strategies to match what the goal requires . . . goals plus feedback is more effective than goals alone.*

Goals also serve as reference points for determining satisfaction in performance. Exceeding the goal tends to provide increased satisfaction; not reaching a goal reduces satisfaction and increases dissatisfaction. The more successful goal attainments a person experiences, the higher her total satisfaction.

Locke and Latham identified three types of goal sources:

1. self-set,
2. assigned, and
3. participatively set.

Locke and Latham have found that performance toward a goal set for a person (assigned) tends to be comparable with performance on a goal in which the person helped define the goal (participatively set), provided that the assigned goal is given with an explanation of the purpose or rationale for the goal. A goal that is set for a person (assigned) without an explanation of its purpose leads to significantly lower performance.

Shilts, Horowitz, and Townsend [82] provide a survey of interventions involving goal-setting as a strategy for promoting physical activity and dietary behavior change. Because much of the goal-setting literature focuses on the workplace, they conducted the survey to determine

goal-setting's effectiveness when used as part of physical activity and dietary interventions with adults ( $\geq 20$  years old), adolescents (12–19), and children ( $< 12$  years old). Regarding goal-setting, Shilts, Horowitz, and Townsend propose two new types of goal sources to add to the three previously identified by Locke and Latham.

4. *guided*, where a practitioner designs multiple goal choices and the person chooses one, and
5. *group-set*, where goals are designed and chosen either by a practitioner or a group of people who are participating together, and goal attainment is contingent on the performance of the group.

To allow for comparison of the effectiveness of goal-setting across studies, Shilts, Horowitz, and Townsend grouped studies into one of three levels of goal-setting support [p. 83]:

- Minimal support — “*Goal was set and no further support was provided regarding goal feedback or goal attainment. No goal-setting theory was mentioned as a guide to the goal-setting process*”;
- Moderate support — “*Goal was set and some but not all aspects of goal-setting were supported (i.e., feedback, barriers, and goal-attainment). Goal-setting theory was used to formulate the goal*”; and
- Full support — “*A majority of the intervention was focused on goal-setting and attainment, with extensive and appropriate support provided (i.e., feedback, contracting, barriers counseling, goal attainment, and skills development). Goal-setting theory was used to formulate the goal and plan and develop the lessons.*”

Shilts et al. concluded that [p. 92]:

*Moderate evidence indicates that implementing goal setting as a dietary or physical activity behavior change strategy is effective with adults, and those studies that*

*fully supported goal setting were more likely to produce positive results.*

There was not enough evidence to support a single type of goal-setting strategy as being most effective at encouraging physical activity and dietary behavior (e.g., self-set, assigned, or participatory — recall that Locke & Latham’s research involving workplace goal-setting found that *prescribed* is least effective unless it is accompanied by an understandable rationale). The intervention evaluation studies included in Shilts, Horowitz, and Townsend’s review showed a positive effect on physical activity and nutrition behavior — however, these studies did not compare interventions with and without goal-setting, but rather interventions with goal-setting versus no intervention. As such, goal-setting theory remains a promising strategy for the design of technologies to support health and wellness behaviors, but many open questions remain as to how to most effectively implement it.

#### 4.1 Goal-Setting in HCI Research

In addition to its use in health sciences and behavioral medicine, goal-setting has been used in HCI research on health and wellness and is also a central strategy in many popular commercial applications (apps) for encouraging physical activity and healthy eating. For example, *Nike+*,<sup>2</sup> *FitBit*,<sup>3</sup> and *Jawbone UP*<sup>4</sup> are just some of the popular mobile apps that use daily and/or weekly goals to motivate users to be more physically active. Similarly, *LoseIt!*,<sup>5</sup> a popular diet app, helps users achieve their weight loss goals by asking users to specify a target weight and how much they want to lose per week (e.g., 1 lb/week, 1.5 lbs/week, etc.). When specifying how much they want to lose per week, *LoseIt!* helps users understand how their choice affects *when* they are likely to reach their target weight. The app then calculates a daily caloric budget (that accounts for calories consumed *and* calories burned) for users to maintain in order to achieve their goal. All of these apps use *self-set*

---

<sup>2</sup><http://nikeplus.nike.com/plus> {Link verified 25 Aug 2013}

<sup>3</sup><http://www.fitbit.com> {Link verified 25 Aug 2013}

<sup>4</sup><https://jawbone.com/up> {Link verified 25 Aug 2013}

<sup>5</sup><http://www.loseit.com> {Link verified 25 Aug 2013}

goals, however some, such as *Jawbone UP*, suggest goals to help get users started (e.g., *UP* suggests 10,000 steps/day and eight hours of sleep per night).

A number of HCI research projects have explored the use of goal sources that are often not implemented in commercial applications. For example, *Fish 'n'Steps* [57] the *Mobile Lifestyle Coach* [35], and *Houston* [19] — our own system, which is described more below — all used *assigned goals*. Participants in the study of *Fish 'n'Steps* had a different daily step count goal for each week of the study. In the evaluation of the *Mobile Lifestyle Coach*, participants had a goal of achieving seven “lifestyle points<sup>6</sup>” each day.

HCI research has also explored the idea of *negotiated goals*. In their work on relational agents, Bickmore et al. [11] had the elderly participants in their study interact with the system’s animated wellness coach, *Laura*, to help set a step count goal for the following day. The negotiation was based on participants’ recent step counts and their medium-term (2-month) step count goal; it was intended to help the elderly participants slowly increase their activity until they reached their medium-term goal. Unlike the daily step count goal, however, the medium-term goals were *assigned* by the system based on the participants’ pre-study baseline.

Finally, like many commercial systems, a number of HCI research projects have used *self-set goals*. The diet application, *PmEB* [90], as well as our own applications *UbiFit* [20] and *GoalPost* [66] — which are described more below — enable users to set their own goals and track progress toward those goals. Like *LoseIt!*, *PmEB* encourages users to lose weight by setting a daily caloric deficit goal. The app then tracks users’ caloric balance based on journaled food intake and exercise data.

We now turn to a deeper discussion of our experiences with goal-setting to encourage physical activity and what we learned through the field studies that we conducted.

---

<sup>6</sup>A *lifestyle point* corresponded to engaging in 10 minutes of moderate physical activity or eating a serving of fruit or vegetables.

## 4.2 Our Experiences with Goal-Setting

We have used various goal-setting strategies in our work on using technology to encourage physical activity. In *Houston* [19], we used a daily step count goal, and in the *UbiFit* [20] and *GoalPost* [66] projects, we used weekly goals that incorporated multiple types of physical activities (e.g., cardio, walking, strength training, and/or flexibility training). In *UbiFit* and *GoalPost*, we also investigated the idea of having two weekly goals — a primary and an alternate.

### 4.2.1 Daily Step Count Goal

In our evaluation of the *Houston* mobile application to encourage people to increase their step count, we set a daily step count goal for study participants that was based on the recommendations in the President’s Council on Physical Fitness and Sports’ *Walking Works* program [73]. The program suggests that people use a pedometer to determine their daily step count every day for one week, then take the highest day’s count and use that as the daily goal for the next two weeks. After the first two weeks with the same daily goal have passed, the program suggests that the user add 500 steps per day to her walking goal at the end of each two-week period provided that on average, she has met her daily goal.<sup>7</sup>

According to the goal sources identified by Locke and Latham [58] and Shilts et al. [82], *Houston* used *assigned/prescribed* goals where each participant was assigned her own goal and was told how the goal had been assigned (i.e., a rationale was provided). Progress toward the goal and goal attainment were shown in many ways. For example, an “\*” next to a step count (either for the user or for any of the members of her group) indicated that the daily goal for that person was met (e.g., in Figure 4.1(a)). A congratulatory message (Figure 4.1(d)) appeared when the user updated her count if she met her goal. Progress could be seen on today’s view for any member of the group (Figure 4.1(b)), when the user updated her count (Figure 4.1(c)), on the daily average

---

<sup>7</sup>Our evaluation of *Houston* was a total of three weeks in the field — a baseline week plus two weeks with the initial goal.



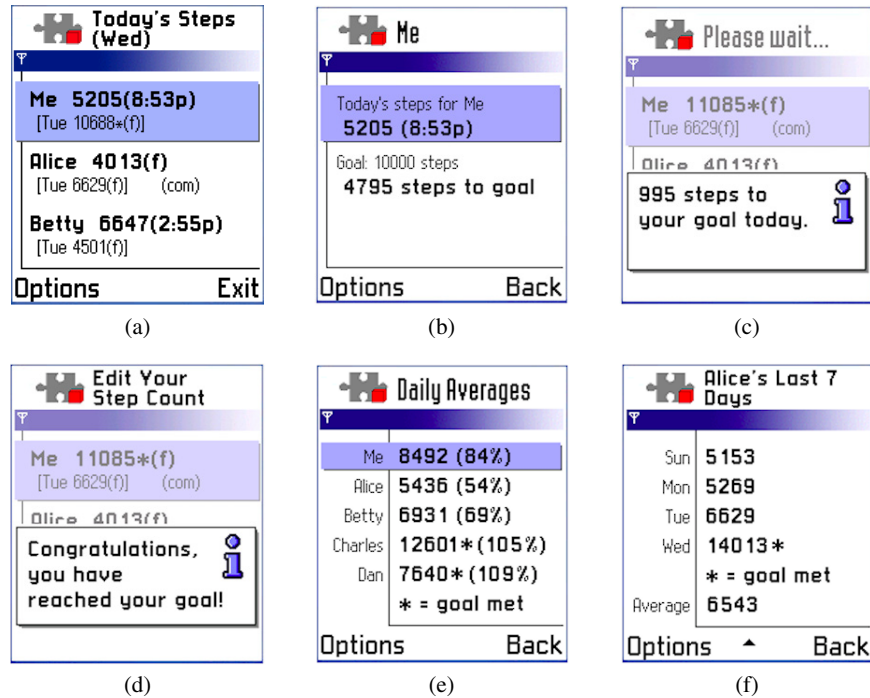


Fig. 4.1 **Viewing goals in *Houston***. *Houston* provided several ways for the user to see goal attainment and progress for herself and the members of her group. In all of the examples, an “\*” next to the step count indicates that the goal was met. In (a), the main screen shows the last updated count for today and yesterday for the user and the members of her group. In (b), the user can see the current count and remaining steps for herself or any of the members of her group. In (c) & (d), when the user updates her step count, a message appears either (c) telling her how many steps she still has to get until she reaches her goal or (d) congratulating her for meeting her goal. In (e), the user can see how she and the other members of her group have been doing on average. In (f), the user can see a summary view of her past 7 days, as well as her daily average (she can also see that view for any of the members of her group).

view for the group (Figure 4.1(e)), or on the view for any member’s last 7 days (Figure 4.1(f)).

The three-week field evaluation of *Houston* was a great learning experience for us. Although *Houston* was positively received and appeared to encourage many of the participants to increase their daily step count, we encountered three primary challenges with how we chose to implement goals: (a) the goal source, (b) the goal’s timeframe, and (c) what the goal targeted.

#### 4.2.1.1 Goal source

Recall that in the Walking Works program, a daily goal is assigned to the user based on her highest daily step count from the first (or baseline) week. However, during the baseline week in our study, many of the participants had one exceptionally high day. Had we followed the rules outlined in the Walking Works program, 10 of the 13 participants would have started their efforts to increase their daily step count with a goal of 12,000 steps per day or higher, four of whom would have started the program with a goal of over 15,000 steps per day.

When we noticed that so many participants had outlier days in their baseline week, we made the decision to set their daily goal based on the second highest count from their baseline week. This resulted in a daily goal for nine of the participants that was anywhere from 1,000 to 8,000 steps lower than what would have been set per the Walking Works program — see Table 4.1). With our slight modification to the goal-setting approach, participants began the program with a daily goal of from 9,000 to 19,000 steps per day.

With our modified goal-setting approach, we hoped that most participants would be on track to meet their daily goal. However,

Table 4.1. **Establishing a daily step count goal.** What the goal should have been according to the *Walking Works* program (i.e., highest day of baseline week), the goal we used (i.e., the second highest day of the baseline week), and the difference ([Highest Day] – [2nd Highest Day]).

ID	Goal based on Highest Day	Goal based on 2nd Highest Day	Difference
P1a	17,000	13,000	<b>4,000</b>
P1b	11,000	11,000	0
P1c	10,000	10,000	0
P1d	9,000	9,000	0
P2a	12,000	12,000	0
P2b	13,000	9,000	<b>4,000</b>
P2c	17,000	9,000	<b>8,000</b>
P2d	16,000	10,000	<b>6,000</b>
P3a	12,000	9,000	<b>3,000</b>
P3b	20,000	19,000	<b>1,000</b>
P3c	13,000	11,000	<b>2,000</b>
P3d	12,000	10,000	<b>2,000</b>
P3e	14,000	10,000	<b>4,000</b>

participants only met their goals on average for 34.06% of the days, ranging from meeting goals as often as 100% of the time (for Participant 1b, or “P1b”) to only 7.69% of the time (for P3d). As to why this rather low goal achievement rate, the data suggests at least two reasons. First, two of the participants specifically indicated in their notes for the day (an optional feature in the *Houston* application that some participants chose to use) that something happened which impacted their efforts to be more active. For P2a, who met her goal on 4 days (or 26.67% of the time over the 15 days she had a goal), she indicated that she started graduate school on the 8th day that she had a goal. All four days that she met her goal took place in the first seven days (i.e., before school started). For P2c, who met her goal on 3 days (or 20.00% of the 15 days that she had a goal), she indicated that she got sick on the sixth day that she had a goal. Incidentally, the 3 days that she met her goal took place in the first four days (i.e., before she got sick). Both of these cases could be explained by Locke & Latham’s warning that conflicting goals may undermine performance if the conflicting goals motivate incompatible behaviors [58], though other reasons could also have been at play (e.g., the novelty of starting a new physical activity program or participating in a study).

Another observation that we made from the data is that it was not uncommon for the first week’s average daily step count to be higher than subsequent weeks. For example, of the 13 participants, six had an average daily step count in the first week (i.e., the baseline week where they had no goal) that was at least 1,000 steps/day higher than in the following two weeks (i.e., the weeks where they had a daily goal) — see Table 4.2. This suggests that for nearly half of the participants, that baseline week may have resulted in artificially high step counts, leading to goals being set that were too high for participants to reasonably achieve.

#### 4.2.1.2 Goal timeframe

Though having a daily goal means more opportunities for succeeding, it also means more chances of failing. Results from our study suggest at least two problems with the daily timeframe of the goals as

Table 4.2. **Average Daily Step Counts.** Average daily step count for week 1 (baseline week, no goal) versus weeks 2 & 3 (with a daily goal), showing the difference ((Weeks 2 & 3 Average) – (Week 1 Average)).

ID	Week 1 Average	Weeks 2 & 3 Average	Difference
P1a	9,493	6,838	–2,655
P1b	8,327	11,323	+2,996
P1c	7,921	11,100	+3,179
P1d	7,565	9,798	+2,233
P2a	10,250	8,850	–1,400
P2b	7,639	6,195	–1,444
P2c	8,529	5,889	–2,640
P2d	9,889	6,677	–3,212
P3a	5,573	5,863	+290
P3b	9,109	14,681	+5,572
P3c	9,026	8,338	–688
P3d	7,056	6,105	–951
P3e	9,530	8,213	–1,317

implemented in *Houston*. First, most participants’ daily step counts did not have the “consistent” pattern that daily goals seem to encourage (a theme that we continued to see throughout our work on encouraging physical activity). Second — and perhaps related — participants who had rather high daily goals did not consider them to be achievable.

With the exception of two participants (i.e., the two who met their goal more often than anyone else), most participants’ step counts were not consistent from day to day. Rather, they tended to have a mix of high, medium, and low days throughout each week. For many participants, their “high” days were the only days that they met their goal, making goal achievement infrequent. Figure 4.2 illustrates the daily step count patterns for three participants in the study — one from each group of friends.

Similarly, participants who had goals that were higher than 10,000 steps per day often felt as if the goal was unreachable, especially on a daily basis. As they explained in the exit interview:

*“The way you guys programmed the goal — you picked a highly active week for me.”* { P1a; goal of 13,000 steps per day }

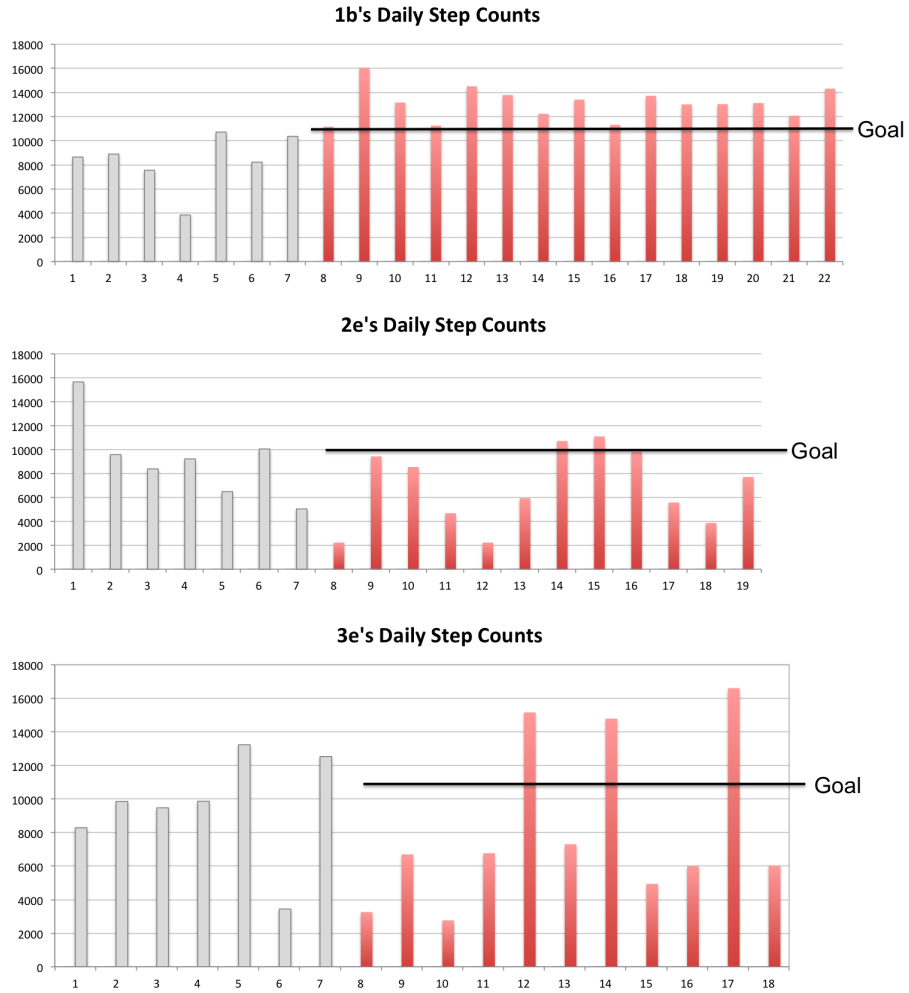


Fig. 4.2 Daily step counts for 3 Houston participants. The relative consistency of P1b's daily step counts compared to P2es and P3es more up and down patterns.

*“I felt like my goal was impossible a lot of the time.... I couldn't walk everywhere one day or if I didn't have time to take a big long walk, then there's just no way and then it became sort of like, well, it's too frustrating.” { P2a; 12,000 steps per day }*

*“I disliked the — the goal that was set. I felt it was too high...well, the first time I got it, I felt defeated.... Like oh man, I’m never gonna be able to make that. And then the days I did — I did meet it, I was like wow, well I was super active today, that’s why.” { P3c; goal of 11,000 steps per day}*

Despite improving from an average daily step count in her baseline week of 9,109 steps per day to 14,681 steps per day for the two weeks she had her goal *and* passing 20,000 steps on three days, P3b felt defeated because she only met her goal of 19,000 steps per day four times.

The participants who were assigned those high goals were often already doing some sort of planned exercise that was detected by the pedometer (e.g., running or fitness walking), which is how they got assigned the rather high goals. However, though they were interested in increasing their physical activity, they were not interested in (or frankly able to commit to) doing *that* much activity — such as a 6-mile run — every single day. Recall that goal-setting theory suggests that in order to meet the goal, the person has to be confident in her ability to achieve it. The participants who ended up with the higher daily goals were not confident that they could achieve their goals, and it often resulted in them feeling defeated or frustrated, which was not a goal of ours.

#### **4.2.1.3 Goal target activity**

Finally, we noticed that our choice to focus on step count — which was being strongly recommended by the health community both in the research literature and the media at the time — actually discouraged participants from performing healthy physical activities in some circumstances. This ranged from participants simply choosing not to perform a physical activity they otherwise would have performed to having participants feel frustrated because they did not receive proper credit from the pedometer for performing healthy physical activities. For example, some participants stopped walking uphill or running because they got less credit for those harder activities than they did for an easy walk, and others decided not to go for a bike ride because they would

not get any credit for the activity. In other cases, participants just got fed up because they performed a physical activity such as cycling or rock climbing and did not get any credit from the pedometer for doing those healthy activities.<sup>8</sup> Though this was a source of frustration for all groups, it was particularly frustrating for those whose step counts were being shared with group members because the step count did not necessarily reflect their day accurately.

Interestingly, the participant whose average daily step count went down the most between the baseline week and the two weeks with the goal (i.e., 2d whose average daily step count went from 9,889 to 6,677) was also the most physically active participant in the study, who regularly does organized fitness events such as triathlons and cycling races. However, the bulk of her physical activity was from cycling — something the pedometer simply did not detect.

In an effort to overcome some of these challenges, when we moved to our next project — *UbiFit* — we chose to change the goal source, goal timeframe, and to account for the range of physical activities that someone could perform.

#### 4.2.2 Weekly Mixed Physical Activities Goal

In *UbiFit* [20] and *GoalPost* [66], which were both designed to encourage people to participate in regular and varied physical activity, users set weekly physical activity goals. Goals ran from Sunday to Saturday and were broken down by category — cardio, strength, flexibility, walking, and other — to promote participation in varied activity (rather than, for example, focusing on step count as we did in *Houston*). Users could specify goals at the category level or for specific activities within the category (e.g., 90 minutes of cardio versus 30 minutes of running and 60 minutes of cycling). Goals could be comprised of any to all of the categories.

---

<sup>8</sup>Incidentally, when we piloted *Houston* with members of the research team and other colleagues prior to the study, the pilot participants used online step count calculators to translate activities such as a bike ride into steps, then included that in their daily counts. Participants in the study, however, did not do that. When we asked about it in the exit interviews, they told us that would feel like cheating or take too much extra effort.

In our evaluations of *UbiFit* and *GoalPost*, participants set their own goals (i.e., the goal source was *self-set*). In *UbiFit*, the evaluators input the goal that the participant set into the system, which then synced with the participants' phones in the *UbiFit* application. Participants could change their goals, but had to work with a researcher to do so. In *GoalPost*, participants set their own goals as part of the *GoalPost* application on their phones. In *GoalPost*, goals were set each week. Participants could reuse a prior goal, edit a prior goal, or create a new goal. The participants could change their goals at any time. Both *UbiFit* (in the three month evaluation) and *GoalPost* supported two goals to be set per week (discussed in the next subsection).

*UbiFit* and *GoalPost* provided various ways of viewing progress toward the goal. In *UbiFit*, goal attainment for the past four weeks could be seen on the glanceable display, which resided on the phone's background screen. A large butterfly in the upper right appeared when this week's goal was met (Figure 4.3(b)), and up to three smaller butterflies appeared to the left to show goal attainment for the prior three weeks (Figures 4.3(a) & 4.3(b)). Within the *UbiFit* application, the *Goal View* showed the goal itself and progress toward the goal (Figure 4.3(c)).

In *GoalPost*, goal attainment and progress was shown in a variety of ways. From the main screen of the application, a bar graph showed progress for each of the goal's activity categories, as well as a percent complete for this week's and last week's overall goals (Figure 4.4(a)). From the *Goals* screen, the user could see the percent complete for each goal. The *Goal Progress View*, accessible from the main screen, included a line chart that showed progress toward the user's goals over the week with a listing of her goal items and activities, as well as how much progress had been made for each activity category (Figure 4.4(b)). The user was able to navigate to previous weeks' data to view her progress over time. Finally, the *Trophy Case* showed trophies and ribbons that represented completed goals and activity categories (Figure 4.4(c)). For each weekly goal that the user completed (i.e., all components of the goal), she received a trophy, and for each category within her weekly goal that she completed — e.g., once she completed the “cardio” portion of her goal for the week — she earned a ribbon. Unlike *UbiFit*'s



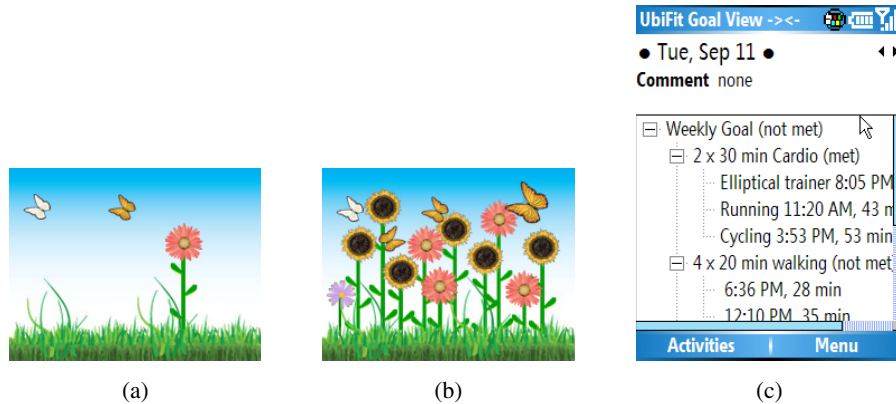


Fig. 4.3 **Viewing goals in UbiFit.** UbiFit showed four weeks worth of goal attainment on the glanceable display and provided a *Goal View* within the application that showed the week’s goal and progress toward it. In (a), the two small butterflies show that the user has met her goal for two of the past three weeks. In (b), the large butterfly at the right shows that the user has met this week’s goal and the three small butterflies show that she has met her goal for the prior three weeks. In (c), the *Goal View* shows that the user has met her *Cardio* portion of her goal for the week (and the activities that she performed to achieve the goal), but she has not yet met the *Walking* portion of her week’s goal.

glanceable display, which appeared on the background screen of the user’s phone, the Trophy Case was accessed from within the *GoalPost* application (see the “Trophy Case” button in the upper right of the main screen as shown in Figure 4.4(a)).

As with the three-week field evaluation of *Houston*, the three-week and three-month field evaluations of *UbiFit* and the four-week field evaluation of *GoalPost* were great learning experiences for us. Both systems were positively received and appeared to encourage most participants to be physically active and in many cases, to include variety in their routines. However, we encountered two primary challenges with how we chose to implement goals: (a) helping participants develop realistic expectations for self-set goals, and (b) determining when small rewards for goal achievement are effective.

#### 4.2.2.1 Realistic expectations

Though participants set their own goals for *UbiFit* and *GoalPost*, we still saw lower goal achievement rates than we expected. One likely

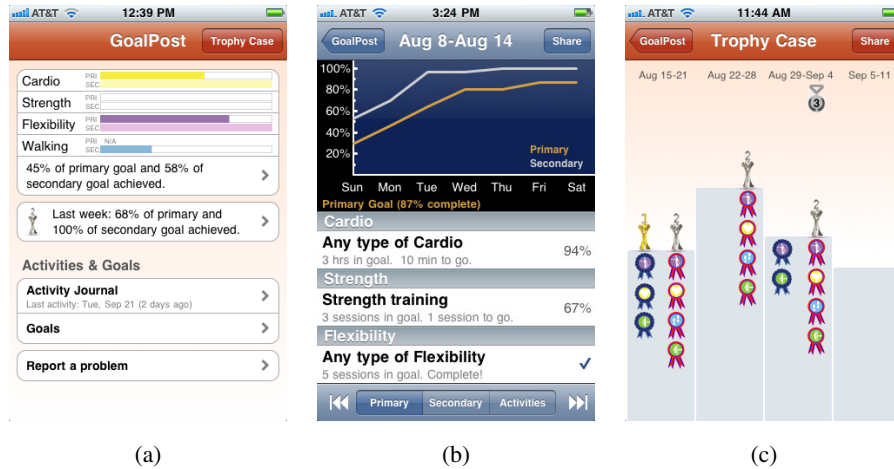


Fig. 4.4 **Viewing goals in *GoalPost*** *GoalPost* provided several ways for the user to view her goal, goal progress, and goal attainment. In (a), the *GoalPost* application’s main screen shows progress bars for each activity category of the user’s goal as well as a percentage of how much of her goal has been achieved and how much of her goal she achieved last week. In (b), the Goal screen included a line chart that showed progress toward the user’s goals over the week with a listing of her goal items and activities, as well as how much progress had been made for each activity category. In (c), the Trophy Case showed trophies and ribbons that represented completed goals and activity categories. The “3” medal shows that the user started a streak by achieving her secondary goal for 3 weeks in a row.

reason is that participants consistently overestimated how active they actually were before they started to keep track of their physical activities. As a result, they often set what seemed to be easy to achieve goals that actually required substantial changes in their lives to achieve — changes that they could not appreciate prior to keeping track of their actual activities. With the 3-month study of *UbiFit*, it became clear that for many participants, it could take many weeks — often more than a month — to realize that substantial lifestyle changes were needed. It was easy for participants to discount a week or two or three of minimal activity as an uncommon exception when in fact the exceptions were the rule. For example, one participant from the three-month study of *UbiFit* explained how it took her about six weeks to realize that she needed to fundamentally change what she was doing in order to meet what she thought was a simple goal:

*What kicked me into gear was realizing that I hadn't met my goals, not once. That was kind of scary, that I could not make that little bit of time... I didn't realize that I didn't make time for myself. That was really shocking that I couldn't do whatever, the 20 minutes for the flexibility and the stretching. I mean, that you could just do watching TV! That was just ridiculous. That just blew my mind. I was shocked... [before] I would start and I'd be like, 'oh, I'm too tired to do it,' and I wouldn't do it... Around December, the week of the 16th, things just kind of turned around for me... a revelation that I just needed to get my act together. {S7}*

#### 4.2.2.2 Small rewards for goal achievement

With both *Houston* and *UbiFit*, most participants mentioned how they really appreciated the little rewards that they received — from the simple “\*” that appeared next to a participant’s step count in *Houston* when she met her daily goal to *UbiFit*’s butterflies for achieving the weekly goal and flowers for performing individual physical activities.<sup>9</sup> In fact, in many cases, they were surprised at how much they appreciated those little rewards.

However *GoalPost*’s trophies and ribbons were not as effective. Most participants were indifferent to *GoalPost*’s rewards; they didn’t find the rewards to be particularly motivating, nor did they find the rewards to be a nuisance or bother. Six of the 23 participants were critical of the rewards, describing them as “lame,” “unnecessary,” or a “gimmick.” However, it is worth noting that only two of those six ever earned a trophy. Only three participants actually seemed to be motivated by *GoalPost*’s small rewards. In fact, one considered lowering her goal mid-week so that she could get the trophy. Interestingly, one other participant thought the rewards would be motivating until he received his first trophy.

<sup>9</sup> However, as we expected, many participants asked for a choice of metaphors with *UbiFit* — this was a more common request in the three-month study compared to the three-week study.

### 4.2.3 Primary & Secondary Goals

In an effort to help users stay active in challenging weeks, both *UbiFit* (in the three month evaluation) and *GoalPost* supported users setting an optional, secondary goal each week. One of the differences in how the alternate goal was implemented in *UbiFit* versus *GoalPost* is how it was described to participants. In *UbiFit*, it was described as being an “alternate” goal — basically, a backup goal in case the week was more challenging than expected. Eighteen of the 28 participants in the 3-month study of *UbiFit* chose to set an alternate weekly goal in addition to their primary goal. However, in *GoalPost*, it was simply described as being a secondary goal, and the participants were free to interpret that however they chose. Nineteen of the 23 participants in the *GoalPost* study set up a secondary goal; 10 used it as a stretch goal, four used it to try to encourage themselves to introduce more variety into their routines, and the others essentially used it as a backup goal. The other major difference between *UbiFit*'s and *GoalPost*'s implementation of secondary goals was that *UbiFit* forced the user to make a choice as to which of the two goals she was going to pursue for the week,<sup>10</sup> while in *GoalPost*, users simultaneously got feedback on and tried to achieve both.

Though we feel as if we've just started to scratch the surface of exploring the use of multiple weekly goals, one issue became apparent with our implementation of alternate goals in *UbiFit*, which led to how goals were implemented in *GoalPost* — that is, forcing a choice between goals. Another issue that came up in the *GoalPost* evaluation was what type of goal should that second goal be?

#### 4.2.3.1 Forcing a choice

In the three-month study of *UbiFit*, participants who set alternate goals often forgot about them, perhaps because the goal defaulted to the primary each week and they actively had to go into the application and switch to their alternate goal, or perhaps because they did not

---

<sup>10</sup>She could switch between the two goals in the *UbiFit* application at any time during the week, but at any given time, she was only trying to achieve one (and each week began by defaulting to the primary goal).

see feedback on and simultaneously try to achieve both. One reason we forced users to choose between the goals in *UbiFit* was that we hoped it would help them more quickly discover if they had set an unachievable primary goal and needed to update it. However, given that so many of them simply forgot about the alternate goal, the feature did not work as intended. That prompted our switch in *GoalPost* to support two, simultaneous weekly goals, which was more effective.

#### 4.2.3.2 Secondary goal type

In *UbiFit*, we described the optional goal to participants as being a backup goal, and thus, that is how they used it. However, in *GoalPost*, when we simply called it a secondary goal, participants chose to use it in one of three different ways: as a backup goal in the event of a busy week, as a stretch goal to push themselves to the next level, or as a goal to incorporate more variety into their physical activity routines (e.g., start strength or flexibility training). It isn't clear if one of these was more effective than the other. In fact, allowing users to appropriate a secondary goal in a way they choose may be one of the most critical insights from our experiences with *GoalPost*.

### 4.3 Open Questions for Supporting Goal-Setting

Recall from Locke and Latham's research that it is important to help the user set achievable goals because not achieving a goal (a) reduces satisfaction and increases dissatisfaction, and (b) the more successful goal attainments a person experiences, the higher her total satisfaction. It is also essential that the goal be something that is important to the user — something where both goal attainment and the outcome that she expects to result from attainment are important.<sup>11</sup>

Our experiences with goal-setting have included technologies that we developed that used goal-setting as a strategy, along with other motivational strategies, to encourage people to participate in regular physical activity. Our evaluations did not systematically test different

---

<sup>11</sup>Based on our own personal experiences, we caution that it can be easy to set achievable physical activity goals that are not very important to the user because the goals are too easy.

implementations of goal-setting, but rather implemented versions of goal-setting based on recommendations from the literature (and later, our own experiences), and studied participants' experiences with those implementations. Many open questions remain about how to best implement goal-setting in mobile technologies that encourage physical activity. In the following, we suggest at least four areas that would benefit from systematic investigation: setting goals, providing feedback on progress and achievement, supporting multiple goals, and supporting goal-setting over time.

### 4.3.1 Setting Goals

One area that needs additional investigation is about how to best help users set goals. Based on our experiences, two areas that we believe are in need of additional exploration are around (a) using baseline data to set goals and (b) understanding what goal sources and timeframes are most effective.

#### 4.3.1.1 Setting goals based on baseline data

As we observed in our studies of *Houston*, *UbiFit*, and *GoalPost*, it can be challenging to help users set achievable goals when goals are being set based on inaccurate baselines. In the case of *Houston*, the baseline was determined from one week's worth of monitoring the user's activity (specifically, step count), and in *UbiFit* and *GoalPost*, the "baseline" was the user's own reflections on her activity level. In both cases, this led to problems for many participants that resulted in unachievable goals being set.

At the time of our *Houston* study in Summer 2005, there was a hypothesis in the health community that adults might change their behavior during baseline-setting weeks simply because those weeks involve wearing a monitoring device like a pedometer. As Tudor-Locke [91] explains:

*Regardless of the data recording specifics, there is always a concern that participants will alter their behavior simply because they are being monitored (also known as*

*reactivity*). Vincent and Pangrazi [97] recently ruled out reactivity in children wearing sealed pedometers. The potential for reactivity using unsealed pedometers has not been well explored yet. A thesis at Arizona State University focused on this problem found preliminary evidence that children did not alter their behavior when monitored by unsealed pedometers compared to sealed ones [70]. At this time we do not know if reactivity is a problem with adults, regardless of whether or not the pedometer is sealed. Additional research is needed to address these niggling issues.

More recently and perhaps not surprisingly, Clemes and Parker [18] found evidence of reactivity in adults when wearing pedometers, especially when unsealed pedometers (i.e., pedometers where the person can see her step count) are used and people are asked to keep a record of their step counts — something that we did in our *Houston* study. Clemes and Parker’s finding could explain how nearly half of the participants in our *Houston* study had a higher average daily step count in the first week of the study when they did not have a daily goal compared to the following two weeks of the study where they did have a daily goal. Clemes and Parker conclude that: “*This has validity implications for short-term pedometer studies investigating habitual free-living activity that require participants to provide a daily log of their step counts.*”

Not only does Clemes and Parker’s finding raise issues with the validity of short-term studies of daily step counts, but it also presents a challenge for those of us who design technologies to support people in their efforts to be more physically active, especially when the focus is on consumers empowering themselves to make a change, rather than people who are participating in a research study or patients (or clients) working with a healthcare professional.

Regarding the problem of users setting their own goals based on an overestimation of their current activity level, it could be the case that people are projecting their “idealistic” activities or what they had planned to do (and forgot that they did not actually follow through) rather than their “realistic” activities. If that is the case, Goffman’s

work on impression management may help explain what is going on [28]. The act of setting a goal could be a type of performance, and users may be attempting to satisfy expectation whether or not they are capable of meeting those expectations. Participating in a research study where study researchers are an audience in addition to the user herself and any social features that the technology supports may exacerbate this.

It remains an open question as to how a technology can help a user set a good physical activity goal that is based on baseline measurements of activity if that baseline is likely to be skewed, whether it be from the initial stages of monitoring or poor recall of one's current behavior. Some ideas to pursue include investigating the use of stabilized baselines [25]. For example, in the case where monitoring is being used to establish a baseline, perhaps instead of fixing the baseline's time period, the technology could wait until it observes that the baseline has stabilized before suggesting that the user set a goal. Another idea is to suggest a change when the goal seems unachievable. For example, if the system detects that the user has failed to achieve her goal a few times, it could suggest that she may want to adjust her goal; this could also be an opportunity to use multiple, simultaneous goals which are discussed more below.

#### **4.3.1.2 Effective goal-sources & timeframes**

Recall that between Locke and Latham [58] and Shilts et al. [82], five types of goal-sources were identified: (a) self-set, (b) assigned/prescribed, (c) participatory/collaborative, (d) guided, and (e) group-set. Our experiences to date have involved assigned/ prescribed and self-set, and it is clear that there is room for improvement in how those goal-sources are implemented. We have also experimented with daily and weekly goals; though the weekly goals seemed to be a reasonable timeframe, there is opportunity for additional investigation of those and other timeframes.

In the exit interview of the 3-month evaluation of *UbiFit*, we proposed several alternative goal-sources as well as timeframes to



participants who then speculated about their preferences [21]. We specifically asked about the following goal-sources and timeframes:

- **Self-set:** the user could set her own goal (i.e., as the participants experienced in the study),
- **Assigned/Prescribed:** the user could be assigned a goal based on expertise from an authorized person or from some identified source, for example:
  - **national recommendations:** where the user could choose from established physical activity guidelines, such as those set by the ACSM, AHA, U.S. Surgeon General, or President’s Council on Physical Fitness,
  - **fitness expert:** a personal trainer could set the goal for the user, or
  - **medical expert:** the user’s medical doctor could set the goal for her,
- **Participatory/Collaborative:** the user could work with an expert such as a personal trainer or medical doctor to help her set her goal,
- **Guided:** the user could choose from a set of goal options that an expert, such as a personal trainer or her medical doctor, prepared for her.
- **Group-set:** the user could work with a group of strangers or members of her social network to set a goal for the group where even if she did her part, if someone else did not do his or her part, the goal might not be achieved.

Although the self-set goals employed in our 3-month study were popular, participants found some of the other options appealing as well, particularly the personal trainer and group-set options. In many cases, participants thought those goal source options could be motivating. However, cost was a common concern for the personal trainer options and certain barriers — such as illness or work deadlines — as well as past experiences with friends and family were concerns for the group-set options. Unpopular options were the “national recommendations” and

“medical expert/doctor” options. However, despite being unpopular, if self-set goals are used in a system, it might be useful to provide the national recommendations for reference — or possibly as default goals — to remind users of the type of variety or general amounts for which they should strive. Additionally, for users who also have health concerns that may impact their ability to perform physical activities, the medical expert/doctor options might be appropriate.

In the same interview we asked about three possible timeframes for these goals:

- **Rolling seven-day window:** a seven-day timeframe that shows the user’s last seven days worth of activity and whether those last seven days qualify as having met her goal (i.e., a rolling seven-day window that never resets),
- **Customizable calendar week:** a goal that resets once per week, where the user specifies the reset day at the start of the study, or
- **Fixed calendar week:** a goal that resets once per week, where the reset day is fixed (i.e., as participants experienced in the study).

Regarding timeframe, the calendar week timeframe that was used in the study was popular with participants. However, slightly more participants would have preferred that the week begin on Monday rather than Sunday. Of course, their preferences were not quite as simple as that. They would have liked the ability to change on which day their week resets (e.g., so that their weekly goal could run from Sunday to Saturday, Wednesday to Tuesday, etc.) and potentially switch to a seven-day rolling window model where the week does not reset, but rather moves forward one day at a time. Of the participants who were interested in the rolling seven-day window idea, most tended to be less active on average or were less consistent with their activity levels. In many cases, their reason for wanting to try the rolling seven-day window model was that the flowers in their garden would last longer. However, they might benefit from a different timeframe altogether until they are more active on a regular basis. Perhaps instead of a seven-day timeframe, a four-week timeframe might provide more motivation for them, even if their

goal resets, so that the little rewards for behavior (e.g., the flowers that we used in *UbiFit*) persist longer. As the user improves the regularity of her behavior, the time window could shorten (e.g., from four to two to one week). See Consolvo et al. [21] for more on these ideas.

Systematically investigating the effectiveness of these and potentially other goal sources and timeframes would be very helpful to the community.

### 4.3.2 Providing Feedback on Progress & Achievement

Another area that needs additional investigation is about how to provide feedback to people on the progress that they're making toward their goal as well as for when they achieve (and exceed) it. Two areas that we believe are in need of additional exploration are around (a) multidimensional goals and (b) providing simple rewards for goal progress and achievement.

#### 4.3.2.1 Multidimensional goals

One of the big changes we made when moving from *Houston* to *UbiFit* and later to *GoalPost* was to evolve from focusing exclusively on encouraging step count to encouraging a broader range of physical activities. When we made this switch, we had to find a way to show progress and achievement toward a more complex goal in an understandable and meaningful way.

In *UbiFit*, we implemented goal progress and feedback in two different ways: (a) as an aesthetic image on the glanceable display, and (b) as a structured list within the interactive application. In *UbiFit*'s glanceable display, butterflies represented goal attainment and the different types of flowers represented different types of physical activities that the user performed. At a glance, the user could easily determine if she was having a generally active or inactive week, if she incorporated variety into her routine, if she met her primary or alternate weekly goal, and if she met her primary or alternate goal recently. If she looked more closely, she would notice that flowers that had stems with leaves represented activities she performed that counted toward her weekly goal, while flowers that had stems without leaves represented activities that

did not count toward the weekly goal, but still served to remind the user of how active she had been. The structured list showed her goal broken down by subcategories and listed what she had done within that subcategory as well as what remained.

*GoalPost* also used multiple ways to represent goal progress and achievement including with bar graphs, a percent complete, a line chart, a structured list, and a trophy case. On the main screen of the *GoalPost* application, goal progress was shown via a bar graph for each of the goal's activity categories, as well as a percent complete for this week's and last week's overall goals. The *Goal Progress View* included a line chart that showed progress toward the user's goals over the week and a structured list similar to that used in *UbiFit*.

From our study interviews, it seemed that our approaches generally worked, though we did not systematically experiment between these and other ways of showing progress and feedback. Systematic experimentation of our and other approaches would be helpful to determine how to best show progress and feedback to users when these more complex goals are being supported.

#### 4.3.2.2 Simple rewards

In our studies of *Houston* and *UbiFit*, the simple rewards that we provided for goal achievement (and in *UbiFit's* case, also for performing activities) were very well received. Participants were often surprised about how good those simple rewards made them feel, from the "\*" next to a step count in *Houston*, to a butterfly or flower in *UbiFit*. In many cases, participants cited those rewards as motivators to get them to do more activity. However, in our study of *GoalPost*, the trophies and ribbons that we provided were no where near as motivating or appreciated, despite the positive reactions we received about them from respondents in a survey that we conducted to help inform the design of *GoalPost*.

Many things could have contributed to the different reactions — none of which have been systematically investigated. For example, perhaps it was something about the nature of the reward. Was it because the trophies and ribbons were too literal compared to the more abstract

rewards provided by *Houston* and *UbiFit*? Or is it more motivating to grow something like a garden than it is to fill a trophy case? Another alternative is that the problem could have been *where* the rewards were provided — *GoalPost* users had to specifically seek out their rewards.<sup>12</sup> Other factors could have also been at play.

Based on our experiences, we suggest that this area needs further investigation. When implemented well, simple rewards can help motivate users to be physically active and remind them to be proud of their healthy accomplishments. However, it is also clear that not all simple rewards are effective.

### 4.3.3 Supporting Multiple Goals

The idea of using multiple, simultaneous goals seems to be promising for supporting health and wellness behaviors — we made clear progress from the way we implemented multiple goals in *UbiFit* to the way we implemented them in *GoalPost*, but more work remains in determining what type of implementation best helps users. We propose at least two trajectories to investigate. First, how many and what types of goals are helpful to the user without becoming overwhelming? For example, should the technology support a main goal and a backup goal? A main goal and a stretch goal? A main goal, backup goal, and stretch goal? Should the type of goal be suggested by the technology or left up to the user to define?

Second, though *GoalPost*'s implementation seemed more successful than *UbiFit*'s, the “choose one” versus simultaneous pursuit was not the only difference — participants in *UbiFit* were not prompted to set their primary and alternate goals each week as participants in *GoalPost* were. It is possible that forcing a choice might be more effective if the user

---

<sup>12</sup>In *UbiFit*, the rewards were provided on the background screen of the phone, so any rewards were seen by participants whenever they used their phones. Though not quite as accessible as *UbiFit*'s rewards, *Houston*'s “\*” reward was attached to individual step counts, so if a participant reviewed her current or past step counts, she got to see her rewards without seeking them out specifically; the members of her group also got to see her rewards. In *GoalPost*, however, participants had to open the app and select the *Trophy Case* button to view their rewards (Figures 4.4(a) & 4.4(c)).

is prompted each week to set goals and chooses then — that remains an open question.

#### 4.3.4 Supporting Goal-Setting Over Time

Helping people maintain a healthy lifestyle likely means managing differing goals at different times in their lives, even after the technology has successfully gotten them started on a good path. Some key questions in this area that we believe are important to investigate include:

- How can the technology maintain the user’s interest over time?
- When should the technology support the user maintaining her current activity level versus intervene to encourage more, less, or different activity?
- If the technology detects a change in the user’s pattern toward more sedentary behavior, how can the technology determine if it should leave the user alone versus provide additional encouragement to help the user get back on track?

A related question deals with how the technology can reliably distinguish between a temporary lapse in activity (e.g., due to a vacation, a minor illness, or a minor injury) versus a real regression toward more sedentary behavior. Basically, how can the technology be helpful and maintain the user’s interest while avoiding being a pest?

#### 4.4 Section 4 Wrap-Up

In this section, we provided an overview of key aspects of goal-setting theory and how goals have been used in mobile technologies from the HCI literature that attempt to encourage health and wellness behaviors. We provided details about how we have implemented goal-setting in our own work on encouraging physical activity and described several challenges that we encountered. Finally, we suggested at least four areas of goal-setting that would benefit from further investigation: setting goals, providing feedback on progress and achievement, supporting multiple goals, and supporting goal-setting over time.

In spite of the many challenges we faced with implementing goal-setting in our work and the low rates that we saw of goal achievement, all three systems that we developed — *Houston*, *UbiFit*, and *GoalPost* — appeared to help participants develop a more accurate reflection of how active they really were and motivate them to incorporate more physical activity into their lives, at least in our multi-week and multi-month field studies. Even though they frequently failed to meet their goals, the goals (combined with the self-monitoring aspects of the system) seemed to help encourage the participants to honestly reflect about their activity levels, develop a more nuanced understanding of the situations in which they were and were not being active, and nudge them to be more active than they had been.

# 5

---

## Moving Forward

---

Throughout the previous sections, we have suggested a number of open research questions that we believe should be investigated in the areas of collecting behavioral data, providing self-monitoring feedback, and supporting goal-setting as they apply to the design and development of mobile technologies to encourage health and wellness. In this section, we discuss other important related areas of future work that need further investigation by the HCI community. One area, based on our own experiences, is how to assess the user’s starting level and progress regarding the target behavior. Other, more forward-looking areas include how to support the user when “stuff” happens and how to support wellness behaviors over the user’s lifespan.

### **5.1 Assessing Starting Level and Progress**

As we argued earlier, in order to adequately support the user in her efforts to be healthy, it is often useful to know the user’s baseline for the target behavior and be able to measure her progress. Such assessments are important for providing the user with feedback and can be particularly helpful if the technology includes goal-setting or coaching components.



Three instruments that we have used to assess starting level and progress for physical activity are (a) the President's Council on Physical Fitness and Sports' *Walking Works* program [73], (b) the American College of Sports Medicine (ACSM) and American Heart Association's (AHA) *Physical Activity Guidelines* [39], and (c) the *Transtheoretical Model of Behavior Change* [74].

Unfortunately, we have faced challenges when using each of these approaches, as we discuss next.

### 5.1.1 The Walking Works Program

As mentioned in Section 4, we used the *Walking Works* program as a guide to set a daily step count goal for participants in our three-week field study of the *Houston* mobile application [19]. To set a daily goal, the program suggests that people track their daily step count every day for one week, then take the highest count from any given day and use that as the daily goal for the next two weeks. We ran into problems with this approach that appear to have stemmed from a variety of reasons, even though we slightly modified the rule and used the second highest count as the basis for the goal. For example, some participants had one or two outlier days in their baseline week due to structured exercise that they were already doing, such as a one-day-a-week 6-mile run or a weekend hike. We also found that nearly half of the participants had a higher average daily step count in the baseline week when compared to the weeks that followed, which may have been due to reactivity of self-monitoring (as described by Tudor-Locke [91] and Clemes & Parker [18]), or by the introduction of a competing goal or interruption to their routine (e.g., the start of graduate school or coming down with a cold). As a result of our experiences with *Houston*, we used a different approach to goal-setting in our subsequent work with *UbiFit* and *GoalPost*.

Our experiences as well as the experiences of others, such as Clemes and Parker [18], suggest that the initial week with an unsealed pedometer (i.e., a pedometer where the user can see her step count) may not provide a true baseline. This can make it difficult for the technology to help users set goals when using approaches like that in the *Walking*

*Works* program. As we suggested in Section 4, some ideas for the community to investigate that may mitigate the risk of using potentially skewed baselines to help users set goals include: (a) waiting until the baseline has stabilized before suggesting a goal<sup>1</sup> [25], or (b) suggesting a goal based on the observed baseline, even if it is inaccurate, and then recommending changes to the goal when it appears that the goal is too difficult or easy to achieve.

### 5.1.2 Physical Activity Guidelines

Rather than using a pedometer as suggested by the *Walking Works* program, a person's level of physical activity could be assessed by comparing her activity to what is recommended by the American College of Sports Medicine (ACSM) and American Heart Association (AHA) [39]:

*To promote and maintain health, all healthy adults aged 18–65 yr need moderate-intensity aerobic physical activity for a minimum of 30 min on five days each week or vigorous-intensity aerobic activity for a minimum of 20 min on three days each week. Also, combinations of moderate- and vigorous-intensity can be performed to meet this recommendation. For example, a person can meet the recommendation by walking briskly for 30 min twice during the week and then jogging for 20 min on two other days. Moderate-intensity aerobic activity, which is generally equivalent to a brisk walk and noticeably accelerates the heart rate, can be accumulated toward the 30-min minimum from bouts lasting 10 or more minutes. Vigorous-intensity activity is exemplified by jogging, and causes rapid breathing and a substantial increase in heart rate. This recommended amount of aerobic activity is in addition to routine activities of daily living of light intensity (e.g., self care, cooking, casual walking*

---

<sup>1</sup>Two questions we have for the stabilized baseline approach are: how long does it typically take for a baseline to stabilize, and how long are users willing to wait for a goal?

*or shopping) or lasting less than 10 min in duration (e.g., walking around home or office, walking from the parking lot).*

The ACSM and AHA suggest, “*persons who wish to further improve their personal fitness, reduce their risk for chronic diseases and disabilities or prevent unhealthy weight gain may benefit by exceeding the minimum recommended amounts of physical activity.*” They also recommend that people perform strength and flexibility training, but they address those activities separately from their minimum physical activity recommendation.

Conveniently, these recommendations are relatively easy for technology to assess. If the *type* of activity performed (e.g., vigorous-intensity cardio or moderate-intensity walking), *when* it is performed, and *for how long* it is performed are tracked by the technology — as they were by *UbiFit* — it is relatively easy for the technology to determine if the user is meeting or exceeding the recommendations each week and by how much.

In our analysis of the data collected during the two *UbiFit* field studies, we classified participants’ activity levels for each full Sunday to Saturday week into one of four categories: inactive, sporadically active, reasonably active, and active. These categories were based on the ACSM and AHA’s recommendation for healthy adults as well as work done by Macera et al. [59] in a phone-based survey study<sup>2</sup> to classify weekly activity levels. Macera et al. classified respondents in their study as being *Inactive* if they reported no activity in a usual week. The ACSM and AHA’s recommendation accounts for a level of activity

<sup>2</sup>Macera et al.’s [59] study was an extension of the *Behavioral Risk Factor Surveillance System* (BRFSS, <http://www.cdc.gov/brfss/> {Link verified 30 Aug 2013}). The traditional BRFSS has usually included questions on physical activity; however, the questions focused on vigorous-intensity types of activities and did not include the more moderate-intensity types of activities, which have also been shown to provide health benefits (e.g., [39]). To address this deficiency, Macera et al. used an additional set of questions in an attempt to “*measure occupational, household, and leisuretime physical activities with a special emphasis on moderate-intensity activities*” [59]. Questions were of the form, “*How many days per week do you do these moderate activities for at least 10 minutes at a time?*” and “*On days when you do moderate activities for at least 10 minutes at a time, how much total time per day do you spend doing these activities?*” [60]. Similar questions were asked about vigorous activities.

that we called *Reasonably Active*, as it is considered to be the *minimum* recommendation to achieve health benefits. However, because a person could fall between the levels of *Inactive* and *Reasonably Active*, we added a third category for our analysis that we called *Sporadically Active*. Finally, we added a fourth category, *Active*, to account for participants who noticeably exceeded the criteria for being *Reasonably Active*. In our definition, *Active* weeks include those where participants either did (a) 30 or more minutes of walking on at least five days of the week *and* 20 or more minutes of cardio on at least three days of the week, or (b) 20 or more minutes of cardio on at least five days of the week. Table 5.1 provides definitions of the four activity level categories that we used in our analysis. Note that as with the basic ACSM & AHA recommendation, these activity levels do not account for any strength or flexibility activities performed.

Unfortunately, we found that it was not as simple as we had hoped to use these activity level definitions to assess participants' level of physical activity. Participants who appeared to be active (at least to us) often did not meet the criteria of being *Reasonably Active* (i.e., the

Table 5.1. **Weekly activity levels as used in the *UbiFit* field study analyses.** No cardio or walking event of less than 10 minutes was considered when determining activity level. The *Reasonably Active* category is equivalent to the ACSM & AHA's minimum physical activity recommendation for healthy adults.

Activity Level	Definition
Inactive	No cardio or walking events of at least 10 minutes in duration
Sporadically Active	At least one cardio or walking event $\geq 10$ min in duration, but does not meet criteria for "Reasonably Active"
Reasonably Active	At least 5 days of $\geq 30$ min of walking, but does not meet criteria for "Active" OR At least 3 days of $\geq 20$ min of cardio, but does not meet criteria for "Active" OR At least 5 days of $\geq 20$ min of cardio AND/OR $\geq 30$ min of walking WITH at least one day of each and does not meet criteria for "Active"
Active	At least 5 days of $\geq 30$ min of walking AND at least 3 days of $\geq 20$ min of cardio OR At least 5 days of $\geq 20$ min of cardio

ACSM and AHA's minimum recommendation for physical activity). Only 11 (30.6%) of the 36 weeks in the 3-week field study of *UbiFit* and 61 (19.9%) of the 306 weeks in the 3-month field study of *UbiFit* met the ACSM and AHA's minimum recommendation for activity (i.e., weeks where participants were *Reasonably Active* or *Active*). Far more weeks fell into our category of being *Sporadically Active*, that is, participants who did at least one cardio or walking event of 10 or more minutes in duration, but did not meet the minimum recommendation of being *Reasonably Active* or *Active*.

The category of *Sporadically Active* not only accounted for a majority of participants' weeks, but it also appeared to us to represent a pretty broad range of activity levels. For example, compare one of P1s and P9s *Sporadically Active* weeks to one of P3s:

- P1: 4 days of *walking*: 74 min, 87 min, 49 min, & 55 min (total: 265 min),  
6 days of *strength training*, and  
5 days of *flexibility training*: 45 min of yoga each of the 5 days<sup>3</sup>
- P9: 6 days of *walking*: 20 min, 46 min, 51 min, 22 min, 66 min, & 51 min (total: 256 min), and  
3 days of *flexibility training*: 20 min, 30 min, & 15 min of general stretching<sup>4</sup>
- P3: 2 days of *walking*: 60 min & 30 min (total: 90 min)

At least to us, P1s and P9s weeks seem substantively different from P3s, even when the strength and flexibility training are ignored. P1 logged 265 minutes of walking that week, and P9 logged 256 minutes; P3, on the other hand, only logged 90 min. However, as it is defined, neither P1s nor P9s weeks met the criteria of being *Reasonably Active*.

---

<sup>3</sup>This week of P1s does meet the criteria for being "Reasonably Active" because there are only four days when P1 walked for at least 30 minutes, and s/he did not perform any vigorous physical activity. P1 would have needed to walk for at least 30 min on one more day that week, or included at least 20 min of vigorous physical activity to meet the ACSM and AHA's minimum recommendation.

<sup>4</sup>This week of P9s does meet the criteria for being "Reasonably Active" because only four of the six days of walking included at least 30 minutes per day, and P9 did not perform any vigorous physical activity.

Imagine how P1 or P9 might have reacted to *UbiFit* had *UbiFit* given them feedback about their activity level based off of that criteria.

Interestingly, in 2008 (i.e., after our analysis was performed), the U.S. Department of Health and Human Services published “*the first comprehensive guidelines on physical activity ever to be issued by the Federal government*” [93]. In these guidelines, the recommendation for adults to achieve “substantial health benefits” is to do:

- at least 150 minutes per week of moderate-intensity aerobic physical activity
- OR at least 75 minutes per week of vigorous-intensity aerobic physical activity
- OR “*an equivalent combination of moderate- and vigorous-intensity aerobic physical activity*” [93].

They clarify that “*aerobic activity should be performed in episodes of at least 10 minutes, and preferably, it should be spread throughout the week*” [93]. Later in the document, they point out that research studies show that activity should be performed on at least 3 days a week. They also classify weekly physical activity levels into four categories: inactive, low, medium, and high (see Table 5.2 for their definitions).

Table 5.2. Weekly activity levels according to the *2008 Physical Activity Guidelines for Americans* [93].

Activity Level	Definition	Summary of Overall Health Benefits
<b>Inactive</b>	No activity beyond baseline activities of daily living	None
<b>Low</b>	Activity beyond baseline activities of daily living, but . . . <150 min of moderate-intensity physical activity OR <75 min of vigorous-intensity physical activity	Some
<b>Medium</b>	150–300 mins of moderate-intensity physical activity beyond baseline activities of daily living OR 75–150 mins of vigorous-intensity physical activity beyond baseline activities of daily living	Substantial
<b>High</b>	>300 min of activity beyond baseline activities of daily living per week	Additional

Though these new guidelines proposed by the USDHHS are rather similar to what the ACSM and AHA recommended, they would lead to a substantial reclassification of activity levels of the *UbiFit* study participants' weeks. For example, P1s and P9s weeks mentioned above would be classified as *Reasonably Active* (or *Medium*, if we used the USDHHS' proposed activity levels) rather than *Sporadically Active*, while P3s week would remain *Sporadically Active* (or *Low*) — classifications that better align with what we would have thought. This example illustrates how a slight reframing of the rules can lead to meaningful differences in the interpretation and classification of behavioral data.

### 5.1.3 Stage of Change

Yet another way to assess a person's level of physical activity is by determining her stage of change as defined by the *Transtheoretical Model of Behavior Change* (or TTM) [74]. In the TTM, Prochaska et al. describe the five stages of change through which people progress to intentionally modify addictive and other problematic behaviors, such as cigarette smoking, alcohol abuse, and obesity. These stages are intended to apply to both self-mediated as well as treatment-facilitated behavior modification. The five stages in the TTM concentrate on *intentional* or *voluntary* change, rather than societal, developmental, or imposed change. The five stages are:

1. “**Precontemplation** is the stage at which there is no intention to change behavior in the foreseeable future” [p. 3, emphasis added]. Though the person tends to be un- or underaware of her problem, those in her social network are often aware. If the person enrolls in some sort of an intervention or treatment program, it is likely due to external pressure from someone in her social network such as a spouse or employer. Though behavior change may be demonstrated by precontemplators, they often revert to their old ways as soon as the external pressure is off. It is possible for a precontemplator to wish to change, however she will have no serious intent to change within the next six months.

2. “**Contemplation** is the stage in which people are aware that a problem exists and are seriously thinking about overcoming it but have not yet made a commitment to take action” [p. 3, emphasis added]. A person can remain a contemplator for long periods of time — even years. The contemplator basically knows what she needs to do, but is not ready to do it. Contemplators tend to struggle over the tradeoffs involved, such as, the effort, cost, energy, and so on that it will take to modify the behavior.
3. “**Preparation** is a stage that combines intention and behavioral criteria. Individuals in this stage are intending to take action in the next month and have unsuccessfully taken action in the past year” [p. 4, emphasis added]. People in the preparation stage have often made some reductions in the problem behavior, but not enough to be in the next stage. For example, the person may have reduced the number of cigarettes she smokes each day, but has not stopped smoking. People in this stage have serious intent to take action in the next month.
4. “**Action** is that stage in which individuals modify their behavior, experiences, or environment in order to overcome their problems. Action involves the most overt behavioral changes and requires considerable commitment of time and energy” [p. 4, emphasis added]. A person in the action stage has successfully altered her behavior to meet the criterion for success for up to six months. For example, in a stop smoking program, a person in the action stage will have successfully stopped smoking for up to six months, not merely cut back the amount she smokes. People tend to receive the most positive external feedback regarding their behavior changes when they are in this stage. The action stage is often erroneously equated with successful change by the person herself, people in her social network, and sometimes even professionals. Prochaska et al. warn that it is important that this does not happen, as important efforts necessary to maintain the change may be overlooked.



5. “**Maintenance** is the stage in which people work to prevent relapse and consolidate the gains attained during action” [p. 4, emphasis added]. In the maintenance stage, the person has successfully modified her behavior to meet the criterion for success for a period of at least six months to an indeterminate period of time; for some behaviors, this stage can last a lifetime. The two hallmarks of the maintenance stage are (a) stabilizing behavior change and (b) avoiding relapse, that is, not reverting to a prior stage.

Linear progression through the stages is possible, but rare, as people typically cycle through stages several times before they ultimately terminate the addictive or problem behavior. Few first attempts to change behavior meet with success. If during a relapse, the person feels embarrassed, ashamed, or guilty, she is likely to return to the “precontemplation” stage. However, most tend to return to the contemplation or preparation stage and develop a new plan for change based on what they learned from their recent, unsuccessful attempt.

Not only have technologies that are designed to encourage and support health and wellness behaviors used the TTM to determine how the user is progressing (e.g., [57]), but they have also used the TTM to tailor behavior change strategies to the user’s current needs. For example, a technology might use different persuasive strategies on a user who is in “contemplation” than it would on a user who is in “action” (e.g., [29]).

We attempted to use the TTM to determine how participants in the 3-month field study of *UbiFit* progressed over the course of the study with respect to moderate and vigorous physical activity. For example, did participants progress along the stages of the model? Remain the same? Or did they regress? To assess participants’ stage of change, they completed a modified version of a survey instrument (i.e., the *Sample Physical Activity Questionnaire to Determine Stage of Change* [96]<sup>5</sup>)

---

<sup>5</sup>The survey from USDHHS et al. [96] is for *moderate*-intensity physical activity. We modified it to also ask about *vigorous*-intensity physical activity. The definition of “vigorous” that we used came from the same book as the survey, and the vigorous questions were repeats of the moderate questions, with the word “vigorous” substituted for “moderate.”

at the beginning of month one, end of month one, and end of month three. The intent of the survey was to serve as a repeated measure to determine how stages of change varied over the three months of the field study.

Unfortunately, our analysis yielded questionable results. First, ten of the 28 participants' changes are problematic according to the stage definitions. In Table 5.3, the bold, italic font highlights changes that did not make sense to us. Interestingly, most represent the participants' stages of change for moderate-, not vigorous-intensity, physical activity.

Table 5.3. **Participants' stage of change for moderate and vigorous physical activity according to the survey instrument** [96] in the 3-month field study of *UbiFit*. 1 = Precontemplation, 2 = Contemplation, 3 = Preparation, 4 = Action, and 5 = Maintenance. Stages in the *bold, italic font* represent inconsistencies in changes.

	ID	Moderate Physical Activity			Vigorous Physical Activity		
		Session 1	Session 2	Session 3	Session 1	Session 2	Session 3
No glanceable display	N1	5	5	1	4	3	3
	N3	<b><i>3</i></b>	<b><i>4</i></b>	<b><i>5</i></b>	3	4	4
	N4	<b><i>2</i></b>	<b><i>5</i></b>	4	2	2	3
	N5	5	5	5	2	1	2
	N6	3	3	3	2	2	2
	N7	<b><i>5</i></b>	<b><i>5</i></b>	<b><i>5</i></b>	<b><i>5</i></b>	<b><i>5</i></b>	<b><i>5</i></b>
	N8	3	3	3	3	3	3
	N9	<b><i>2</i></b>	<b><i>5</i></b>	5	2	3	1
	N10	3	3	3	3	3	3
	No fitness device	S1	5	5	5	5	3
S2		3	2	3	3	1	1
S3		5	5	5	5	4	5
S4		3	3	3	3	2	3
S5		<b><i>3</i></b>	<b><i>4</i></b>	<b><i>5</i></b>	3	3	3
S7		4	2	4	2	2	3
S8		3	<b><i>2</i></b>	<b><i>5</i></b>	<b><i>4</i></b>	<b><i>4</i></b>	<b><i>4</i></b>
S9		3	5	5	2	2	3
S10		5	5	4	5	3	4
Full system		F1	3	3	3	1	1
	F2	5	<b><i>3</i></b>	<b><i>5</i></b>	2	2	1
	F3	3	3	3	2	1	1
	F4	3	3	2	2	1	2
	F5	5	5	3	4	2	3
	F6	5	<b><i>3</i></b>	<b><i>5</i></b>	3	2	4
	F7	<b><i>2</i></b>	<b><i>4</i></b>	<b><i>5</i></b>	2	2	2
	F8	5	5	5	3	1	1
	F9	<b><i>3</i></b>	<b><i>5</i></b>	5	3	4	4
	F10	5	3	3	3	3	3

From the beginning of month one to the end of month three, N3 and S5 jumped from *Preparation* to *Maintenance* according to the survey. However, it does not make sense to go from serious intent to take action to consistently having modified the behavior for six or more months in only a three-month period. F2, F6, and F9 similarly jumped from *Preparation* to *Maintenance*, but in even less time. N4, N9, S8, and F7 made even bigger jumps, going from *Contemplation* to *Maintenance*. This jump would suggest that in the space of one to three months, those four participants went from thinking about changing behavior to successfully having changed behavior for six or more months.

Second, our analysis revealed an inconsistency in how participants responded to a survey question that we had not expected to change from session to session — that is, “*In the past, I was regularly physically active in [moderate | vigorous] activities for at least 3 months: yes/no*” [96]. Twenty-three of 28 participants answered the moderate activity version of this question consistently all three times the survey was implemented, and only 19 answered consistently for the vigorous activity version of the survey. For example, some participants originally said that they had been regularly physically active in the past, and then later said that they had not (and vice versa).

Third, we had inconsistencies when comparing the survey results to the levels of activity that we determined from the activity logs. For example, N7 was consistently scored as being in the *Maintenance* stage for both moderate- and vigorous-intensity physical activities according to the survey. However, of N7’s 11 full Sunday — Saturday weeks in the field study, he had six *Sporadically Active* weeks and five *Inactive* weeks. Similarly, S8, who at the end of the three months was scored as being in *Maintenance* for moderate activity and *Action* for vigorous activity also had six weeks of *Sporadic Activity* and five weeks of *Inactive*. Although the survey results make sense when considered alone — that is, by definition, a person could certainly be in *Maintenance* for three straight months — they do not make sense when compared to the patterns of behavior recorded in participants’ daily activity logs.

A number of issues could explain our results. For example, the common self-report concern of *social desirability bias* could be the culprit, as was also suggested by Macera et al. [59] as a potential limitation of

their survey-based study. This could explain our third example above where the two participants whose activity logs classified them as being *Sporadically Active* or *Inactive* were assessed to be in either the *Action* or *Maintenance* stages of change for both moderate- and vigorous-intensity physical activity according to their survey responses. That is, the participants may have responded in a way that would present themselves favorably to the researchers.

Another possible explanation is that participants did not consistently read the instructions carefully, that is the *misclassification bias* suggested by Macera et al. [59]. The survey instructions defined important terms such as “regular,” “moderate,” and “vigorous.” For example, [96]:

*For each of the following 10 statements, please answer ‘yes’ or ‘no.’ For these 10 statements, ‘**moderate physical activity or exercise**’ includes such activities as walking (slower than 12 min/mile), gardening, hiking, bicycling between 5–10 mph, and heavy house-cleaning. ‘**Vigorous physical activity or exercise**’ includes such activities as jogging or running, racewalking (12 min/mile or faster), walking briskly up a hill, and bicycling faster than 10 mph.*

*For activity to be **regular**, it must **add up to 30 or more minutes per day and be done at least 5 days per week**. For example, in one day you could achieve 30 minutes of moderate activity by taking a 10-minute walk, raking leaves for 10 minutes, and washing your car for 10 minutes. [emphasis added, but was included in the instructions on the survey instrument that participants completed]*

If a participant did not carefully read the instructions at any given session, it is easy to imagine how results could be inconsistent. For example, F2 may have read the instructions carefully at Session 2 where

he was assessed as being in *Preparation* for moderate physical activity, but applied his own definition of “regular” at Session 3 where he was assessed as being in *Maintenance*.

Yet another possible explanation has to do with the variability of patterns of physical activity behavior that we observed in the activity logs. Survey questions are of the form, “*I currently participate in [moderate | vigorous] physical activity: yes/no,*” and “*I currently engage in regular [moderate | vigorous] physical activity: yes/no.*” It is not clear how participants interpreted the word “currently,” nor is it clear what the survey meant by the term, as it is not explicitly defined. Participants’ patterns of activity often varied from week to week when using the criteria of regularity that this survey instrument employs (i.e., the same criteria that was used to develop the activity level classifications described above). Perhaps the participants were answering accurately for that particular week or for very recent weeks only.

It is also possible that *recall bias* was at play. That is, before participants used the system for several weeks, they were unaware of how active (or rather inactive) they actually were. Most participants admitted to believing that they were more active than they actually were until they started keeping a record of their physical activity behavior — a finding that is consistent with results from the 3-week field study of *UbiFit* as well as our prior work with *Houston*. This would mean that responses to surveys completed at the beginning of the study were likely overestimates for most participants. This could explain participants whose stage of change regressed.

All of this points to problems for using the TTM as an activity assessment mechanism, especially if the user’s stage is being assessed via a self-report mechanism like a survey. Versions of the above biases could be at play if systems try to use the TTM for individualized self-assessments. We anticipate a similar problem if a technology were to simply apply the TTM definitions to assess users over longer periods of time. What happens if she has a temporary lapse while in the *Maintenance* stage? If she comes down with the flu or has a major deadline at work and breaks her routine for a week, does that mean that she regresses from *Maintenance* to *Action* and has to start her six-month

clock over again? What if her break is two or three weeks long? Basically, at what point is her behavior no longer considered *regular*? Similar to the problems with the guidelines above, what are the precise rules that a technology should use to make these assessments?

We would like to note that though we faced problems with using the TTM as a way to assess participants' starting levels and measure their progress, we did find it helpful when we were designing *UbiFit* and our other systems. It served as an important (though arguably obvious) reminder that people are different and need different strategies to support them as they attempt to change their health and wellness behaviors. Also, it reminded us that even the same person may need to be supported with different strategies as she attempts to change her behavior over time — which we address further in our section below on how to support the user over her lifespan.

## 5.2 Supporting the User When “Stuff” Happens

After several years of researching and designing technologies to encourage participation in regular physical activity, one thing has become very clear (though is admittedly not surprising): *stuff happens*. Life is complicated, and therefore designing technologies to support health and wellness behaviors is complicated. As such, any technology that is being developed to encourage health and wellness must fit into people's complicated lives while respecting their competing priorities. Even the best, most devoted person may need an occasional break from her routine. People get sick every now and again, as do people for whom they provide care (such as their children or elderly parents). In fact, according to the National Institute of Allergy and Infectious Diseases [67], the common cold is a leading cause of missed days from work and school. Work or school deadlines come up, people change jobs, have new relationships, take vacations, travel for business, or even move. In some cases, occasional “breaks” in behavior can actually help people maintain the wellness behavior over time. Therefore, maintaining certain behaviors such as being physically active or eating well is often more about the person's overall pattern of behavior than a strict adherence to a routine. Successfully maintaining a physically

active lifestyle may not need to look the same as being a successful non-smoker.<sup>6</sup>

How the technology should support the user's goal of being healthy when life gets in the way remains an open question. How can the technology sustain the user's interest in the health and wellness behavior? How can the technology help her get back on track as soon as possible? How can it facilitate these goals without becoming a pest or otherwise causing the user to abandon the technology (or worse, abandon her health goals)?

### **5.3 Supporting the User Over Her Lifespan**

Finally, a key characteristic of wellness behaviors, such as being physically active and eating a healthy diet, is that their value predominantly arises from performing these behaviors regularly and over time. People certainly benefit from one-time health behaviors such as a vaccination or a screening test. However, the benefits of physical activity, diet, sleep, mindfulness practices and other such activities accrue from their practice, day in and day out, over the course of the person's lifespan. A couple of months of physical activity is better than not doing any at all, but most of the benefits — having a healthy cardiovascular system, maintaining a healthy weight, having energy and focus — require regular physical activity over the long-term.

This requires that wellness behaviors be maintained through the many changes and transitions that constitute a human life: growing up, moving away from home, going to school, getting and changing jobs, having children, getting sick and recovering, and so on. At different periods of a person's life, wellness behaviors may be more or less of her focus, and the resources (time, energy, strength, money) that she can devote to maintaining wellness will vary. From the perspective of health promotion, the goal is to maintain at least some focus on wellness across these changes.

---

<sup>6</sup>We note that the usefulness of the strict application of rules is domain dependent, even within health and wellness (e.g., medication compliance vs. healthy eating). Our experience is mainly in using technology to encourage adults to participate in regular physical activity to support a healthy lifestyle.

Health interventions typically do not approach wellness from this long-view perspective. Although the importance of wellness behaviors across the lifespan is clear, the concrete interventions that have been developed to support wellness — including those that use mobile technology — only focus on one aspect of this process or another: encouraging teenagers to be more physically active (e.g., Toscos et al.’s *Chick Clique* [88]), helping adults to manage weight through diet and exercise (e.g., Denning et al.’s *BALANCE* [26]; Tsai et al.’s *PmEB* [90]), helping heart disease patients to recover and recondition following a heart attack [36], or helping improve wellness behaviors of the elderly (e.g., [11]). As helpful as such systems may be, they do not account for the changes in users’ lives over time. The implicit assumption — both behind technological and non-technological interventions — seems to be that the role of the intervention is to help the user initially make the needed behavioral changes, and that what is needed after that is either more of the same or that things will continue to work on autopilot. The poor adherence rates, even in populations for whom maintaining wellness behaviors can literally be a question of life or death (e.g., heart disease patients following a myocardial infarction [65]), suggests that the traditional approaches have not been sufficiently effective.

We believe that supporting wellness behaviors across the lifespan is an area where technology — and particularly mobile technology — can make significant contributions. Many of the characteristics that make mobile technology a good match for wellness interventions in general — being close at hand, the ability to monitor users’ context and behaviors, always-on connectivity, and the sophistication of programming interfaces [53] — also support the development of long-term wellness solutions. In addition, what makes the current technological landscape a particularly promising starting point for the development of long-term wellness applications is that mobile applications are increasingly designed not as standalone systems but as components of suites of services, fundamentally cloud-based, that span different types of technologies. Any true long-term wellness solution has to be able to travel with the user as she changes phones and computers and as new types of sensors and wellness devices come onto the market. Until recently, only a small number of services had the ability to follow users across



applications and technological platforms — most notably email. The emergence of cloud-based applications with rich native mobile clients makes this model, and the promise of long-term use, a possibility for a much broader range of applications, including wellness applications.

How to design applications that work effectively over years or decades of use is an open question. Although we do not know what forms such long-term use applications will take, we believe that they will likely need to employ strategies for supporting healthy behaviors that are rarely used in current systems. The following three strategies seem to us as likely candidates, but others will surely emerge over time.

### **5.3.1 Deep Personalization**

A key promise of long-term-use mobile technologies is that they can, over time, learn their users' patterns. By capturing information about users' health-related behaviors and the various contexts in which health-promoting and health-damaging behaviors take place, future wellness applications will be able to create and refine highly personal, dynamic models of the users' behavior and the factors that influence it. An application can use such models to determine which behaviors and determinants to focus on at different times, when the user is most likely to slip (and when, for example, a just-in-time intervention is needed [42]), and which changes in behavior are likely to be temporary (e.g., a person's physical activity level dropped because she has a deadline at work or got the flu) and which ones are indicative of a deeper trend of which the user should be made aware.

Realizing this type of personalization will require contributions from a number of disciplines, including sensing, machine learning, security, and agent modeling, among others. But getting the personalization right will require a great deal of work from the HCI community as well. For instance, while mobile technologies are able to automatically detect an increasing number of human behaviors and features of the environment, not everything can be sensed. Many important determinants of human behavior, from cognitive constructs such as goals and self-efficacy to the meaning of places and relationships (knowing that Person X is the user's colleague but Person Y is the user's manager), are

not the kinds of things that can be completely inferred from physical signals. An accurate model of a person's behavior will need to incorporate such determinants, which means that the information about them and their changes will need to be collected on an ongoing basis. How that can be done in an unobtrusive, privacy-observant, secure, low-effort way, so that such data collection can be maintained over many years is an area that HCI researchers and designers will need to investigate. Similarly, how exactly an application should respond to a change in a person's behavior, or what types of interventions users will be willing to use over long periods of time are all questions that will need to be answered by the HCI community.

### **5.3.2 Adaptation**

As a person's behavior and circumstances change, to provide optimal support, a long-term-use technology will need to adapt. Consider self-monitoring of diet, for instance. As we mentioned, detailed tracking of individual foods is a laborious process that users are rarely able to maintain for very long. Yet, research indicates that such tracking is very helpful for changing one's diet and for initiating weight loss. Designers of current wellness applications choose one way to track diet (e.g., detailed tracking in *LoseIt*), and that's how the application continues to work whether it's used for two weeks or for two years. Applications intended for long-term use might want to take a different approach. When a person is initiating a change in diet, the application could default to detailed tracking to provide the user with a better understanding of her current dietary behavior and to help her make the initial changes in how she eats. If, however, the application notices that the user's diet has stabilized and that she is consistently logging healthy foods, the application might switch to a less-laborious mode where the user logs only the categories of foods but not individual food items. If the user's diet remains on track, the self-monitoring could potentially turn off altogether. If at some point later the user's weight begins to creep back up or the application notices that the user is increasing how often she goes to fast-food restaurants, the application could notify her about the trend and the diet tracking module could turn on again.

Other types of adaptations are possible as well. For instance, an application could track how effective different strategies are for a particular user and then adapt its behavior to use the most effective strategies. To encourage wellness, the system might suggest, over time, that a user set different types of goals: to decrease sedentary time, to increase walking time, to reduce total calorie intake, to increase intake of fruit and vegetables, or combinations of such goals. The system could then track how well the user does in terms of diet and physical activity with different types of goals, and depending on the results, it could adjust its goal-setting suggestions toward the kinds of goals that are most effective for that user.

As with personalization, developing effective adaptive technologies will require contributions from many areas. The use of reinforcement learning [85], for example, could enable the kind of adaptive goal-setting we just described. However, questions such as what the action choices need to be, or how to transition from one mode of functioning to another without confusing and frustrating the user, are fundamentally HCI issues that will need to be explored by HCI researchers and designers.

### 5.3.3 **Alternate Motivational Strategies**

Finally, for a wellness system to be effective over many years, it will need to understand — and adapt to — the changing motivations of its users. At different life stages and in different situations people care about different things. A healthy 20-year-old may have few worries about her health and have difficulty fully grasping how her long-term health will be affected if she doesn't take care of herself now. For her, a focus on staying healthy — or even maintaining a healthy weight (if she is not overweight) — may not resonate deeply with her. However, that same 20-year-old might be very concerned about the environment and the impact that people are having on the Earth. She also might be strapped for money and looking for places to cut her expenses. For such a person, framing the value of walking and biking in terms of saving money or reducing her carbon footprint might be a very effective strategy for motivating physical activity. A wellness application that knows

the user's broader goals and aspirations could adjust its behavior to provide feedback on health and wellness in terms of these other factors that the user cares about.<sup>7</sup> Later in life, however, losing weight might become a much stronger motivation than protecting the environment and the system could adapt accordingly. Still later, the person's deep motivation for staying active might be to have enough energy to keep up with her small children and other obligations. And so on. For a different person, the motivations at the same ages might be very different and to support her wellness, a long-term-use application would need to act differently.

We are proposing that one way that a wellness system can truly accompany a person through her life is by knowing and adapting to the various things that she deeply cares about, even if these things might not be directly about health. Maintaining a robust user profile that includes information about such goals and aspirations would be necessary for this strategy to work, as would finding non-trivial ways in which healthy activities could be framed in terms of those goals and aspirations. How to do either of these things well, and in a privacy-observant, secure way, is currently uncharted territory. With their expertise in understanding users' needs and experience, and in translating those needs into compelling designs, HCI researchers and designers are uniquely suited to make progress on this challenge. The payback for those efforts, we believe, would be a new class of applications that support health and wellness in deeper and more personal ways.

## 5.4 Section 5 Wrap-Up

In this section, we suggested several challenges in the space of designing mobile technologies to support health and wellness that need further investigation by the HCI community. We discussed open questions that remain in figuring out how to best assess the user's starting level and progress regarding the target behavior, drawing from our own experiences with the President's Council on Physical Fitness and Sports'

---

<sup>7</sup>We made a first attempt at doing something similar to what we are proposing here with our colleagues in UbiGreen [33].

*Walking Works* program [73], the American College of Sports Medicine (ACSM) and American Heart Association's (AHA) *Physical Activity Guidelines* [39], and the *Transtheoretical Model of Behavior Change* [74]. We also suggested several other opportunities for future work including how to support the user when “stuff” happens, and how to support the user over her lifespan.

## 5.5 Final Thoughts

As the rates of lifestyle diseases such as obesity, diabetes, and heart disease continue to rise, the development of effective tools that can help people adopt and sustain healthier habits is becoming ever more important. We believe that for people who decide to pursue a healthy lifestyle, mobile technology can play a key role in enabling them to take control of their health. These powerful, context-aware, connected, personal, and always close-at-hand devices can reach a person in ways that no previous method for supporting behavior change could.

Yet, for the potential of mobile wellness systems to be realized, they need to be well designed. In this monograph we outlined important considerations that the design of such tools involves — with a focus on collecting behavioral data, providing self-monitoring feedback, and supporting goal-setting. We also suggested many open questions in those areas as well as in other areas that still need to be investigated for mobile health and wellness technologies to be truly effective. What is clear from our work and the work of many others is that this is a complicated space for which to design. Though plenty of great and helpful work has already been done, much more remains, and the HCI community is well-positioned to make important contributions to mobile health and wellness technologies.

## Acknowledgments

---

We would like to thank the many people who helped make this work happen. In particular,

- Kate Everitt and Ian Smith — our co-authors on *Houston*;
- Daniel Avrahami, Richard Beckwith, Mike Chen, Tanzeem Choudhury, Jon Froehlich, Beverly Harrison, Jeff Hightower, Anthony LaMarca, Louis LeGrand, Ryan Libby, and Keith Mosher, — our co-authors on *UbiFit*;
- Sean Munson — our co-author on *GoalPost*;
- Jimi Huh — for her helpful comments on Section 3;
- Dan Russell — for his helpful comments;
- and the many other friends, family, colleagues, pilot testers, study participants, paper reviewers, and members of the press who have contributed along the way.

## References

---

- [1] A. Ahtinen, E. Mattila, A. Vaatanen, L. Hynninen, J. Salminen, E. Koskinen, and K. Laine, “User experiences of mobile wellness applications in health promotion: User study of Wellness Diary, Mobile Coach and SelfRelax,” in *Proceedings of Pervasive Computing Technologies for Healthcare: Pervasive Health '09*, pp. 1–8, London, UK, 2009.
- [2] I. Anderson, J. Maitland, S. Sherwood, L. Barkhuus, M. Chalmers, M. Hall, B. Brown, and H. Muller, “Shakra: Tracking and sharing daily activity levels with unaugmented mobile phones,” *Mobile Networks and Applications*, vol. 12, no. 2–3, pp. 185–199, 2007.
- [3] A. H. Andrew, G. Borriello, and J. Fogarty, “Simplifying mobile phone food diaries,” in *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare: Pervasive Health '13*, Venice, Italy, 2013.
- [4] J. Anhøj and C. Møldrup, “Feasibility of collecting diary data from asthma patients through mobile phones and SMS (short message service): Response rate analysis and focus group evaluation from a pilot study,” *Journal of Medical Internet Research*, vol. 6, no. 4, p. e42, 2004.
- [5] E. Arsand, N. Tatara, G. Ostengen, and G. Hartvigsen, “Mobile phone-based self-management tools for type 2 diabetes: The few touch application,” *Journal of Diabetes Science and Technology*, vol. 4, no. 2, pp. 328–336, 2010.
- [6] E. Barrett-Connor, “Nutrition epidemiology: How do we know what they ate?,” *The American Journal of Clinical Nutrition*, vol. 54, no. (1 Supplement), pp. 182S–187S, 1991.

- [7] J. S. Bauer, S. Consolvo, B. Greenstein, J. O. Schooler, E. Wu, N. F. Watson, and J. A. Kientz, "ShutEye: Encouraging awareness of healthy sleep recommendations with a mobile, peripheral display," in *Proceedings of the Conference on Human Factors in Computing Systems: CHI 2012*, pp. 1401–1410, Austin, TX, USA, 2012.
- [8] E. P. Baumer, S. J. Katz, J. E. Freeman, P. Adams, A. L. Gonzales, J. Pollak, D. Retelny, J. Niederdeppe, C. M. Olson, and G. K. Gay, "Prescriptive persuasion and open-ended social awareness: Expanding the design space of mobile health," in *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work: CSCW '12*, pp. 475–484, Seattle, WA, 2012.
- [9] F. Bentley and K. Tollmar, "The power of mobile notifications to increase wellbeing logging behavior," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI '13*, pp. 1095–1098, Paris, France, 2013.
- [10] E. S. Berner and J. Moss, "Informatics challenges for the impending patient information explosion," *Journal of the American Medical Informatics Association: JAMIA*, vol. 12, no. 6, pp. 614–617, 2005.
- [11] T. W. Bickmore, L. Caruso, and K. Clough-Gorr, "Acceptance and usability of a relational agent interface by older urban residents," in *Proceedings of the Conference on Human Factors in Computing Systems: CHI '05*, pp. 1212–1215, Portland, OR, USA, 2005.
- [12] J. P. Bigham, C. Jayant, H. Ji, G. Little, A. Miller, R. C. Miller, R. Miller, A. Tatarowicz, B. White, S. White, and T. Yeh, "Vizwiz: Nearly real-time answers to visual questions," in *Proceedings of the Annual ACM Symposium on User Interface Software and Technology: UIST '10*, pp. 333–342, New York City, NY, USA, 2010.
- [13] K. Blondon and P. Klasnja, "Designing supportive mobile technology for stable diabetes," in *Proceedings of the International Conference on Human-Computer Interaction: HCI International '13*, 2013.
- [14] B. Brown, M. Chetty, A. Grimes, and E. Harmon, "Reflecting on health: A system for students to monitor diet and exercise," in *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, pp. 1807–1812, Quebec, Canada, 2006.
- [15] Centers for Disease Control and Prevention and U.S. Department of Health and Human Services, "Preventing obesity and chronic diseases through good nutrition and physical activity," *Preventing Chronic Diseases: Investing Wisely in Health* Retrieved from <http://www.cdc.gov/nccdphp/publications/factsheets/prevention/pdf/obesity.pdf>, 2003.
- [16] C. B. Chan, E. Spangler, J. Valcour, and C. Tudor-Locke, "Cross-sectional relationship of pedometer-determined ambulatory activity to indicators of health," *Obesity Research*, vol. 11, no. 12, 2003.
- [17] T. Choudhury, S. Consolvo, B. Harrison, J. Hightower, A. LaMarca, L. LeGrand, A. Rahimi, A. Rea, G. Borriello, B. Hemingway, P. Klasnja, K. Koscher, J. A. Landay, J. Lester, D. Wyatt, and D. Haehnel, "The mobile sensing platform: An embedded activity recognition system," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 32–41, 2008.



- [18] S. A. Clemes and R. A. Parker, "Increasing our understanding of reactivity to pedometers in adults," *Medicine and Science in Sports and Exercise*, vol. 41, no. 3, pp. 674–680, 2009.
- [19] S. Consolvo, K. Everitt, I. Smith, and J. A. Landay, "Design requirements for technologies that encourage physical activity," in *Proceedings of the Conference on Human Factors and Computing Systems: CHI '06*, pp. 457–466, Quebec, Canada, 2006.
- [20] S. Consolvo, P. Klasnja, D. W. McDonald, D. Avrahami, J. Froehlich, L. LeGrand, R. Libby, K. Mosher, and J. A. Landay, "Flowers or a robot army? encouraging awareness and activity with personal, mobile displays," in *Proceedings of the International Conference on Ubiquitous Computing: UbiComp '08*, pp. 54–63, Seoul, Korea, 2008.
- [21] S. Consolvo, P. Klasnja, D. W. McDonald, and J. A. Landay, "Goal-Setting considerations for persuasive technologies that encourage physical activity," in *Proceedings of the International Conference on Persuasive Technology: Persuasive '09*, Claremont, CA, USA, 2009.
- [22] S. Consolvo, D. W. McDonald, T. Toscos, M. Chen, J. E. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. A. Landay, "Activity sensing in the wild: A field trial of *UbiFit* garden," in *Proceedings of the Conference on Human Factors in Computing Systems: CHI '08*, pp. 1797–1806, Florence, Italy, 2008.
- [23] Consumer Health Information Corporation, "Motivating patients to use smartphone health apps," McLean, VA, Retrieved from <http://www.prweb.com/releases/2011/04/prweb5268884.htm>, April 25, 2011.
- [24] L. T. Cowan, S. A. Van Wagenen, B. A. Brown, R. J. Hedin, Y. Seino-Stephan, P. Cougar Hall, and J. H. West, "Apps of steel: Are exercise apps providing consumers with realistic expectations?: A content analysis of exercise apps for presence of behavior change theory," *Health Education and Behavior*, 2012.
- [25] J. Dallery, R. N. Cassidy, and B. R. Raiff, "Single-Case experimental designs to evaluate novel technology-based health interventions," *Journal of Medical Internet Research*, vol. 15, no. 2, p. e22, 2013.
- [26] T. Denning, A. Andrew, R. Chaudhri, C. Hartung, J. Lester, G. Borriello, and G. Duncan, "BALANCE: Towards a usable pervasive wellness application with accurate activity inference," in *Proceedings of the Workshop on Mobile Computing Systems and Applications*, pp. 1–6, 2009.
- [27] A. K. Dey, K. Wac, D. Ferreira, K. Tassini, J. Hong, and J. Ramos, "Getting closer: An empirical investigation of the proximity of user to their smart phones," in *Proceedings of the International Conference on Ubiquitous Computing: UbiComp '11*, pp. 163–172, Beijing, China, 2011.
- [28] E. E. Goffman, *The Presentation of Self in Everyday Life*. New York, NY, USA: Doubleday Anchor, 1959.
- [29] R. Farfanzar, S. Frishkopf, J. Migneault, and R. Friedman, "Telephone-linked care for physical activity: A qualitative evaluation of the use patterns of information technology program for patients," *Journal of Biomedical Informatics*, vol. 38, no. 3, pp. 220–228, 2005.

- [30] G. M. Fitzsimons and J. A. Bargh, "Automatic self-regulation," in *Handbook of Self-Regulation: Research, Theory, and Applications*, (K. D. Vohs and R. F. Baumeister, eds.), pp. 151–170, New York: The Guilford Press, 2004.
- [31] B. S. Fjeldsoe, A. L. Marshall, and Y. D. Miller, "Behavior change interventions delivered by mobile telephone short-message service," *American Journal of Preventive Medicine*, vol. 36, no. 2, pp. 165–173, 2009.
- [32] S. Fox and M. Duggan, "Mobile Health 2012, Washington, DC: Pew Internet and American Life Project," Retrieved from <http://www.pewinternet.org/Reports/2012/Mobile-Health.aspx>, 2012.
- [33] J. Froehlich, T. Dillahunt, P. Klasnja, J. Mankoff, S. Consolvo, B. Harrison, and J. A. Landay, "UbiGreen: Investigating a mobile tool for tracking and supporting green transportation habits," in *Proceedings of the International Conference on Human Factors in Computing Systems: CHI '09*, pp. 1043–1052, Boston, MA, USA, 2009.
- [34] M. Galesic and R. Garcia-Retamero, "Graph literacy: A cross-cultural comparison," *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, vol. 31, no. 3, pp. 444–457, 2011.
- [35] R. Gasser, D. Brodbeck, M. Degen, J. Luthiger, R. Wyss, and S. Reichlin, "Persuasiveness of a mobile lifestyle coaching application using social facilitation," in *Proceedings of the International Conference on Persuasive Technology*, pp. 27–38, Eindhoven, The Netherlands, 2006.
- [36] V. Gay, P. Leijdekkers, and E. Barin, "A mobile rehabilitation application for the remote monitoring of cardiac patients after a heart attack or a coronary bypass surgery," in *Proceedings of the International Conference on Pervasive Technologies Related to Assistive Environments: PETRA '09*, pp. 1–7, Corfu, Greece, 2009.
- [37] A. Grimes, M. Bednar, J. D. Bolter, and R. E. Grinter, "EatWell: Sharing nutrition-related memories in a low-income community," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing: CSCW '08*, San Diego, CA, USA, 2008.
- [38] A. Grimes, V. Kantroo, and R. E. Grinter, "Let's play!: Mobile health games for adults," in *Proceedings of the ACM International Conference on Ubiquitous Computing: UbiComp '10*, pp. 241–250, 2010.
- [39] W. L. Haskell, I.-M. Lee, R. R. Pate, K. E. Powell, S. N. Blair, B. A. Franklin, C. A. Macera, G. W. Heath, P. D. Thompson, and A. Bauman, "Physical activity and public health: Updated recommendation for adults from the american college of sports medicine and the american heart association," *Circulation*, vol. 116, pp. 1081–1093, 2007.
- [40] E. T. Higgins, "Knowledge activation: Accessibility, applicability, and salience," in *Social Psychology: Handbook of Basic Principles*, (E. T. Higgins and A. W. Kruglanski, eds.), pp. 133–168, New York: Guilford Press, 1996.
- [41] C. Hoffman, D. Rice, and H. Y. Sung, "Persons with chronic conditions. Their prevalence and costs," *JAMA: The Journal of the American Medical Association*, vol. 276, no. 18, pp. 1473–1479, 1996.
- [42] S. S. Intille, "Ubiquitous computing technology for just-in-time motivation of behavior change," *Studies in Health Technology and Informatics*, vol. 107, no. Pt 2, pp. 1434–1437, 2004.

- [43] A. E. Järvi, B. E. Karlström, Y. E. Granfeldt, I. E. Björck, N. G. Asp, and B. O. Vessby, “Improved glycemic control and lipid profile and normalized fibrinolytic activity on a low-glycemic index diet in type 2 diabetic patients,” *Diabetes Care*, vol. 22, no. 1, pp. 10–18, 1999.
- [44] T. Joutou and K. Yanai, “A food image recognition system with multiple kernel learning,” in *Proceedings of the International Conference on Image Processing: ICIP '09*, pp. 285–288, 2009.
- [45] A. E. Kazdin, “Reactive self-monitoring: The effects of response desirability, goal setting, and feedback,” *Journal of Consulting and Clinical Psychology*, vol. 42, pp. 704–716, 1974.
- [46] E. T. Kennedy, J. Ohls, S. Carlson, and K. Fleming, “The healthy eating index: Design and applications,” *Journal of the American Dietetic Association*, vol. 95, no. 10, pp. 1103–1108, 1995.
- [47] A. C. King, E. B. Hekler, L. A. Grieco, S. J. Winter, J. L. Sheats, M. P. Buman, B. Banerjee, T. N. Robinson, and J. Cirimele, “Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults,” *PLoS ONE*, vol. 8, no. 4, p. e62613, 2013.
- [48] D. E. King, A. G. Mainous, M. Carnemolla, and C. J. Everett, “Adherence to healthy lifestyle habits in US adults, 1988–2006,” *The American Journal of Medicine*, vol. 122, no. 6, pp. 528–534, 2009.
- [49] K. Kitamura, C. de Silva, T. Yamasaki, and K. Aizawa, “Image processing based approach to food balance analysis for personal food logging,” in *Proceedings of the 2010 IEEE International Conference on Multimedia and Expo: ICME '10*, pp. 625–630, 2010.
- [50] P. Klasnja, S. Consolvo, T. Choudhury, R. Beckwith, and J. Hightower, “Exploring privacy concerns about personal sensing,” in *Proceedings of the International Conference on Pervasive Computing: Pervasive 2009*, pp. 176–183, Nara, Japan, 2009.
- [51] P. Klasnja, B. L. Harrison, L. LeGrand, A. LaMarca, J. Froehlich, and S. E. Hudson, “Using wearable sensors and real time inference to understand human recall of routine activities,” in *Proceedings of the International Conference on Ubiquitous Computing: UbiComp '08*, Seoul, Korea, 2008.
- [52] P. Klasnja, A. Hartzler, C. Powell, and W. Pratt, “Supporting cancer patients’ unanchored health information management with mobile technology,” in *AMIA Annual Symposium Proceedings*, 2011.
- [53] P. Klasnja and W. Pratt, “Healthcare in the pocket: Mapping the space of mobile-phone health interventions,” *Journal of Biomedical Informatics*, vol. 45, no. 1, pp. 184–198, 2012.
- [54] J. Kopp, “Self-monitoring: A literature review of research and practice,” *Social Work Research and Abstracts*, vol. 24, pp. 8–20, 1988.
- [55] W. J. Korotitsch and R. O. Nelson-Gray, “An overview of self-monitoring research in assessment and treatment,” *Psychological Assessment*, vol. 11, no. 4, pp. 415–425, 1999.
- [56] N. D. Lane, M. Mohammad, M. Lin, X. Yang, H. Lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, and A. Campbell, “BeWell: A smartphone

- application to monitor, model and promote wellbeing,” in *International Conference on Pervasive Computing Technologies for Healthcare: Pervasive Health 2011*, Dublin, Ireland, 2011.
- [57] J. L. Lin, L. Mamykina, S. Lindtner, G. Delajoux, and H. B. Strub, “Fish’n’steps: Encouraging physical activity with an interactive computer game,” in *Proceedings of the International Conference on Ubiquitous Computing: Ubicomp 2006*, pp. 261–278, Orange County, CA, USA, 2006.
- [58] E. A. Locke and G. P. Latham, “Building a practically useful theory of goal setting and task motivation: A 35-year Odyssey,” *American Psychologist*, vol. 57, no. 9, pp. 705–717, 2002.
- [59] C. A. Macera, S. A. Ham, M. M. Yore, D. A. Jones, B. E. Ainsworth, D. Kimsey, and H. W. Kohl III, “Prevalence of physical activity in the United States: Behavioral risk factor surveillance system, 2001,” *Preventing Chronic Disease: Public Health Research, Practice, and Policy*, vol. 2, no. 2, 2005.
- [60] C. A. Macera, D. A. Jones, M. M. Yore, S. A. Ham, H. W. Kohl, C. D. Kimsey, and D. Buchner, “Prevalence of physical activity, including lifestyle activities among adults — United States, 2000–2001,” *Morbidity and Mortality Weekly Report: MMWR*, vol. 52, no. 32, pp. 764–769, 2003.
- [61] A. Macvean and J. Robertson, “IFitQuest: A school based study of a mobile location-aware exergame for adolescents,” in *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services: MobileHCI '12*, pp. 359–368, ACM, 2012.
- [62] L. Mamykina, A. D. Miller, C. Grevet, Y. Medynskiy, M. A. Terry, E. D. Mynatt, and P. R. Davidson, “Examining the impact of collaborative tagging on sensemaking in nutrition management,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: CHI '11*, pp. 657–666, Vancouver, BC, Canada, 2011.
- [63] L. Mamykina, E. Mynatt, P. Davidson, and D. Greenblatt, “MAHI: Investigation of social scaffolding for reflective thinking in diabetes management,” in *Proceeding of the SIGCHI Conference on Human Factors in Computing Systems: CHI '08*, pp. 477–486, Florence, Italy, 2008.
- [64] E. Mattila, J. Pärkkä, M. Hermersdorf, J. Kaasinen, J. Vainio, K. Samposalo, J. Merilahti, K. Kolari, M. Kulju, R. Lappalainen, and I. Korhonen, “Mobile diary for wellness management—results on usage and usability in two user studies,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 4, pp. 501–512, 2008.
- [65] R. R. Miller, A. E. Sales, B. Kopjar, S. D. Fihn, and C. L. Bryson, “Adherence to heart-healthy behaviors in a sample of the U.S. Population,” *Preventing Chronic Disease*, vol. 2, no. 2, p. A18, 2005.
- [66] S. A. Munson and S. Consolvo, “Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity,” in *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare: Pervasive Health '12*, San Diego, CA, USA, 2012.
- [67] National Institute of Allergy and Infectious Diseases *Common Cold*, Retrieved from <http://www.niaid.nih.gov/topics/commoncold/Pages/overview.aspx>.

- [68] R. O. Nelson, "Assessment and therapeutic functions of self-monitoring," in *Progress in Behavior Modification*, (M. R. Hersen, M. Eisler, and P. M. Miller, eds.), New York: Academic Press, 1977.
- [69] J. Noronha, E. Hysen, H. Zhang, and K. Z. Gajos, "PlateMate: Crowdsourcing nutritional analysis from food photographs," in *Proceedings of the Annual ACM Symposium on User Interface Software and Technology: UIST '11*, pp. 1–12, Santa Barbara, CA, USA, 2011.
- [70] R. Ozdoba, C. B. Corbin, and G. Le Masurier, "Does reactivity exist in children when measuring activity levels with unsealed pedometers?," *Pediatric Exercise Science*, vol. 16, no. 2, pp. 158–166, 2004.
- [71] E. Peters, "Beyond comprehension: The role of numeracy in judgments and decisions," *Current Directions in Psychological Science*, vol. 21, no. 1, pp. 31–35, 2012.
- [72] J. Pollak, G. Gay, S. Byrne, E. Wagner, D. Retelny, and L. Humphreys, "It's time to eat! Using mobile games to promote healthy eating," *IEEE Pervasive Computing*, vol. 9, no. 3, pp. 21–27, 2010.
- [73] President's Council on Physical Fitness and Sports. *Walking Works: The Blue Program for a Healthier American*, 2004.
- [74] J. O. Prochaska, C. C. DiClemente, and J. C. Norcross, "In search of how people change: Applications to addictive behaviors," *American Psychologist*, vol. 47, no. 9, pp. 1102–1114, 1992.
- [75] N. P. Pronk, L. H. Anderson, A. L. Crain, B. C. Martinson, P. J. O'Connor, N. E. Sherwood, and R. R. Whitebird, "Meeting recommendations for multiple healthy lifestyle factors. Prevalence, clustering, and predictors among adolescent, adult, and senior health plan members," *American Journal of Preventive Medicine*, vol. 27, no. 2 Suppl, pp. 25–33, 2004.
- [76] H. Rachlin, "Self-control," *Behaviorism*, vol. 2, pp. 94–108, 1974.
- [77] A. Raij, A. Ghosh, S. Kumar, and M. Srivastava, "Privacy risks emerging from the adoption of innocuous wearable sensors in the mobile environment," in *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems: CHI '11*, pp. 11–20, Vancouver, BC, Canada, 2011.
- [78] L. Rainie, "Cell phone ownership hits 91% of adults," *Pew Research Center's FactTank: News in the Numbers*, Retrieved from <http://www.pewresearch.org/fact-tank/2013/06/06/cell-phone-ownership-hits-91-of-adults/>, 2013.
- [79] T. A. Ryan, *Intentional Behavior*. New York: Ronald Press, 1970.
- [80] T. S. Saponas, J. Lester, J. E. Froehlich, J. Fogarty, and J. A. Landay, "iLearn on the iPhone: Real-Time Human Activity Classification on Commodity Mobile Phones," UW CSE Technical Report. Retrieved from <http://dub.washington.edu/djangosite/media/papers/UW-CSE-08-04-02.pdf>, 2008.
- [81] D. A. Schoeller, L. G. Bandini, and W. H. Dietz, "Inaccuracies in self-reported intake identified by comparison with the doubly labelled water method," *Canadian Journal of Physiology and Pharmacology*, vol. 68, no. 7, pp. 941–949, 1990.
- [82] M. K. Shilts, M. Horowitz, and M. S. Townsend, "Goal setting as a strategy for dietary and physical activity behavior change: A review of the literature," *American Journal of Health Promotion*, vol. 19, no. 2, pp. 81–93, 2004.

- [83] K. A. Siek, K. H. Connelly, Y. Rogers, P. Rohwer, D. Lambert, and J. L. Welch, "When do we eat? An evaluation of food items input into an electronic food monitoring application," in *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare: Pervasive Health '06*, pp. 1–10, Innsbruck, Austria, 2006.
- [84] A. Smith, "Smartphone ownership 2013," *Pew Internet and American Life Project*, Retrieved from <http://pewinternet.org/Reports/2013/Smartphone-Ownership-2013/Findings.aspx>, 2013.
- [85] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.
- [86] M. Tentori, G. R. Hayes, and M. Reddy, "Pervasive computing for hospital, chronic, and preventive care," *Foundations and Trends® in Human-Computer Interaction*, vol. 5, no. 1, pp. 1–95, 2011.
- [87] K. Tollmar, F. Bentley, and C. Viedma, "Mobile health mashups: Making sense of multiple streams of wellbeing and contextual data for presentation on a mobile device," in *Proceedings of the International Conference on Pervasive Computing Technologies for Healthcare: Pervasive Health '12*, pp. 65–72, San Diego, CA, USA, 2012.
- [88] T. Toscos, A. Faber, S. An, and M. P. Gandhi, "Chick clique: Persuasive technology to motivate teenage girls to exercise," in *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, pp. 1873–1878, Montreal, Quebec, Canada, 2006.
- [89] F. A. Treiber, T. Baranowski, D. S. Braden, W. B. Strong, M. Levy, and W. Knox, "Social support for exercise: Relationship to physical activity in young adults," *Preventive Medicine*, vol. 20, pp. 737–750, 1991.
- [90] C. C. Tsai, G. Lee, F. Raab, G. J. Norman, T. Sohn, W. Griswold, and K. Patrick, "Usability and feasibility of pmeb: A mobile phone application for monitoring real time caloric balance," *Mobile Networks and Applications*, vol. 12, no. 2–3, pp. 173–184, 2007.
- [91] C. Tudor-Locke, "Taking steps toward increased physical activity: Using pedometers to measure and motivate," *President's Council on Physical Fitness and Sports: Research Digest*, vol. 3, no. 17, 2002.
- [92] C. Tudor-Locke and D. R. Bassett Jr, "How many steps/day are enough?," *Sports Medicine*, vol. 34, no. 1, pp. 1–8, 2004.
- [93] U.S. Department of Health and Human Services *2008 Physical Activity Guidelines for Americans*. Retrieved from <http://www.health.gov/paguidelines/>, 2008.
- [94] U.S. Department of Health and Human Services and U.S. Department of Agriculture, "Dietary Guidelines for Americans, 2005," Retrieved from <http://www.health.gov/dietaryguidelines/dga2005/document/pdf/DGA2005.pdf>, 2005.
- [95] U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion and The President's Council on Physical Fitness and Sports. *Physical Activity and Health: A Report of the Surgeon General*, 1996.

- [96] U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Nutrition and Physical Activity, *Promoting Physical Activity: A Guide for Community Action*. Champaign, IL: Human Kinetics, 1999.
- [97] S. D. Vincent and R. P. Pangrazi, "Does reactivity exist in children when measuring activity levels with pedometers?," *Pediatric Exercise Science*, vol. 14, no. 1, pp. 56–63, 2002.
- [98] D. Walters, A. Sarela, A. Fairfull, K. Neighbour, C. Cowen, B. Stephens, T. Sellwood, B. Sellwood, M. Steer, M. Aust, R. Francis, C.-K. Lee, S. Hoffman, G. Brealey, and M. Karunanithi, "A mobile phone-based care model for outpatient cardiac rehabilitation: The care assessment platform (CAP)," *BMC Cardiovascular Disorders*, vol. 10, no. 1, 2010.
- [99] M. H. Whaley, P. H. Brubaker, and R. M. Otto, eds., "General principles of exercise prescription." *ACSM's Guidelines for Exercise Testing and Prescription*. Baltimore, MD: Lippincott Williams and Wilkins, 7th ed., 2006.
- [100] L. Zepeda and D. Deal, "Think before you eat: Photographic food diaries as intervention tools to change dietary decision making and attitudes," *International Journal of Consumer Studies*, vol. 32, no. 6, pp. 692–698, 2008.